# Analysis and Linear Algebra for Finance: Part I

Patrick Roger

Patrick Roger

# Analysis and Linear Algebra for Finance: Part I

Patrick ROGER
LaRGE Research Center
EM Strasbourg Business School
University of Strasbourg

Analysis and Linear Algebra for Finance: Part I
First edition

ISBN 978-87-403-0412-1

# Contents


360°
thinking.

## Introduction

This book is part of a sequence of books[1] written as a technical support for students in finance. In this book and the companion book[2], we complete the presentation of the mathematical tools useful to understand the theory of finance. Here, we focus on mathematical analysis and linear algebra. Part I essentially recalls concepts and properties the reader should know after (and for some students, before) an undergraduate program in economics and finance. We encourage the reader to download the four books because in any of the four books we often refer to the others.

The book is divided in four chapters. Chapter 1, entitled "Preliminaries" recalls elementary definitions and results about sets, mappings and real numbers. The second chapter deals with functions of one variable. Even if most financial models include functions of several variables, it is essential to be comfortable with concepts like continuity, limits, logarithms, exponentials, and derivatives. Chapter three explains integrals, and more precisely how Riemann integrals are built. Understanding (and keeping in mind) these integrals is also important not to be too surprised when starting the study of stochastic processes and Itô integrals that appear in continuous-time financial models.

Finally, chapter four recalls the rules of matrix calculus and elementary algebraic operations on matrices. It ends by studying the way systems of linear equations are solved. The companion book (Part II) develops vector spaces and linear mappings, functions of several variables and non-linear optimization. Therefore, students who have a clear understanding and remembering of the basic tools exposed in Part I can go quickly to Part II.

[1] Roger. P (2010a), Probability for Finance, Roger. P (2010b), Stochastic Processes for Finance, www.bookboon.com

[2] Roger. P (2013), Analysis and Linear Algebra for Finance, Part II.

# Chapter 1

# Preliminaries

This chapter recalls definitions and properties of sets, relations, mappings and sequences. Section 1.1 deals with sets and subsets, their specification, the way they are built and some elementary operations on them like union, intersection, or inclusion. Sets are fundamental mathematical objects in finance, first at the theoretical level through probability theory, and second at the empirical level, because many studies focus on some categories of stocks (that is subsets of the set of stocks traded on the financial market) or on some subsets of investors (fund managers, individual investors, young investors, etc.).

Relations are also an important concept, especially in microeconomics. In this domain, the most important relations are preference relations over bundles of goods or over lotteries when risk is introduced. In finance models, preference relations are also used on the set of portfolios. For example, in standard portfolio choice theory of Markowitz[1], a portfolio is characterized by its expected return and variance of returns. An investor is supposed to prefer a higher expected return and a lower variance of returns. She is then assumed to be able to rank portfolios and, of course, to select the best one.

Mappings also appear as an intuitive tool to establish a link between two

[1]Harry Markowitz (1952), Portfolio Selection, *The Journal of Finance*, 7(1), 77-91.

sets. For example, a lot was said recently about ratings given by agencies like Standard & Poor's, Moody's or Fitch. The mapping "Moody's" could be defined as the pair (Country, Rating) where a country name is associated to the grade given by Moody's to this country.

## 1.1 Sets and subsets

As said above, it is important for finance students to be comfortable with sets and subsets. Most models in finance are based on probability theory; probability measures are defined on subsets of a "big" set, coined by economists "the set of states of nature". This set is an abstract way to consider all possible economic situations that can show up in the future.

But sets and subsets also appear in much more concrete situations. For example, many empirical studies distinguish large market capitalizations, medium caps and small caps which are distinct subsets of the set of traded firms. Eugene Fama and Kenneth French[2] generalized the Capital Asset Pricing Model to include other factors than the return on the market portfolio. Mainly, these factors are size and book-to-market factors. It follows that in many studies in finance, firms are sorted or double sorted (by quintiles or deciles) on size and/or book-to-market.

### 1.1.1 Definition and properties

**Definition 1** *A set $E$ is a collection (not necessarily finite) of well defined objects. Members of $E$ are called* ***elements.***

*- If an object $a$ is an element of the set $E$, we write $a \in E$ ($a$ belongs to $E$).*

[2] Fama E. F., French, K. R.(1992), The Cross-Section of Expected Returns, *Journal of Finance,* 47 (2), 427-465.
Fama, E. F. and K.R. French (1993), Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics,* 33, 3-56.

*- If a is not a member of the set E, we write* $a \notin E$ *(a does not belong to E).*

**Example 2** 1. *The set of students having downloaded this book since its publication.*

2. *The set of stocks traded on the New York Stock Exchange at the end of December this year.*

3. *The set of positive integers* $\mathbb{N} = \{0; 1; ...; n; ...\}$ *or the set of all integers (positive or negative) denoted* $\mathbb{Z} = \{... - n; -n + 1; ..0; 1; ...; n; ...\}$

4. *The set of clients of Citigroup.*

5. *The set* $\mathbb{Q}$ *of rational numbers (that is numbers written* $\frac{m}{n}$ *where m,n are elements of* $\mathbb{Z}$ *,* $n \neq 0$*).*

A set can also be defined by some properties of its elements. Let $E$ be the set of solutions of the following equation:

$$E = \left\{x \in \mathbb{Q} \text{ so that } x^2 - 1 = 0\right\} \tag{1.1}$$

It appears that this set is defined implicitly because $E$ is not defined as a list of elements. In other words, the definition of $E$ necessitates neither to be able to calculate the solutions of the equation nor to know if solutions exist. Of course, finding the number of elements of $E$ may be quite tricky when the definition is implicit.

**Definition 3** *The **cardinality** of a set E, denoted* $Card(E)$ *or* $\#(E)$ *is the number of elements of E.*

We can easily prove that $Card(E) = 2$ when $E$ is the set defined by (1.1). The equation $x^2 - 1 = 0$ has two solutions, respectively equal to $-1$ and $1$. But it is not always that simple. Many sets have an infinite cardinality. It is the case for the sets $\mathbb{N}$ (positive integers) and $\mathbb{Q}$ (rational numbers). We then write $Card(\mathbb{N}) = +\infty$, $Card(\mathbb{Q}) = +\infty$ . There also exists a set with no element, called the empty set.

**Definition 4** *A set with no elements is called the* ***null*** *or* ***empty set*** *. It is denoted* $\emptyset$.

### 1.1.2 Comparison of sets: inclusion, union, intersection and complement

**Definition 5** *Let $A$ and $B$ be two sets; $A$ is* ***included*** *in $B$, and we write $A \subset B$, if any element of $A$ belongs to $B$. $A$ is called a* ***subset*** *of $B$.*

*The set of all subsets of $B$ is denoted $\mathcal{P}(B)$.*

1. The set of odd positive integers is a subset of $\mathbb{N}$, the set of positive integers.

2. If $B$ is the set of students having downloaded this book, the set $A$ of students having downloaded the book and clicked on all ads it contains is included in $B$.

3. The set of U.S stocks with a market cap greater than $1 billion is a subset of all stocks traded on the U.S market.

4. The three sets $\mathbb{N}, \mathbb{Z}$ and $\mathbb{Q}$ are related by the following inclusions:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$$

It is worth to notice that the empty set is included in any set. It implies that $\emptyset$ is always an element of the set of subsets of $B$, whatever the set $B$ under consideration. In other words, $\emptyset$ is an element of $\mathcal{P}(B)$. We come back later on to this property.

The inclusion relationship allows to define the equality of two sets.

**Definition 6** *Two sets $A$ and $B$ are equal, and we write $A = B$, if $A \subset B$ and $B \subset A$.*

We are now ready to define some elementary operations on sets, in particular the union (intersection) of two sets. These are quite natural concepts often used in empirical studies. For example, when studying the behavior of individual investors, it is interesting to know if their portfolio diversification choices have consequences on the return they obtain. A natural way to address this issue empirically is to double sort the set of investors, according to the number of different stocks they hold, and to the range of returns they obtain. Investors then belong to categories, each investor being in one and only one category.

In statistics, the independence $\chi^2$-test is also based on a classification of a sample according to two variables. If the variables are not independent, the frequencies in subsets are far from what is expected under the independence hypothesis.

Let now $E$ be a set and $A, B$ denote two subsets of $E$.

**Definition 7** *1) The* ***intersection*** *of $A$ and $B$, denoted $A \cap B$, is the set defined by:*

$$A \cap B = \{a \in E \text{ so that } a \in A \textbf{ and } a \in B\} \tag{1.2}$$

*2) The* ***union*** *of $A$ and $B$, denoted $A \cup B$, is the set defined by:*

$$A \cup B = \{a \in E \text{ so that } a \in A \textbf{ or } a \in B\}$$

*3) The* ***complement*** *of $A$ in $E$, denoted $A^c$, is the set defined by:*

$$A^c = \{a \in E \text{ so that } a \notin A\}$$

Part (3) of the above definition implies that an element $a$ cannot be at the same time in $A^c$ and in $A$. But an element $a$ belongs to $A$ or to $A^c$. This remark implies:

$$A \cap A^c = \emptyset \text{ and } A \cup A^c = E$$

**Proposition 8** *Let $A, B, C$ denote three subsets of a set $E$; intersection and* *union are distributive in the following sense:*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Simple relations are also established between the complement of an intersection (union) and the union (intersection) of the complements.

**Proposition 9** *Let $A$ and $B$ denote two subsets of a set $E$; we have:*

$$(A \cap B)^c = A^c \cup B^c$$
$$(A \cup B)^c = A^c \cap B^c$$

The first relation is easily understandable when looking at figure 1.1. When $a \notin A \cap B$, either it belongs to $A^c$ or it belongs to $B^c$. Finally, $a$ is an element of $A^c \cup B^c$. The interpretation of the second equality is similar.

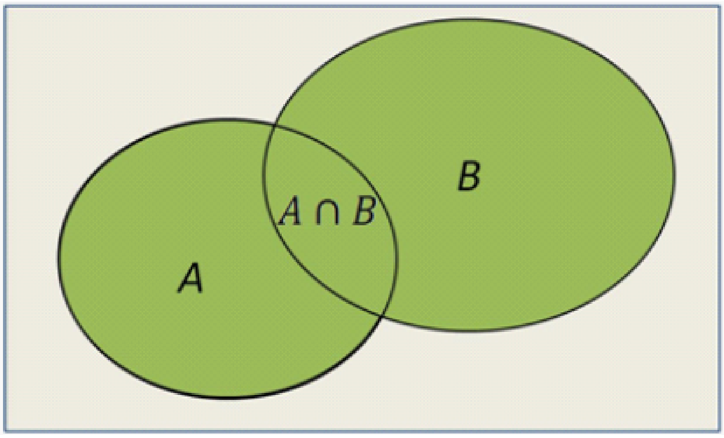


Figure 1.1: Two subsets $A$ and $B$ and their intersection

Suppose now you want to count the number of U.S firms with more than 1,000 employees (subset $A$) or with revenues greater than \$1 billion (subset $B$). If you simply add the number of firms with more than 1,000 employees and the number of firms with revenues higher than \$1 billion, you count twice the firms belonging to the two categories, that is $A \cap B$. Consequently, it is necessary to subtract the number of these firms in $A \cap B$. This calculation is written in the general case as follows.

**Proposition 10** *Let $A$ and $B$ be two subsets of a set $E$. We then have:*

$$Card(A \cup B) = Card(A) + Card(B) - Card(A \cap B) \tag{1.3}$$

Consider for example the two sets $A = \{1, 4, 7, 9\}$ and $B = \{2, 4, 5, 9\}$. We observe $Card(A) = Card(B) = 4$ but $A \cup B = \{1, 2, 4, 5, 7, 9\}$ leading to $Card(A \cup B) = 6$. Of course, $Card(A \cap B) = 2$ because $A \cap B = \{4, 9\}$.

Definition 7 can be generalized to any finite number of sets.

**Definition 11** *Let $E$ be a set and $A_1, A_2, ..., A_n$ be subsets of $E$.*

*1) The **union** of sets $A_i, i = 1, ..., n$, denoted $\cup_{i=1}^n A_i$, is the set defined by:*

$$\cup_{i=1}^n A_i = \{a \in E \text{ so that } \exists i \in \{1, 2, ..., n\} \text{ with } a \in A_i\}$$

*2) The **intersection** of sets $A_i, i = 1, ..., n$, denoted $\cap_{i=1}^n A_i$ is the set defined by:*

$$\cap_{i=1}^n A_i = \{a \in E \quad \text{so that } \forall i \in \{1, 2, ..., n\},\ a \in A_i\}$$

It is important to note that for the union of sets, we use the quantifier $\exists$ (it exists), which is associated to the logical connective OR in computer programming, whereas for the intersection we use $\forall$ (whatever) which is associated to the connective AND.

### 1.1.3 Partitions

In the particular case where pairwise intersections of subsets are empty, proposition 10 can be simplified. If, in addition, the union of sets under consideration is the complete set $E$, we get what is called a partition of $E$.

**Definition 12** *A* ***partition*** *of the set* $E$ *is a family of subsets* $A_1, A_2, ..., A_n$ *satisfying:*

$$\begin{aligned} \cup_{i=1}^{n} A_i &= E \\ \forall (i,j) &\in \{1,2,...,n\}, i \neq j \Rightarrow A_i \cap A_j = \emptyset \end{aligned}$$

**Example 13** *Consider the set of firms rated by Standard and Poor's (denoted* $E$*) and define subsets according to the grades received by firms in* $E$*. The union of all subsets is the entire set of rated firms and a firm cannot be in two subsets at a given moment. The rating system naturally defines a partition of the set of rated firms.*

### 1.1.4 The set of subsets of $E$

Remember that in financial and economic models, the possible future economic situations are represented by a set of states of nature. When this set has been defined (we still denote it $E$), it is convenient to give probabilities of occurrence to subsets of this "big" set of states[3]. In fact, in standard models where the number of elements of $E$ is finite, a probability measure is defined on all possible subsets of $E$. As mentioned before, this set of subsets is denoted $\mathcal{P}(E)$. The following proposition gives the number of elements[4] in $\mathcal{P}(E)$.

**Proposition 14** *Let $E$ denote a set satisfying $Card(E) = n$; then $\mathcal{P}(E)$ has $2^n$ elements (including the empty set).*

**Proof.** In order to demonstrate this proposition, we use a proof by induction. It is decomposed in two steps.

1) showing that the statement is true for $n = 1$;

2) showing that if the statement is true for a given $n$, it is also true for $n + 1$.

These two steps are sufficient to prove that the statement is always true.

When the set $E$ contains only one element, we have $\mathcal{P}(E) = \{\emptyset, E\}$ and thus $Card(\mathcal{P}(E)) = 2 = 2^1$. Assume now that $Card(E) = n$ and $Card(\mathcal{P}(E)) = 2^n$. We have to prove $Card(\mathcal{P}(G)) = 2^{n+1}$ where $G$ is a set with $n + 1$ elements built by adding one element (say $e_{n+1}$) to $E$.

Subsets of $G$ are divided in two categories, some subsets do not contain $e_{n+1}$ and the others contain this element. The first category is simply equal to $\mathcal{P}(E)$; its cardinality is then $2^n$. But the second one contains all the elements which write $A \cup \{e_{n+1}\}$ where $A$ is an element of $\mathcal{P}(E)$. There are also $2^n$

[3] see Roger, 2010a, *Probability for Finance,* chapter 1.

[4] In probability theory, it is important to distinguish elements of a set from subsets of cardinality 1. Indeed, when we write $E = \{1, 2, 3\}$, 2 is an element of $E$. However, when we write $\{2\}$, it is a subset of $E$, that is an element of $\mathcal{P}(E)$.

subsets in this category. Finally, we get:

$$Card(\mathcal{P}(G)) = 2^n + 2^n = 2 \times 2^n = 2^{n+1}$$

■

## 1.2 Binary relations

**Definition 15** *Let $E$ and $F$ denote two sets;* ***a binary relation*** $\mathcal{R}$ *between $E$ and $F$ is a property satisfied by a subset of couples $(x, y) \in E \times F$; we then write $x\mathcal{R}y$.*

The following definition give a few properties that a binary relation $\mathcal{R}$ might satisfy.

**Definition 16** *Let a relation $\mathcal{R}$ be defined between $E$ and $E$ (we then say $\mathcal{R}$ is defined on $E$) . $\mathcal{R}$ is said:*

- ***reflexive*** *if* $\forall\ x \in E,\ x\mathcal{R}x$;
- ***symmetric*** *if* $\forall(x, y) \in E \times E,\ x\mathcal{R}y \Leftrightarrow y\mathcal{R}x$;
- ***antisymmetric*** *if* $\forall(x, y) \in E \times E, \{x\mathcal{R}y \text{ and } y\mathcal{R}x\} \Rightarrow y{=}x$;
- ***transitive*** *if* $\forall(x, y, z) \in E^3,\ \{x\mathcal{R}y \text{ and } y\mathcal{R}z\} \Rightarrow x\mathcal{R}z$.

We examine more closely ordering and equivalence relations hereafter because they are the most common relations appearing in financial and economic models. Each category of relations (ordering or equivalence) satisfies a subset of the properties stated in definition 16.

### 1.2.1 Orderings

**Definition 17** *a) A relation $\mathcal{R}$ on $E$ is a* ***pre-ordering*** *on $E$ if $\mathcal{R}$ is reflexive and transitive.*

*b) $\mathcal{R}$ is an* ***ordering*** *relation if $\mathcal{R}$ is reflexive, antisymmetric and transitive.*

*c) An ordering (or a pre-ordering) is said to be* ***complete*** *if for all couples of elements $(x, y)$ in $E \times E$ we have $x\mathcal{R}y$ or $y\mathcal{R}x$.*

**Example 18** *The most intuitive ordering relation is the inequality $\leq$ (or $\geq$) on the set $\mathbb{R}$ of real numbers. It is easy to check that it is indeed a complete ordering relation.*

In microeconomic models, preferences of consumers over bundles of goods are pre-ordering relations. Within utility theory, we associate to each set of goods a number (utility) measuring the consumer welfare obtained through consumption of this set of goods. In this case, it is easy to say if a given bundle is preferred to another one by comparing utilities with the relation $\leq$ . The preference relation is then a complete pre-ordering on the baskets of goods since any pair of baskets can be ranked. However, this preference relation is not an ordering as two different baskets can generate the same utility without being identical. This relation is not antisymmetric in general.

Below are two examples of orderings that are not complete.

**Example 19** *Even simple relations may not be complete; for example, consider the following relation on the set of couples of real numbers (denoted $\mathbb{R}^2$):*

$$(x_1, x_2)\ \mathcal{R}\ (y_1, y_2) \text{ if } \{x_1 \geq y_1 \text{ and } x_2 \geq y_2\}$$

*Two couples $x = (x_1, x_2)$ and $y = (y_1, y_2)$ satisfy $x\mathcal{R}y$ if and only if $x$ is located in the north-east of $y$ on a graphical representation where the first number is the coordinate on the horizontal axis and the second is the coordinate on the vertical axis. It is then easy to define couples that are not in relation such as the elements with coordinates $(3; 2)$ and $(1; 4)$. We have $3 \geq 1$ but $2 \leq 4$. The relation $\mathcal{R}$ is then a partial ordering as some couples cannot be ranked.*

**Example 20** *In standard portfolio choice theory of Markowitz*[5]*, investors select portfolios according to the trade-off between expected return and variance (or standard deviation). A portfolio is then characterized by a pair* $(\mu, \sigma)$*. Investors are looking for a high expectation* $\mu$ *and a low standard deviation* $\sigma$*. In this framework, we could define an ordering relation by*

$$(\mu_1, \sigma_1) \mathcal{R} (\mu_2, \sigma_2) \text{ if } \{\mu_1 \geq \mu_2 \text{ and } \sigma_1 \leq \sigma_2\} \tag{1.4}$$

*meaning that* $(\mu_1, \sigma_1)$ *dominates* $(\mu_2, \sigma_2)$ *because portfolio 1 has a higher expected return and a lower variance.*

### 1.2.2 Equivalence relations

**Definition 21** *A relation* $\mathcal{R}$ *is an* ***equivalence*** *relation if it is reflexive, symmetric and transitive.*

Equality is an equivalence relation on the set of real numbers. It is easy to check that for every triple $(x, y, z)$ of real numbers, we have:

$$\begin{aligned} x &= x \\ x &= y \Leftrightarrow y = x \\ \text{If } x &= y \text{ and } y = z \text{ then } x = z \end{aligned}$$

Consider now an economy with two goods. We say that an agent is indifferent between $(x_1, x_2)$ and $(y_1, y_2)$ if and only if $(x_1, x_2) \succeq (y_1, y_2)$ and $(y_1, y_2) \succeq (x_1, x_2)$. We usually write $(x_1, x_2) \sim (y_1, y_2)$, meaning that the agent is indifferent between the two pairs. The relation "$\sim$" is an equivalence relation. It is reflexive because $\succeq$ is reflexive, it is symmetric by definition, and transitive as $\succeq$ is itself transitive[6].

[5] Markowitz, H. 1952, Portfolio Selection, *The Journal of Finance*, 7(1), 77-91.

[6] In the companion book "Probability Theory" (Roger, 2010), you can understand why equivalence relations are important. In fact, when dealing with continuous random vari-

## 1.3 Mappings

### 1.3.1 Definitions and general properties

A binary relation $\mathcal{R}$ is characterized by three elements; the set $E$ (input set), the set $F$ (output set) and the subset $\mathcal{G}$ of $E \times F$ defined by:

$$\mathcal{G} = \{(x, y) \in E \times F \text{ so that } x\mathcal{R}y\}$$

$\mathcal{G}$ is called the **graph of relation** $\mathcal{R}$. Consequently, a relation $\mathcal{R}$ is completely defined by the triple $(E, F, \mathcal{G})$.

**Definition 22** *A **mapping** is a triple $(E, F, \mathcal{G})$ so that every element of $E$ is associated to one, and only one, element of $F$. A simple notation for a mapping is $f$ with input set $E$ and output set $F$ so that:*

$$x \in E \to f(x) \in F \tag{1.5}$$

$x \to f(x)$ means that $x$ and $f(x)$ are related, that is $x\mathcal{R}f(x)$. The graph of the mapping $f$ is then:

$$\mathcal{G} = \{(x, f(x)), x \in E\}$$

For a given mapping $f$, any $x \in E$ has only one **image** $f(x) \in F$. This does not mean that two different elements of $E$ cannot have the same image.

**Definition 23** *Let $f$ be a mapping from $E$ to $F$; the **range** of $f$, denoted $f(E)$ is the subset of $F$ defined by:*

$$f(E) = \{y \in F \text{ so that } \exists x \in E \text{ with } y = f(x)\}$$

ables characterized by a density, we see that a density can be changed on a finite number of values without changing any moment of the random variable. In other words, it is easier to identify under an equivalence relation all variables that are equal almost everywhere.

It turns out that for any mapping from $E$ to $F$, $f(E) \subset F$.

**Proposition 24** *Let $E_1$ and $E_2$ be two subsets of $E$ and $f$ be a mapping from $E$ to $F$; we then have:*

*1)* $E_1 \subset E_2 \Rightarrow f(E_1) \subset f(E_2)$

*2)* $f(E_1 \bigcup E_2) = f(E_1) \bigcup f(E_2)$

*3)* $f(E_1 \bigcap E_2) \subset f(E_1) \bigcap f(E_2)$

Although the notation $f(E_1)$ is explicit, one needs to keep in mind that $f$ associates elements of $F$ to elements of $E$ and not subsets of $F$ to subsets of $E$.

**Definition 25** *Let $F_1$ be a subset of $F$; we call* ***reciprocal range*** *of $F_1$ under $f$, and denote $f^{-1}(F_1)$, the subset of $E$ defined by:*

$$f^{-1}(F_1) = \{x \in E \text{ so that } f(x) \in F_1\}$$

$f^{-1}(F_1)$ is thus the set of elements of $E$ whose image is in $F_1$.

The reciprocal range has also interesting properties which are especially important in the definitions of random variables.

**Proposition 26** *Let $F_1$ and $F_2$ be two subsets of $F$. We have:*

*a)* $F_1 \subset F_2 \Rightarrow f^{-1}(F_1) \subset f^{-1}(F_2)$

*b)* $f^{-1}(F_1 \bigcup F_2) = f^{-1}(F_1) \bigcup f^{-1}(F_2)$

*c)* $f^{-1}(F_1 \bigcap F_2) = f^{-1}(F_1) \bigcap f^{-1}(F_2)$

*d)* $f\left[f^{-1}(F_1)\right] \subset F_1$

*e) If* $E_1 \subset E,\ E_1 \subset f^{-1}\left[f(E_1)\right]$

### 1.3.2 Injective, surjective, bijective mappings

**Definition 27** *A mapping $f$ from $E$ to $F$ is:*

-***injective*** *if :*

$$\forall (x, x') \in E \times E, x \neq x' \Rightarrow f(x) \neq f(x')$$

-***surjective*** *if :*

$$\forall y \in F, \exists x \in E \quad \text{such that} \quad y = f(x)$$

-***bijective*** *if $f$ is both injective and surjective. It is also called a* ***one-to-one*** *mapping.*

The definition means that injective mappings preserve separateness. Two different elements of $E$ always lead to two different elements of $F$.

A surjective mapping is characterized by $f(E) = F$. When a mapping is bijective, every element of $E$ has a unique image and every element of $F$ is the image of a unique element of $F$. It allows to define what is called the inverse mapping of $f$.

**Definition 28** *Let $f$ be a bijective mapping from $E$ to $F$; there exists an* ***inverse mapping*** *from $F$ to $E$, noted $f^{-1}$, so that for all elements $y$ of $F$, $f^{-1}(y) = x$ with $y = f(x)$.*

For example, the mapping $x \in \mathbb{R} \to f(x) = x^5$ is bijective. The inverse mapping $f^{-1}$ is defined by:

$$f^{-1}(y) = y^{\frac{1}{5}}$$

We then have:

$$x \in E \to f(x) \in F \to f^{-1}(f(x)) = x \in E$$

**Remark 29** *The reader can easily observe that mappings $x \to x^n$ are not bijective when $n$ is an even integer since $f(x) = f(-x)$. Figure 1.2 illustrates this remark.*

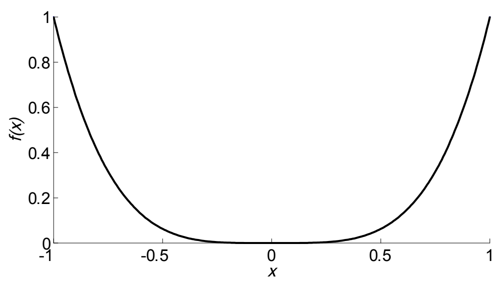


Figure 1.2: The mapping $x \to f(x) = x^4$

### 1.3.3 Compounding mappings

**Definition 30** *Let $E, F, G$ be three sets, $f$ be a mapping from $E$ to $F$ and $g$ be a mapping from $F$ to $G$ ; the triplet $(E, G, \mathcal{R})$ defines a mapping from $E$ to $G$ where the relation $\mathcal{R}$ is characterized by:*

$$x\mathcal{R}z \text{ if } z = g(y) \text{ with } y = f(x)$$

Denote $h$ the mapping composed as follows:

$$x \in E \to f(x) \in F \to g(f(x)) = h(x) \in G$$

The composition of two mappings is also written $g \circ f$. We can thus write:

$$g \circ f(x) = g\left(f(x)\right) = h(x)$$

**Definition 31** *Let $E$ be a set; the **identity mapping** of $E$ denoted $i_E$ is the mapping defined from $E$ to $E$ by:*

$$\forall x \in E, i_E(x) = x$$

When a mapping $f$ from $E$ to $F$ is bijective, we know that there exists an inverse mapping $f^{-1}$ from $F$ to $E$ so that $f^{-1}(f(x)) = x$, consequently $f^{-1} \circ f = i_E$. Of course, the mapping $f \circ f^{-1}$ is equal to $i_F$ as the input set is $F$ and the output set is $E$.

**Proposition 32** *Let $E, F, G$ be three sets, $f$ a bijective mapping from $E$ to $F$ and $g$ a bijective mapping from $F$ to $G$; the map $g \circ f$ is bijective and we have:*

$$\left(g \circ f\right)^{-1} = f^{-1} \circ g^{-1}$$

The intuition behind this result is straightforward; every element $z$ of $G$ has only one predecessor $y \in F$ under the mapping $g$. Also, $y$ has only one predecessor $x \in E$ under the mapping $f$. It follows that $z$ has only one predecessor under the composition $g \circ f$.

## 1.4 Topology of $\mathbb{R}$

"Open, closed, interior, frontier" are words of everyday life. These words have a precise mathematical meaning in a field of mathematics named topology. Here, we focus on the simple case of $\mathbb{R}$, the set of real numbers. We are first going to give mathematical definitions of the abovementioned words and of some others and we then justify the definition of the set $\mathbb{R}$.

The first subsection explains why it is necessary to build a set like $\mathbb{R}$, compared to "smaller" sets like $\mathbb{N}, \mathbb{Z}$ or $\mathbb{Q}$. Properties of sequences of real numbers appear in subsection 2. Sequences appear in many finance problems, the most simple example being discounting. When a financial security generates a sequence of future cash-flows, valuing this asset necessitates the use of a sequence of discount factors. More generally, standard equilibrium models are built around a tâtonnement process. Proposing a price generates demand and supply. When demand is greater than supply, a higher price is proposed. The process restarts for a second step and possibly many successive steps to reach the equilibrium where demand equals supply. The sequence of successive prices converges to the equilibrium price. It is the reason why we present the convergence properties of sequences of real numbers in this section.

### 1.4.1 The set of real numbers

In section 1.1 we saw the following inclusions:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$$

But the Pythagorean theorem shows that some numbers do not belong to $\mathbb{Q}$. Consider a triangle $ABC$ right at $A$ such that the lengths of $AB$ and $AC$ are equal to 1 (see figure 1.3). The question is: what is the length of $BC$? We know by the Pythagorean theorem the square of the length of $BC$ is 2. Consequently $\sqrt{2}$ denotes the length of $BC$.

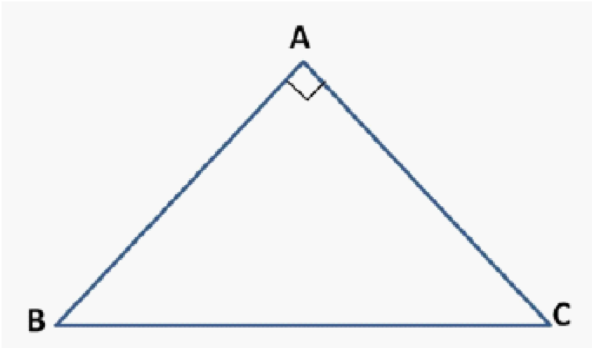


Figure 1.3: Right triangle at $A$ with length 1 for $AB$ and $AC$

Therefore it is easy to check if $\sqrt{2}$ is a rational number. If $\sqrt{2}$ was written as an irreducible fraction $n/m$, we should get

$$\frac{n^2}{m^2} = 2 \Leftrightarrow n^2 = 2m^2$$

This equality implies that $n^2$ is an even number. But, in this case, $n$ is also an even number. Consequently, $n^2$ can be divided by 4 and $m^2$ is then even. But if $n$ and $m$ are even numbers, there is a contradiction because the fraction $n/m$ is not irreducible. It shows that we need a larger set of numbers (compared to $\mathbb{Q}$) to measure some lengths. This set is the set of real numbers, denoted $\mathbb{R}$.

### Upper and lower bounds

Subsets of real numbers are characterized by properties requiring the introduction of the notions of bounds. They are also fundamental to study sequences of real numbers.

**Definition 33** *a) Let $E$ be a subset of $\mathbb{R}$. A real number $M \in \mathbb{R}$ is an* ***upper bound*** *of $E$ if:*

$$\forall x \in E, \quad x \leq M$$

*b) $m \in \mathbb{R}$ is a* ***lower bound*** *of $E$ if:*

$$\forall x \in E, \quad x \geq m$$

Upper or lower bounds do not always belong to $E$. For example if $E = \{x \in \mathbb{R} \text{ such that } x > 0 \text{ and } x < 1\}$, any upper (lower) bound of $E$ is greater or equal to 1 (lower or equal to 0) and does not belong to $E$.

It may happen that a set has no upper/lower bound. The set of real numbers, $\mathbb{R}$, is such a set. Moreover, when a subset $E$ of $\mathbb{R}$ has upper(lower) bounds, one of these is lower (greater) than the others. It is called the **supremum (infimum)** of $E$.

**Definition 34** *a) The **supremum** of a subset $E$ of $\mathbb{R}$ is the lowest upper bound of $E$. In the same way, the **infimum** of $E$ is the greatest lower bound of $E$.*

*b) A subset $E$ is **bounded** if it has a supremum and an infimum.*

0 is the infimum of $\mathbb{N}$ but this set has no supremum. $\mathbb{Q}$ has neither a supremum nor an infimum. These notions allow to characterize $\mathbb{R}$, the set of real numbers.

**Proposition 35** *Any nonempty subset $E$ of $\mathbb{R}$ that has an upper (lower) bound has a supremum (infimum) in $\mathbb{R}$.*

This proposition is the essential distinction between rational numbers and real numbers. Let $A = \left\{x \in \mathbb{Q} \text{ such that } x < \sqrt{5}\right\}$; $A$ has no supremum in $\mathbb{Q}$ since the lowest upper bound of $A$ is $\sqrt{5}$. In other words, for any rational upper bound of $A$, you can find a lower one (closer to $\sqrt{5}$).

**Remark 36** *The proposition has important consequences for empirical issues. If $E$ has a supremum in $\mathbb{R}$, say $b$, it is unique. Though $b$ is not necessarily in $E$, it means that you can always find in $E$ a number which is as close as you want of $b$. More precisely:*

$$\forall \varepsilon > 0, \exists x \in E \text{ such that } 0 < b - x < \varepsilon$$

### Intervals

The most usual subsets of $\mathbb{R}$ are intervals, defined as follows.

**Definition 37** *An interval is a subset $I$ of $\mathbb{R}$ defined by:*

$$\forall (x, y) \in I \times I, \forall \alpha \in [0; 1], \alpha x + (1 - \alpha) y \in I$$

The interpretation of this definition is as follows. As soon as $a$ and $b$ belong to $I$, every number between $a$ and $b$ is also an element of $I$. Geometrically speaking[7], when $\alpha$ goes from 0 to 1, $\alpha x + (1-\alpha)y$ goes from $y$ to $x$. In a more general framework, we will see later on that this property characterizes convex sets.

Intervals can be open, closed or semi-open(closed), depending on the fact that the ends of the interval belong or not to the it.

Table 1.1 gives some examples of each category of intervals.

| **Type** | **Example** |
|---|---|
| open | $]0;10[ = \{x \in \mathbb{R} \ / \ 0 < x < 10\}$ |
| semi-open(closed) | $[-1;3[ = \{x \in \mathbb{R} \ / \ -1 \leq x < 3\}$ |
| closed | $[3;5] = \{x \in \mathbb{R} \ / \ 3 \leq x \leq 5\}$ |

Table 1.1: Examples of intervals

It may happen that one of the two ends of an interval is infinite. For example, the interval $[y;+\infty[$ denotes all numbers greater or equal to $y$. It is a semi-closed interval but remember that $+\infty$ does not belong to $\mathbb{R}$.

### Open and closed sets

When dealing with optimization problems (issues addressed in Part II), we will see that knowing if the function to be optimized is defined on an open or on a closed set is important.

**Definition 38** *Let $E$ be a subset of $\mathbb{R}$. $y \in E$ is **interior** to $E$ if there exists $\varepsilon > 0$ such that $]y-\varepsilon; y+\varepsilon[ \subset E$. $E$ is then called a **neighborhood** of $y$.*

The interval $]y-\varepsilon; y+\varepsilon[$ can also be written

$$\{x \in \mathbb{R} \text{ such that } |x-y| < \varepsilon\}$$

[7]When $\alpha \in [0;1]$, the combination $\alpha x + (1-\alpha)y$ is called a convex combination. When there are no constraints on $\alpha$, $\alpha x + (1-\alpha)y$ is called an affine combination.

**Example 39** *Consider the set* $E = [0; 5]$*; any* $y \in ]0; 5[$ *is interior to* $E$*. To see why, define* $\varepsilon = \frac{1}{3} \min (y; 5 - y)$ *; it is easy to verify that* $]y - \varepsilon; y + \varepsilon[ \subset E$*. On the opposite* $0$ *and* $5$ *are not in the interior of* $E$ *because any open interval centered on one of these numbers is not included in* $E$*.*

We can now provide the general definition of open and closed sets in $\mathbb{R}$.

**Definition 40** *a) A subset* $E$ *of* $\mathbb{R}$ *is* ***open*** *if any element of* $E$ *is interior to* $E$*.*

*b) A subset* $F$ *of* $\mathbb{R}$ *is* ***closed*** *if its complement* $F^c$ *(in* $\mathbb{R}$*) is open.*

Remark that open (closed) intervals are open (closed) subsets[8] of $\mathbb{R}$. The interval $[3; +\infty[$ is closed according to the above definition because the complement $]-\infty; 3[$ is open; more generally, intervals like $]-\infty; a]$ or $[b; +\infty[$ are closed sets. It follows that the (complete) set $\mathbb{R}$ is simultaneously closed and open. But the complement $\mathbb{R}^c$ is equal to $\varnothing$. Consequently the empty set is also closed and open.

**Definition 41** *a) The* ***interior*** *of a subset* $E$ *of* $\mathbb{R}$*, denoted* $\overset{\circ}{E}$*, is the set of elements in* $E$ *which are interior to* $E$*; it is also the largest open subset included in* $E$*.*

*b) The* ***closure*** *of a subset* $E$ *of* $\mathbb{R}$ *denoted* $\overline{E}$ *is the smallest closed set containing* $E$*.*

*c) The* ***exterior*** *of a subset* $E$ *of* $\mathbb{R}$ *is the interior of the complement* $E^c$*.*

*d) The* ***frontier*** *of a subset* $E$ *of* $\mathbb{R}$ *is the set of elements which are neither interior nor exterior to* $E$*. It is also called the* ***boundary*** *of* $E$

The interior of an interval $[a; b[$ is the open interval $]a; b[$ . This result is quite intuitive; but look at what is going on when $E = \mathbb{Q}$. It turns out that the interior of $\mathbb{Q}$ in $\mathbb{R}$ is the empty set. For any rational number $x$ (written

[8]For more general sets, for example sets of functions, openness and closedness are not so intuitive, contrary to what we note here.

$\frac{m}{n}$ where $m$ and $n$ are integers) and any $\varepsilon > 0$, the interval $\left]\frac{m}{n} - \varepsilon; \frac{m}{n} + \varepsilon\right[$ contains irrational numbers which are not elements of $E = \mathbb{Q}$. Consequently $\overset{\circ}{E} = \varnothing$. We also deduce that the closure of $\mathbb{Q}$ is the set of all real numbers $\mathbb{R}$. It explains why calculators or spreadsheets do not make large errors though they use only rational numbers.

**Proposition 42** *A subset $E$ of $\mathbb{R}$ is* ***bounded*** *if there exists an interval $]a; b[$ such that $E$ is included in $]a; b[$.*

In other words, a set $E$ is bounded when it has a lower bound and an upper bound. However it is not mandatory that the bounds belong to $E$. An open interval $]a; b[$ is an example of such a bounded set.

**Definition 43** *A subset $C$ of $\mathbb{R}$ is* ***compact*** *if it is at the same time closed and bounded.*

**Remark 44** *This definition is true for sets in $\mathbb{R}$ or $\mathbb{R}^n$ (the set of $n$-tuples of real numbers $(x_1, ..., x_n)$). In more general frameworks, this definition has to be changed (this issue is addressed in part II). Compact sets are important to establish some properties of continuous functions, described later on in chapter 2.*

## 1.4.2 Sequences of real numbers

**Example 45** *Assume you invest \$100 in a savings account on January, 1, next year, at a 4% interest rate. Interests are paid at the end of the year and are annually compounded (you get $x_1 = \$100 \times 1.04 = \$104$ at the end of the first year). As interests are compounded, you will get $x_2 = \$104 \times 1.04 = 100 \times 1.04^2$ at the end of the second year. More generally, your investment is worth $x_n = 100 \times 1.04^n$ at the end of year $n$.*

*$x_n$ can also be written $x_n = x_{n-1} \times 1.04$ and the ratio $\frac{x_{n+1}}{x_n}$ is constant over time, equal to 1.04. We will see later on that this characterizes a geometric sequence.*

*Assume now that the banker says that he sells a product paying a 4 % yearly interest rate. The meaning of this sentence depends on the frequency at which interests are compounded. With a monthly frequency, interest received after one month produce interests in the future. In this case, an investment of $100 at the beginning of the year is worth*

$$100 \times \left(1 + \frac{0.04}{12}\right)^{12} = 104.07$$

*at the end of the first year.*

*More generally, when interests are paid* $n$ *times a year, the final amount is:*

$$100 \times \left(1 + \frac{0.04}{n}\right)^{n}$$

*The question is to know what is the amount obtained when interests are continuously compounded, that is when* $n$ *tends to infinity. The notion of limit of a sequence of real numbers, and more generally the concept of convergence of sequences, is necessary to answer this question.*

### Convergence of sequences

Before defining precisely limits and convergence, we characterize what we call a sequence of real numbers.

**Definition 46** *A **sequence** of real numbers is a mapping* $f$ *from* $\mathbb{N}$ *to* $\mathbb{R}$ *denoted* $n \rightarrow f(n)$. *A simplified notation is* $(x_n, n \in \mathbb{N})$ *where* $f(n) = x_n$.

Consider the following examples:

1. $(x_n, n \in \mathbb{N}) = (2n, n \in \mathbb{N})$
2. $(x_n, n \in \mathbb{N}) = \left(\frac{1}{n^3}, n \in \mathbb{N}\right)$
3. $(x_n, n \in \mathbb{N}) = ((-1)^n, n \in \mathbb{N})$

4. $(x_n, n \in \mathbb{N}) = \left(\left(\frac{-1}{n}\right)^n, n \in \mathbb{N}\right)$

5. $\begin{cases} x_0 = 1 \\ x_n = \frac{x_{n-1}+1}{3}, n \in \mathbb{N}^* \end{cases}$

6. $\begin{cases} x_0 = 0; x_1 = 1 \\ x_n = x_{n-1} + x_{n-2}, n \geq 2 \end{cases}$

The two last sequences are different from the first four ones because they are not defined directly as functions of $n$. They are defined by induction, that is the $n$-th element is defined as a function of the preceding elements.

These sequences behave differently when $n$ increases. The first one increases without limits. The second one has positive elements that decrease to 0 when $n$ increases. The elements of the third one alternate in sign and are always equal to 1 or $-1$. Elements of the fourth one also alternate in sign but their absolute value decreases to 0. Finally, it is difficult to know the behavior of the fifth one without calculations.

The last sequence is the so-called "Fibonacci sequence"; it was described by Leonard from Pisa in the *Liber abaci* written in 1202[9]. This sequence increases without bounds but it is famous because the ratio of two successive terms $x_n/x_{n-1}$ tends to the gold number (approximately 1.618) when $n$ tends to infinity[10].

**Definition 47** *a) A sequence* $(x_n, n \in \mathbb{N})$ ***converges*** *to a limit* $a \in \mathbb{R}$ *if:*

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \;/\; \forall n > N,\; |x_n - a| < \varepsilon$$

*We then write* $\lim_{n \to +\infty} x_n = a$.

[9] Leonardo Fibonacci (trad. Laurence E. Sigler), Fibonacci's Liber abaci : A translation into modern English of Leonardo Pisano's Book of calculation, Springer-Verlag, 2002

[10] Today, the Fibonacci sequence is used in finance in some methods of technical analysis (see for example: Deron Wagner (2012), Advanced Technical Analysis of ETFs: Strategies and Market Psychology for Serious Traders, chapter 4, John Wiley & Sons Inc.

*b) A **Cauchy** sequence is a sequence* $(x_n, n \in \mathbb{N})$ *such that* $|x_i - x_j|$ *tends to 0 when* $i$ *and* $j$ *tend to* $+\infty$.

The first part of this definition shows that convergence does not change if we delete a finite number of elements in the sequence. This remark comes from the fact that $N$ is not specified in the definition (it may be 10 or 10,000,000 or any number). The only important fact is that $N$ exists.

**Proposition 48** *1) Every Cauchy sequence of real numbers converges to a limit.*

*2) If a sequence* $(x_n, n \in \mathbb{N})$ *converges to a limit, this limit is unique.*

In the 6 abovementioned examples, it is easy to see that sequences 1 and 3 do not converge. The first one increases without bounds (we say it tends to $+\infty$) and the second one takes alternatively values 1 and $-1$. On the opposite, sequence 2 converges to 0. Using definition 47 we observe that being given $\varepsilon$, it is sufficient to choose $N$ as the smallest integer greater than $\frac{1}{\sqrt[3]{\varepsilon}}$. Sequence 4 is specific because it takes alternatively positive and negative values but eventually converges to 0.

Sequence 5 is defined by induction so studying convergence may be more complicated. However, in this case, we know that $x_0 = 1$, and $x_n < x_{n-1}$. It is then strictly decreasing and always positive; the intuition says that it should converge. The next proposition shows it is the case by giving a clear convergence criterion.

**Proposition 49** *Let $(x_n, n \in \mathbb{N})$ a sequence of real numbers :*

*- If there exist $\alpha \in ]0;1[$ and $N \in \mathbb{N}$ such that $\left|\frac{x_n}{x_{n-1}}\right| < \alpha$ for $n > N$, the sequence $(x_n, n \in \mathbb{N})$ converges to 0.*

*- If there exist $\alpha > 1$ and $N \in \mathbb{N}$ such that $\left|\frac{x_n}{x_{n-1}}\right| > \alpha$ for $n > N$, $|x_n|$ tends to $+\infty$ when $n$ tends to $+\infty$ .*

We can also complete this proposition by the following result: any convergent sequence of real numbers is included in a closed interval. Moreover, remember that we defined compact sets as closed and bounded sets. In general spaces, a compact set is a set in which any sequence has a convergent subsequence. It is a kind of reciprocal of the above result. We cannot say that any sequence in a closed interval is convergent but one can find a convergent subsequence because closed intervals are compact.

### Some specific and useful sequences

Two special cases are very useful when dealing with practical problems, namely arithmetic and geometric sequences. These two types of sequences have become famous because they are essential elements of the theory of

Reverend Thomas Robert Malthus (1798). He assumed that available food increases in an arithmetical ratio, but simultaneously, the population doubles every twenty-five years, that is according to a geometrical progression. Consequently, if population growth is not limited, a food shortage should come in a few decades!

**Definition 50** *a) An **arithmetic** sequence with ratio* $a \in \mathbb{R}$ *is a sequence* $(x_n, n \in \mathbb{N})$ *such that for* $n > 0$*, we have :*

$$x_n = x_{n-1} + a$$

*where the first term* $x_0$ *may be any real number.*

*b) A **geometric** sequence with ratio* $c \in \mathbb{R}$ *is a sequence* $(x_n, n \in \mathbb{N})$ *such that for* $n > 0$*, we have :*

$$x_n = cx_{n-1}$$

*where the first term* $x_0$ *may be any real number.*

For these specific sequences convergence criteria are quite intuitive. They are summarized in the following proposition.

**Proposition 51** *a) Any arithmetic sequence with a non zero ratio* $a$ *does not converge. It increases (decreases) without limits if* $a > 0$ $(a < 0)$*.*

*b) A geometric sequence converges to 0 if the absolute value of the ratio c is strictly lower than 1, otherwise it does not converge. It increases (decreases) without bounds if* $x_0 > 0$ *and* $c > 1$ $(x_0 < 0$ *and* $c > 1)$*. When* $c \leq -1$*, the successive elements alternate in sign and the sequence does not converge.*

**Proposition 52** *a) Let* $(x_n, n \in \mathbb{N})$ *denote a geometric sequence with ratio* $c$*, different from 0 and 1; we have:*

$$\sum_{n=0}^{N} x_n = x_0 \frac{1 - c^{N+1}}{1 - c}$$

*b) If* $|c| < 1$*, the limit is given by:*

$$\lim_{N\to+\infty} \sum_{n=0}^{N} x_n = \frac{x_0}{1-c} \tag{1.6}$$

**Proof.** The sum of the first $N$ terms is:

$$\begin{aligned} \sum_{n=0}^{N} x_n &= x_0 + cx_0 + .... + c^N x_0 \\ &= x_0 \left(1 + c + .... + c^N\right) \end{aligned}$$

We know that:

$$\left(1 + c + .... + c^N\right)(1-c) = 1 - c^{N+1}$$

it then follows:

$$1 + c + .... + c^N = \frac{1 - c^{N+1}}{1-c}$$

The result of the proposition follows immediately. ■

In real data, it is common to deal with monotonic sequences, defined as follows.

**Definition 53** *A sequence* $(x_n, n \in \mathbb{N})$ *is* ***increasing*** *if for any* $n > 0, x_n \geq x_{n-1}$*. It is* ***decreasing*** *if the reverse inequality is true.*

Convergence criteria of monotonic sequences are simpler than the general criteria.

**Proposition 54** *a) An increasing sequence* $(x_n, n \in \mathbb{N})$ *which is bounded above is convergent. The limit a is the lowest upper bound of the set of values* $\{x_n, n \in \mathbb{N}\}$*.*

*b) A decreasing sequence* $(x_n, n \in \mathbb{N})$ *which is bounded below is convergent. The limit a is the largest lower bound of the set of values* $\{x_n, n \in \mathbb{N}\}$*.*

The following example 55 uses the properties of geometric sequences and proposition 54 as well.

**Example 55** ***Consol bonds***

*Consol bonds are bonds first issued by the British government during the 19th century. They are now known as perpetual bonds because they have no maturity date (in other words they are never reimbursed). The holder of a consol receives a given amount of interest every year for ever.*

*Assume you can invest your money at a constant yearly rate* $r$. *To get 1 dollar in one year you need to invest* $1/(1+r)$ *dollars today. To get one dollar in two years you need to invest* $1/(1+r)^2$ *today, and so on.*

*More generally investing* $x_n = \frac{1}{(1+r)^n}$ *today at a yearly rate* $r$ *provides one dollar in* $n$ *years. The sequence* $(x_n, n \in \mathbb{N})$ *is obviously decreasing and positive, that is with a lower bound equal to 0.*

*Proposition 54 implies it converges and it is easy to see that the limit is 0. Consider now the sequence* $y_n$ *defined by:*

$$\begin{aligned} y_0 &= x_0 = 1 \\ y_n &= y_{n-1} + x_n \end{aligned}$$

$y_n$ *is in fact the amount to invest today to get 1 dollar every year up to year* $n$. *The sequence* $(y_n, n \in \mathbb{N})$ *is increasing because* $y_n = y_{n-1} + x_n$ *and* $x_n$ *is positive. The limit of* $y_n$ *is obtained if we remark that :*

$$y_n = \sum_{k=0}^{n} \left(\frac{1}{1+r}\right)^k$$

*In other words it is the sum of the* $n+1$ *first terms of a geometric sequence with ratio* $c = \frac{1}{1+r} < 1$. $y_n$ *can be rewritten as:*

$$y_n = \sum_{k=0}^{n} \left(\frac{1}{1+r}\right)^k = \frac{1 - \left(\frac{1}{1+r}\right)^{n+1}}{1 - \frac{1}{1+r}}$$

*using the relationship* $1-d^{n+1}=(1-d)\sum_{k=0}^{n}d^k$ *for any number* $d$. *We then get:*

$$\lim_{n\to+\infty}\frac{1-\left(\frac{1}{1+r}\right)^{n+1}}{1-\frac{1}{1+r}}=\frac{1}{1-\frac{1}{1+r}}=\frac{1+r}{r}$$

*In fact* $\lim_{n\to+\infty}y_n$ *is the necessary amount to be invested today to receive one dollar every year for ever! It is then the current price of a consol paying one dollar of interest every year. In most cases the bond is paid now but the first interest is received in one year. It means that the price of the consol will be* $\frac{1+r}{r}-1=\frac{1}{r}$.

**Operations on sequences**

If it is natural to add, subtract or multiply numbers, we may wonder if simple arithmetics has unexpected consequences on limits of sequences. For example, if two sequences $u_n$ and $v_n$ converge respectively to $a$ and $b$, is it true that the product $u_nv_n$ converges to the product $ab$. The proposition hereafter answers this kind of question.

**Proposition 56** *Denote* $(u_n, n \in \mathbb{N})$ *and* $(v_n, n \in \mathbb{N})$ *two convergent sequences with limits* $a$ *and* $b$*. We then have:*

*1)* $lim_{n\to+\infty}(u_n + v_n) = a + b$

*2)* $lim_{n\to+\infty}(u_n v_n) = ab$

*3)* $\forall(x, y) \in \mathbb{R} \times \mathbb{R}, lim_{n\to+\infty}(xu_n + yv_n) \ = ax + by$

*4)* $lim_{n\to+\infty} |u_n| = |a|$

*5) If* $b \neq 0,\ \ lim_{n\to+\infty} \frac{u_n}{v_n} = \frac{a}{b}$

Of course, the proposition is valid only if the two sequences $(u_n, n \in \mathbb{N})$ and $(v_n, n \in \mathbb{N})$ converge. In fact, suppose that one of the two is not convergent (increasing or decreasing without limits for example). In this case, some of the above questions cannot be answered. For example, in part (2) of the proposition, let us assume $u_n \to 0$ and $v_n \to +\infty$. Anything can happen; if $v_n = n$ and $u_n = \frac{1}{n}$, the product is always equal to 1 (and then convergent). If $u_n = \frac{1}{n^2}$, the product converges to 0 but, if $u_n = \frac{1}{\sqrt{n}}$, the product is $\sqrt{n}$ and does not converge. Table 1.2 summarizes the different cases. More generally, prudence is necessary when applying proposition 56.

| $v_n$ | $u_n$ | $u_n v_n$ | $\lim_{n\to+\infty} u_n v_n$ |
|---|---|---|---|
| $n$ | $1/n$ | 1 | 1 |
| $n$ | $1/n^2$ | $1/n$ | 0 |
| $n$ | $1/\sqrt{n}$ | $\sqrt{n}$ | $+\infty$ |

Table 1.2: Undeterminate cases

### Adjacent sequences

A standard method to evaluate an unknown quantity $b$ is to define sequences $x_n$ and $y_n$ such that :

$$x_n \leq b \leq y_n$$

If $(x_n, n \in \mathbb{N})$ is increasing and $(y_n, n \in \mathbb{N})$ decreasing, proposition 54 allows to conclude that these two sequences converge. In addition, if the difference $x_n - y_n$ tends to 0, we are sure that the common limit is the number

$b$ we were trying to evaluate. This approach will be used to define Riemann integrals in chapter 3.

**Definition 57** ***Adjacent*** *sequences are sequences* $(x_n, n \in \mathbb{N})$ *and* $(y_n, n \in \mathbb{N})$ *such that the first (second) one is increasing (decreasing), and* $\lim_{n\to+\infty}(y_n - x_n) = 0$.

We deduce from this definition that $x_n \leq y_n$ for any $n$ and get the following result.

**Proposition 58** *Two adjacent sequences* $(x_n, n \in \mathbb{N})$ *and* $(y_n, n \in \mathbb{N})$ *converge to the same limit.*

In fact, there is a contradiction between $\lim_{n\to+\infty}(y_n - x_n) = 0$ and the assumption that the limits can be different.

# Chapter 2

# Functions of one variable

In finance and, more generally, in economics, the word "function" is very common. One naturally speaks about profit functions, cost functions; portfolio managers usually recall that the return on a portfolio is a function of the risk it bears. Most often, a given quantity depends on several others. For example, the return depends on a given measure of risk but also depends on the horizon of the investor, on market liquidity, and possibly on many other variables. However, it is often useful to start with a simple analysis where a quantity under consideration is assumed to depend on only one variable. We call these relationships *functions of one variable*.

In this chapter we start by definitions and elementary properties of functions. Section 2.2 is devoted to limits, continuity and the intermediate value theorem. In section 2.3, we define and characterize derivatives before developing their main properties. The next section is devoted to logarithms and exponential functions. These functions are possibly the most important in economics and finance. Logarithms are concave and therefore widely used as models of utility functions. They are also used to calculate continuous returns on financial markets. Exponential functions are convex and are the natural tool to discount future amounts of money in continuous-time models.

Finally, the last section of the chapter is devoted to Taylor expansions allowing to approximate any sufficiently regular function by a polynomial. These approximations are also a fundamental tool of economic models. For example, the Arrow-Pratt coefficient of risk aversion is obtained through a reasoning using Taylor expansions.

## 2.1 Definitions and notations

**Definition 59** *A **function** of a real variable $x$ is any map $f$ defined on a set $D$ (input set) included in $\mathbb{R}$ and taking values in $\mathbb{R}$(output set). We write*:

$$x \in D \to f(x) = y \in \mathbb{R}$$

*or*

$$x \to f(x) = y$$

The input set is called **the domain** of $f$ and the set of possible outputs $f(x)$ is the **range** of $f$. In relation to statistical studies and to econometrics, $x$ is named the **independent** variable and $y = f(x)$ the **dependent** variable.

**Example 60** *Linear and affine functions*

$$D = \mathbb{R} \;;\; x \to f(x) = ax + b \text{ where } a, b \in \mathbb{R}$$

*When $b = 0$, $f$ is a **linear function.** When $b \neq 0$, $f$ is an **affine function.** When the relationship between two economic variables is affine, it is easy to calculate the sensitivity of $y = f(x)$ to variations in the independent variable $x$ because this sensitivity is exactly equal to $a$. More precisely, considering two numbers $x_1$ and $x_2$, we write the sensitivity as*

$$\frac{f(x_1) - f(x_2)}{x_1 - x_2} = \frac{ax_1 + b - ax_2 - b}{x_1 - x_2} = a \tag{2.1}$$

**Definition 61** *1) Let $f$ and $g$ be two functions defined on the same domain $D$; their sum $f+g$ is defined as follows:*

$$\forall x \in D, (f+g)(x) = f(x) + g(x)$$

*2) The product of a function $f$ by a number $\alpha \in \mathbb{R}$, written $\alpha f$, is defined by:*

$$\forall x \in D, (\alpha f)(x) = \alpha f(x)$$

*3) The product $fg$ of two functions $f$ and $g$, defined on the same domain $D$, is defined by:*

$$\forall x \in D, (fg)(x) = f(x)g(x)$$

*4) The inequality $f \leq g$ means:*

$$\forall x \in D, \ f(x) \leq g(x)$$

In further chapters devoted to linear algebra, we will see that parts (1) and (2) in the above definition are essential properties of vectors, meaning that some sets of functions are in fact vector spaces. The inequality defined in the last part of the definition is frequently used in microeconomics when $f$ and $g$ are cumulative distribution functions of payoffs of financial securities. Such an inequality defines first-order stochastic dominance[1].

**Definition 62** *Let us assume that the domain $D$ of a function $f$ is such that if $x \in D$ then $-x \in D$.*

*- $f$ is said* ***even*** *if for all $x \in D$, $f(-x) = f(x)$*

*- $f$ is said* ***odd*** *if $f(-x) = -f(x)$.*

**Remark 63** *The graph of an even function is symmetric with respect to the y-axis. The graph of an odd function is symmetric with respect to the origin (0;0).*

[1] see Roger, Probability for Finance, 2010a, chapter 1.

Figures 2.1 and 2.2 illustrate this remark with the even function $f(x) = x^2$ and the odd function $g(x) = x^3$. More generally, any function $f(x) = x^n$ is even if $n$ is an even integer and odd if $n$ is odd.

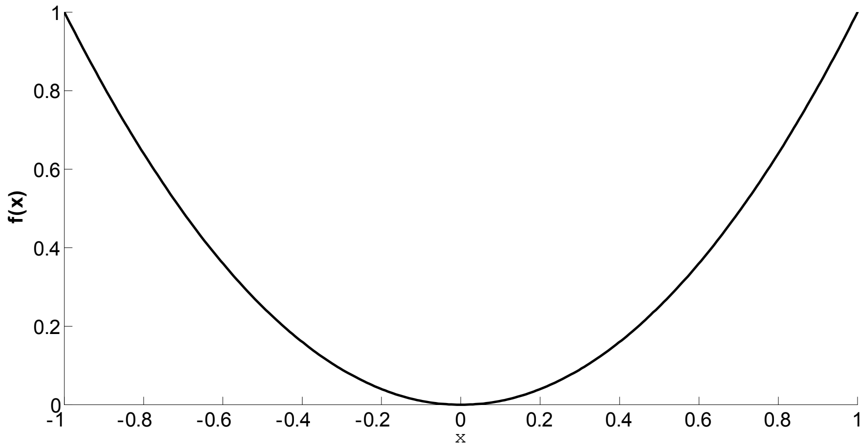


Figure 2.1: The function $f(x) = x^2$

### 2.1.1 Increasing and decreasing functions

**Definition 64** *Let $f$ be a function defined on $D \subset \mathbb{R}$ and $B$ a subset of $D$. $f$ is* ***increasing (decreasing)*** *on $B$ if:*

$$\forall (x_1, x_2) \in B \times B, x_1 \geq x_2 \Rightarrow f(x_1) \geq (\leq)\ f(x_2)$$

*If the inequalities are strict then $f$ is said strictly increasing (decreasing) on $B$.*

When the set $B$ is the entire domain $D$, we simply say that $f$ is increasing (decreasing). A function is said to be monotonic if it varies in only one direction , that is, if it is solely increasing or decreasing on the entire domain $D$.

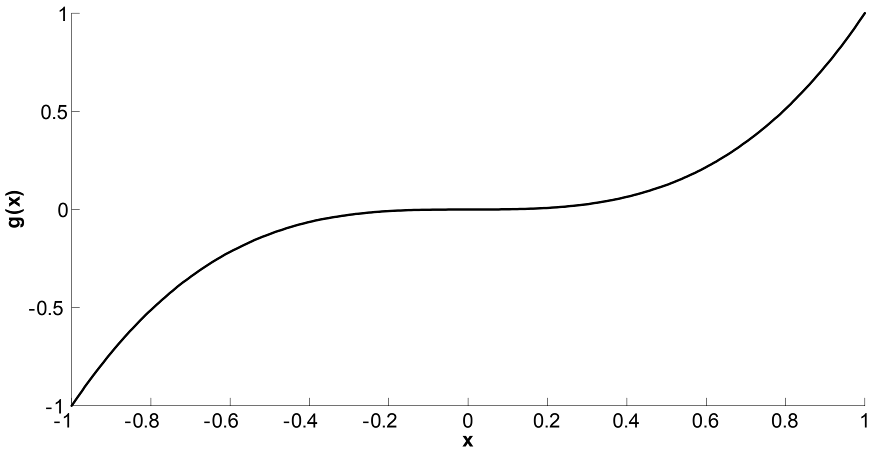


Figure 2.2: The function $g(x) = x^3$

**Example 65** *a) A standard hypothesis in economic models is to assume that the demand* $d$ *for a given good is a decreasing function of the price* $p$ *of the good. The demand function is written* $p \to d(p)$ *and we have:*

$$p_1 \leq p_2 \Rightarrow d(p_1) \geq d(p_2)$$

*b) When you buy a given quantity* $b$ *of stocks, your global cost is an affine function like* $c(x) = bx + a$ *(where* $b$ *is the number of stocks and* $a$ *the fixed cost per trade).* $x$ *is the unit stock price including the proportional transaction cost.* $c(x)$ *is increasing in* $x$ *because* $b$ *is positive. One can also describe this global cost as a function of the number of stocks in the portfolio. In this case, we can write* $c^*(y) = dy + a$ *where* $d$ *is the unit price,* $y$ *is the number of stocks and* $a$ *is still the fixed cost per trade.*

**Proposition 66** *Let* $f$ *and* $g$ *be two functions defined on* $D$ *:*

*1) If* $f$ *and* $g$ *are increasing (decreasing),* $f + g$ *is increasing (decreasing).*

*2) If* $f$ *and* $g$ *are positive and increasing on* $D$*, the product* $fg$ *is an increasing function.*

*3) If $f$ and $g$ are both increasing or both decreasing, the compound function $g \circ f$ is increasing. If one is increasing while the other is decreasing, then $g \circ f$ is decreasing.*

Of course, if $\alpha$ and $\beta$ are two positive numbers and if $f$ and $g$ are increasing (decreasing), the combination $\alpha f + \beta g$ is increasing (decreasing).

### 2.1.2 Extremum of a function

Optimization, that is the search of extrema of functions, is a rather important tool in finance and, more generally, in economics. In microeconomics, agents are supposed to maximize their expected utility. In portfolio choice theory, investors maximize the expected return of their portfolio under some constraints linked to the quantity of risk they are ready to bear. Similarly, investors may want to minimize the variance of their portfolio return, being given a level of expected return. These examples show why extrema (minima and maxima) of functions are so important in financial models.

### Upper and lower bounds

**Definition 67** *Let $f$ be a function defined on a domain $D \subset \mathbb{R}$:*

*- A number $M \in \mathbb{R}$ is an **upper bound** of $f$ on $D$ if for all $x \in D$, $f(x) \leq M$. If such a number exists, $f$ is **bounded above**.*

*- A number $m \in \mathbb{R}$ is a **lower bound** of $f$ on $D$ if for all $x \in D$, $f(x) \geq m$. If such a number exists, $f$ is **bounded below**.*

*- A function $f$ is **bounded** if there are both an upper and a lower bound..*

**Proposition 68** *Let $f$ and $g$ be two functions defined on a domain $D$. If they are both bounded above (below), the sum $f + g$ is also bounded above (below). Moreover, if $f$ and $g$ are positive and bounded above, the product $fg$ is bounded above.*

### Local extremum and global extremum

**Definition 69** *Let $f$ be a function defined on a domain $D$ :*

*1) $x_0$ is called a **local minimum** of $f$ if:*

$$\exists \alpha > 0, \forall x \in ]x_0 - \alpha; x_0 + \alpha[\,, \quad f(x) \geq f(x_0)$$

*2) $x_0$ is called a **local maximum** of $f$ if:*

$$\exists \alpha > 0, \forall x \in ]x_0 - \alpha; x_0 + \alpha[\,, \quad f(x) \leq f(x_0)$$

This notion of minimum (maximum) is said to be local because $f(x_0)$ is lower (higher) than $f(x)$ only for $x$ in an interval containing $x_0$. As we used "$\exists \alpha > 0$", it means that the range of values over which $f(x) \geq f(x_0)$ may be really narrow.

**Definition 70** *Let $f$ be a function defined on $D$ :*

*1) $x_0$ is called a **global minimum** of $f$ if:*

$$\forall x \in D, \quad f(x) \geq f(x_0)$$

*2) $x_0$ is called a **global maximum** of $f$ if:*

$$\forall x \in D, \quad f(x) \leq f(x_0)$$

Here, the extremum is global because we used "$\forall x \in D$"; for a global minimum (maximum) $x_0$, the function cannot take lower (higher) values than $f(x_0)$ all over the domain $D$. In financial problems, finding a local extremum is not always satisfying because you can be far from the global extremum you were looking for. However, for general functions, there is no simple way (and sometimes no way at all) to find a global extremum. Reasonable conditions are obtained when functions exhibit "nice" properties like convexity or concavity. These properties will be presented later on in this chapter.

## 2.2 Limits and continuity

### 2.2.1 Pointwise limits

**Definition 71** *Let $f$ be a function defined on a domain $D$. The function $f$ has a **limit** $b \in \mathbb{R}$ at $a \in D$ if:*

$$\forall \varepsilon > 0, \exists \lambda > 0 \text{ such that } |x - a| < \lambda \implies |f(x) - b| < \varepsilon$$

*The limit is usually denoted:*

$$\lim_{x \to a} f(x) = b$$

By extension, it is possible, in some cases, to compute the limit of $f(x)$ when $x$ tends to $a$, even if $a$ is not in the domain but on the boundary of the domain. In this case we write:

$$\lim_{x \to a, x \neq a} f(x) = b$$

Another way to define the above limit is to say that for any sequence ($x_n$, $n \in \mathbb{N}$) converging to $a$, the sequence ($f(x_n)$, $n \in \mathbb{N}$) converges to $b$.

**Example 72** *Let us consider the function $f$ defined by:*

$$f(x) = \frac{x-9}{\sqrt{x}-3}$$

*The domain of $f$ is reduced to $\mathbb{R}_+ \setminus \{9\}$ because the denominator is 0 when $x = 9$ and the number under the square root must be positive. However, it is still possible to calculate $\lim_{x \to 9, x \neq 9} f(x)$ because the numerator of $f$ can be written as:*

$$x - 9 = (\sqrt{x} - 3)(\sqrt{x} + 3)$$

*Therefore $f(x)$ can be simplified because $\sqrt{x} - 3$ appears in the denominator and in the numerator of $f$. It turns out that $\lim_{x \to 9, x \neq 9} f(x) = \lim_{x \to 9}(\sqrt{x} + 3) = 6$.*

**Proposition 73** *If a function $f$ has a limit $b$ at $a \in D^{\circ}$ (the interior of $D$), this limit is unique.*

This proposition is easily understandable by assuming that there are two different limits $b$ and $b'$ in definition 71. We immediately get a contradiction.

### 2.2.2 One-sided limits

**Definition 74** *Let $f$ be a function defined on $D$ and $a \in D^{\circ}$. $f$ has a **right-limit**, $b \in \mathbb{R}$, at $a \in D^{\circ}$ if:*

$$\forall \varepsilon > 0, \exists \lambda > 0 \text{ such that } 0 \leq x - a < \lambda \implies |f(x) - b| < \varepsilon$$

*This limit is denoted:*

$$\lim_{x \to a^+} f(x) = b$$

*In the same way, a* ***left-limit*** $b \in \mathbb{R}$*, at* $a \in D^\circ$ *is defined by:*

$$\forall \varepsilon > 0, \exists \lambda > 0, \text{ we have } 0 \leq a - x < \lambda \implies |f(x) - b| < \varepsilon$$

*This limit is denoted:*

$$\lim_{x \to a^-} f(x) = b$$

We also aim at defining the limit of a function $f$ when $x$ tends to $\mp\infty$. It can only be a one-sided limit (left-limit at $+\infty$ and right-limit at $-\infty$).

We say that $f$ tends to $b$ when $x$ tends to $+\infty$ if, for any $\varepsilon > 0$, there exists $\alpha > 0$ such that:

$$x > \alpha \Rightarrow |f(x) - b| < \varepsilon$$

This limit is written:

$$\lim_{x \to +\infty} f(x) = b$$

In the same spirit, we say that $f$ tends to $b$ when $x$ tends to $-\infty$ if, for all $\varepsilon > 0$, there exists $\alpha > 0$ such that:

$$x < \alpha \Rightarrow |f(x) - b| < \varepsilon$$

This limit is written:

$$\lim_{x \to -\infty} f(x) = b$$

### 2.2.3 Case of infinite limits

In some cases limits do not exist or are not finite.

**Definition 75** $f$ *tends to* $+\infty$ *when* $x$ *tends to* $a$ *(we note* $\lim_{x \to a} f(x) = +\infty$*)*

*if, for all* $L > 0$, *there exists* $\alpha > 0$ *such that :*

$$|x - a| < \alpha \Rightarrow f(x) > L$$

The definition is similar (with $f(x) < -L$) when $f$ tends to $-\infty$.

In the general case where $f$ is defined over the entire real line $\mathbb{R}$, we can face the situation where $f$ tends to $+\infty$ when $x$ tends to $+\infty$. In that case, we write:

$$\lim_{x \to +\infty} f(x) = +\infty$$

**Example 76** *Denote* $f$ *the following function*

$$f(x) = 1/x \tag{2.2}$$

*The domain* $D$ *is* $\mathbb{R}^*$ *and we easily obtain:*

$$\lim_{x \to 0^+} f(x) = +\infty \tag{2.3}$$
$$\lim_{x \to 0^-} f(x) = -\infty \tag{2.4}$$

*Remember that* $x \to 0^+$ *means* $x \to 0$ *and* $x > 0$*; of course* $x \to 0^-$ *means* $x \to 0$ *and* $x < 0$.

### 2.2.4 Essential properties of limits

The following propositions show that limits of sums/products of functions are intuitively deduced of the limits of the functions entering the sums/products, except when some of these limits are infinite.

**Proposition 77** *Let* $f$ *and* $g$ *be two functions defined on a domain* $D$*. We assume that* $f$ *and* $g$ *have finite limits* $b$ *and* $b'$ *when* $x$ *tends to* $a$*, that is* $\lim_{x \to a} f(x) = b$ *and* $\lim_{x \to a} g(x) = b'$*. We then have the following properties:*

*1)* $\lim_{x \to a} (f + g)(x) = b + b'$

*2)* $\lim\limits_{x\to a}(fg)(x) = bb'$
*3) For any real number* $\alpha$*,* $\lim\limits_{x\to a}\alpha f(x) = \alpha b$
*4) If* $b' \neq 0$*,* $\lim\limits_{x\to a}\frac{f(x)}{g(x)} = \frac{b}{b'}$
*5) If* $\lim\limits_{x\to a} f(x) = b$ *then* $\lim\limits_{x\to a} |f|(x) = |b|$

As mentioned above, the rules provided in proposition 77 work only when the limits $b$ and $b'$ are finite. In the general case (where at least one limit may be infinite) the computations are a little bit more involved.

**Proposition 78** *With the same notations as before, assume that* $b = +\infty$ *; we then have:*
*1)* $\lim_{x\to a}\frac{1}{f(x)} = 0$
*2) If* $g$ *is bounded below, then* $\lim_{x\to a}(f(x)+g(x)) = +\infty$
*3) If* $g$ *has a strictly positive lower bound,* $\lim_{x\to a} f(x)g(x) = +\infty$

Though the two preceding propositions allow to calculate many limits, some cases remain unclear. If, for example, the limits of the functions $f$ and $g$ are respectively $-\infty$ and $+\infty$, the limit of the sum depends on the rate at which each function tends toward infinity.

If $f(x) = -1/x$ and $g(x) = 1/x^2$, the sum tends toward $+\infty$ when $x$ tends toward 0. Indeed, we write:

$$f(x) + g(x) = -\frac{1}{x} + \frac{1}{x^2} = -\frac{1}{x}\left(1 - \frac{1}{x}\right)$$

We can write the sum $f + g$ as the product of two terms that both tend to $-\infty$ when $x$ tends to 0. Thus, the product tends toward $+\infty$.The reason is simply that $g$ goes to infinity much more quickly than $f$ because of the power 2 in the definition of $g$.

### 2.2.5 Continuous functions

The notion of "continuous function" is very intuitive. When one represents the graph of a function $f$, saying that this graph is "continuous" simply means that there is no break in the curve representing $f(x)$. The mathematical definition of continuity says nothing else. However, it allows to consider the case of continuity at a given point, meaning that there is no break in the graph around that point.

**Definition 79** *A function $f : D \to \mathbb{R}$ is* ***continuous*** *at $x_0 \in D^\circ$ if*

$$\lim_{x \to x_0} f(x) = f(x_0) \tag{2.5}$$

*A function $f : D \to \mathbb{R}$ is* ***right-continuous*** *at $x_0 \in D^\circ$ if :*

$$\lim_{x \to x_0^+} f(x) = f(x_0)$$

*A function $f$ is **left-continuous** at $x_0 \in D^\circ$ if :*

$$\lim_{x \to x_0^-} f(x) = f(x_0)$$

Right and left continuity are often used to describe trajectories of stochastic processes[2]. For example, the path of the price process of a stock or an index is often assumed to be a continuous function of time. Figures 2.3 and 2.4 represent the evolution of the S&P500 index for two periods of time. On the first figure, the evolution covers a ten-year period (2002-2011) with daily data. The second figure represents one day of tick-by-tick variations. In the two cases, the continuous approximation does not look as a too artificial assumption. However, looking more closely to the first figure raises the question of what happened in September 2008. Maybe a right-continuity assumption could be justified according to the sudden drop (remember the Lehman Brothers bankruptcy) in the value of the index. The two assumptions (continuity or right-continuity) are encountered in financial models. However, it is worth to mention that introducing the possibility of jumps in price paths leads to much more complicated mathematical models of financial markets.

**Proposition 80** *Let $f$ and $g$ denote two functions defined on $D$, continuous at $a \in D$, and let $c$ be a real number. The sum $f+g$, the products $cf$ and $fg$ and, if $g(a)$ is different from $0$, the quotient $f/g$, are continuous functions at $a$.*

### 2.2.6 Intermediate value theorem

**Proposition 81** *Let $f : [a;b] \to \mathbb{R}$ be a continuous function; if $f(a) < 0 < f(b)$, there exists $c \in ]a;b[$ so that $f(c) = 0$.*

[2] Roger. P (2010b), *Stochastic Processes for Finance, www.bookboon.com*

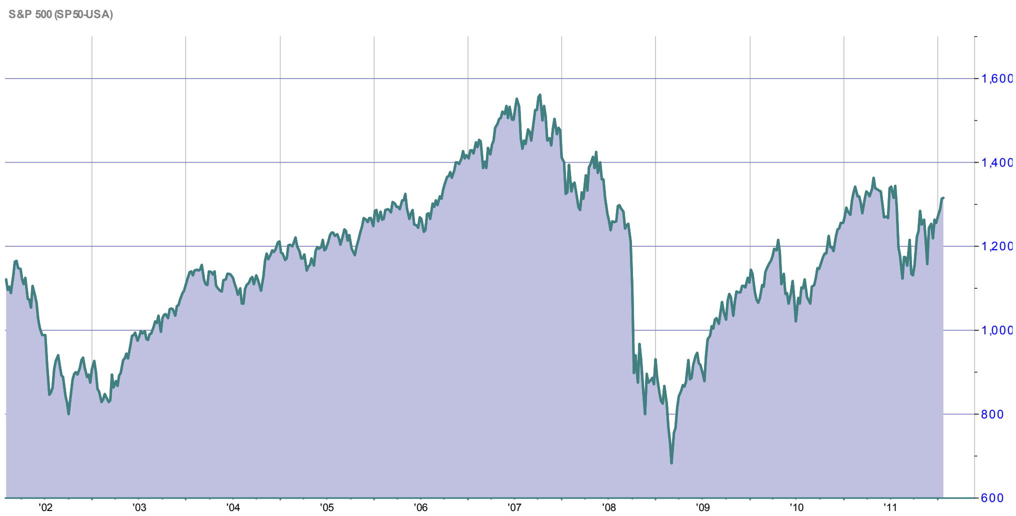


Figure 2.3: Daily evolution of the S&P500 index from 2002 to 2011

Consider figure 2.5 which represents the function $f(x) = x^2 - 4$ on the interval $[0; 4]$. This function satisfies $f(0) = -4 < 0$ and $f(4) = 12 > 0$. As it is of course continuous (and moreover monotonic) the graph crosses the horizontal axis at $x = 2$.

This proposition could also be written by replacing 0 by a real number $\alpha$, then writing "if $f(a) < \alpha < f(b)$, there exists $c \in ]a; b[$ so that $f(c) = \alpha$ ". For example, on figure 2.3 consider an index value of 1000. As the minimum value of the index is around 700 and the maximum at 1,500, there exists at least one point in time at which the value of the index is 1,000, if we assume that the index value moves continuously. This example shows that the intermediate value theorem is very intuitive. It is also easy to understand why continuity is a necessary assumption of proposition 81. If jumps are allowed in the evolution of the S&P500 index, we could have an index value of 1,030 at the market close and an opening value of 990 the day after, without any moment with an index value equal to 1,000.

**Proposition 82** *If $f$ is defined, continuous and strictly monotonic on the*

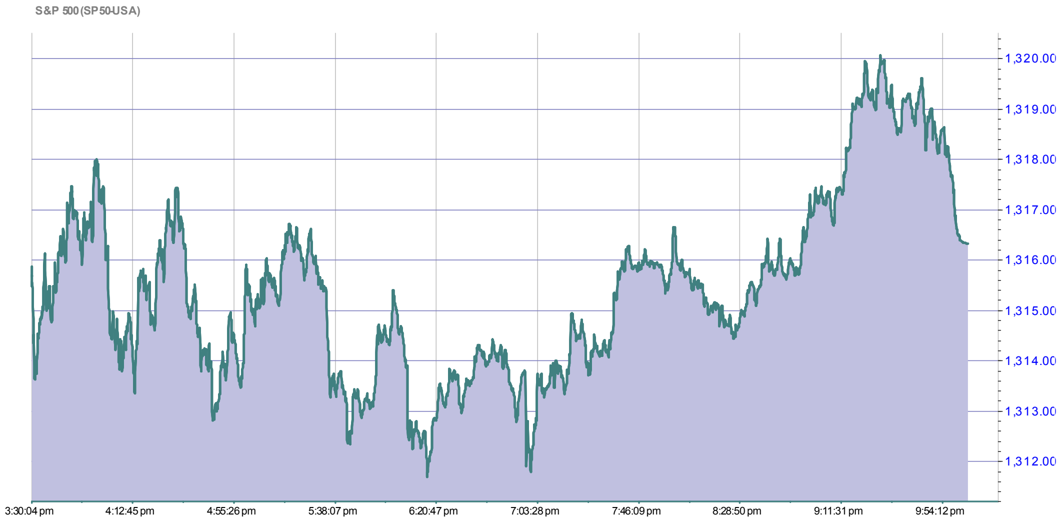


Figure 2.4: Tick by tick evolution of the S&P500 on 01/28/2012 (European time on the X-axis)

*interval* $I = [a; b]$ *with* $I \subset D$, *we have :*

*1)* $f(I)$ *is an interval ending at* $f(a)$ *and* $f(b)$.

*2)* $f$ *is a bijection from* $I$ *to* $f(I)$.

*3) The inverse of* $f$, *denoted* $f^{-1}$, *is continuous, strictly monotonic, and varies in the same direction as* $f$.

**Proof.** (1) is straightforward as $f$ is strictly monotonic.

(2) Assume that $f$ is strictly increasing; $f$ is then injective. Indeed, for two real numbers $x_1$ and $x_2$ satisfying $x_1 < x_2$, we have $f(x_1) < f(x_2)$. Similarly, $f$ is surjective from $I$ to $f(I) = [f(a); f(b)]$ as $f$ is strictly increasing.

(3) follows from the fact that the graph of $f^{-1}$ is symmetric to the graph of $f$ with respect to the first bisector. For example, figure 2.6 illustrates this property. The graphs of $f(x) = x^2$ and $f^{-1}(x) = \sqrt{x}$ on the interval $[0; 1]$ are symmetric with respect to the first bisector represented by the dashed line. ■

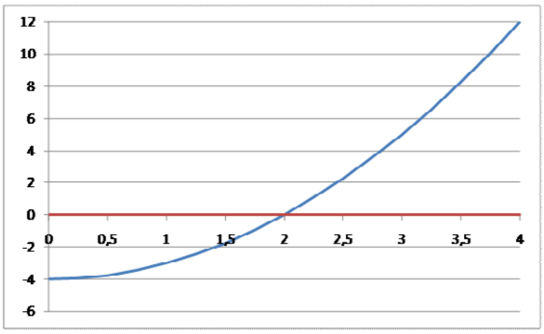


Figure 2.5: $F(x) = x^2 - 4$

When you study option pricing, you learn that option prices are strictly increasing functions of volatility. Proposition 82 says that to a given option price corresponds one and only one volatility level. The consequence is that professionals often value options by referring to the volatility corresponding to the observed price, called the implicit volatility. The advantage is that referring to implicit volatilities allow comparisons between options with different characteristics (maturity, exercise price, etc.) as long as the underlying stock is the same for the options under consideration.

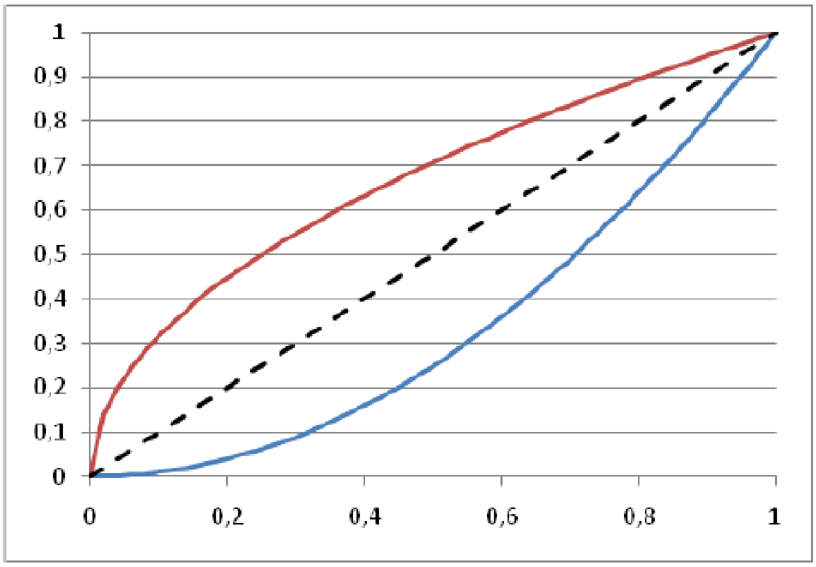


Figure 2.6: Graphs of $f(x) = x^2$ (lower curve) and $f^{-1}(x) = \sqrt{x}$ (upper curve) over the interval $[0; 1]$

**Proposition 83** *Let $f$ be a continuous function defined on a compact domain $D$. $f$ has a maximum and a minimum on $D$ and reaches its bounds.*

Remember that a compact set in $\mathbb{R}$ is a closed and bounded set (see chapter 1). So, consider a sequence $(x_n, n \in \mathbb{N})$ included in a compact set; there exists a convergent subsequence. Denote for example $(x_{n_k}, k \in \mathbb{N})$ this subsequence and $x^*$ the limit. Since $f$ is continuous, $f(x_{n_k})$ converges to $f(x^*)$. Applying this reasoning and assuming that $f$ does not reach its bounds leads to a contradiction.

### 2.2.7 Convex and concave functions

Convexity and concavity are the most widely used properties of functions in economic analysis and finance. When utility functions are introduced in microeconomics lectures, one of the first reasonable behavioral assumptions is the decrease of the marginal utility of consumption. It simply means that utility of consuming one unit of a given good decreases with the number of units already consumed. Concavity is the technical translation of this assumption. Convexity appears naturally when compounding interests. The fact that interest obtained on an investment in period 1 generates interest in period 2 implies that the marginal increase of the value of your investment is increasing. This is nothing else than convexity.

**Definition 84** *1) Let $f$ be a function defined on an interval $[a;b]$. $f$ is* ***convex*** *if for all $\alpha \in [0;1]$ and all couples $(x,y) \in [a;b] \times [a;b]$ we have:*

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

*2) Under the same hypotheses, $f$ is* ***concave*** *if the inequality is reversed.*

*3) If the inequality is strict in (1), the function is said to be* ***strictly convex****. If the inequality is strict in (2), the function is said to be* ***strictly concave****.*

Figure 2.6, used to illustrate inverse functions represented two functions defined on the interval $[0;1]$ and taking values 0 at 0 and 1 at 1. The first function is $f(x) = x^2$; it is convex. On the figure, it can be seen that $f$ is below the dashed line joining (0,0) and (1,1). It is the main geometrical property of convex functions. Lines joining two points on the curve are above the curve. The second function on figure 2.6 is $f^{-1}(x) = \sqrt{x}$. It is concave and above the dashed line.

## 2.3 Differentiation

### 2.3.1 Derivatives: definitions

In this section, all functions are defined on a domain $D$ and $I$ is an open interval included in $D$.

**Definition 85** *1) The **derivative** of $f$ at $x_0 \in I$, denoted $f'(x_0)$, is the limit defined by:*

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

*If the limit $f'(x_0)$ exists, $f$ is said differentiable at $x_0$.*

*2) When $f'(x_0)$ exists for all $x_0$ in $D$, $f$ is simply said to be **differentiable**. The function $x \to f'(x)$ is called the derivative of $f$.*

If $f$ is differentiable at $x_0$, $f(x_0 + h)$ can be approximated as:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \varepsilon(h)h \tag{2.6}$$

where $\varepsilon$ is a continuous function that satisfies $\lim_{h \to 0} \varepsilon(h) = 0$. It follows that any function differentiable at $x_0$ is continuous at $x_0$.

The geometric interpretation of the above equality is that, in the neighborhood of $x_0$, the graph of $f$ is close to the tangent (at $x_0$) to the curve representing $f$.

**Definition 86** *A differentiable function $f$, is **continuously differentiable** if the derivative $x \to f'(x)$ is continuous. $f$ is then called a $C^1$-function.*

**Proposition 87** *If the derivative of a **continuously differentiable** function $f$ is strictly positive (negative) on $I$, $f$ is increasing (decreasing) on $I$. Symmetrically, if $f$ is increasing (decreasing) on $I$, the derivative $f'$ is positive (negative) or equal to 0 at any $x_0 \in I$.*

The following remarks on derivatives are important:

a) The derivative of a function does not necessarily exists. A derivative at $x_0$ can be defined only if $x_0$ is interior to $D$, because $x_0 + h$ must belong to $D$ for $h$ sufficiently small. For example, the function $g(x) = \sqrt{x}$ is not differentiable at 0 as $g$ is not defined on the left of 0. Consequently, for all negative $h$ , $x_0 + h$ is not in the domain if $x_0 = 0$.

b) For a function to be differentiable, the limit given in the definition must be the same (and finite) when $h > 0$ and $h < 0$. It follows that the function $f(x) = |x|$ is not differentiable at $x_0 = 0$ because:

$$\lim_{h \to 0^+} \frac{|h|}{h} = 1 \quad (2.7)$$

$$\lim_{h \to 0^-} \frac{|h|}{h} = -1 \quad (2.8)$$

c) A function which is not continuous cannot be differentiable. A simple example to illustrate this remark is the following function:

$$f(x) = \begin{cases} 1 \text{ if } x > 0 \\ -1 \text{ if } x \leq 0 \end{cases}$$

For any $x \neq 0$, $f'(x) = 0$ but $f'(0)$ is not defined because the sign of $h$ is not specified in definition 85.

**Definition 88** *1) The **right-derivative** of $f$ at $x_0$, denoted $f'_d(x_0)$ is the limit (if it exists) defined as:*

$$f'_d(x_0) = \lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0}$$

*2) The **left-derivative** of $f$ at $x_0$, denoted $f'_d(x_0)$ is the limit (if it exists) defined as:*

$$f'_l(x_0) = \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0}$$

### 2.3.2 Properties of derivatives

The properties of derivatives are not as intuitive as the properties linked to continuity, especially when considering products of functions. However, these properties are of constant use and must be perfectly known.

**Proposition 89** *Let $f$ and $g$ be two functions differentiable at $x_0 \in D$ and $c \in \mathbb{R}$ :*

1. $(f+g)'(x_0) = f'(x_0) + g'(x_0)$
2. $(cf)'(x_0) = c \times f'(x_0)$
3. $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$
4. *If* $g(x_0) \neq 0$, $\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g^2(x_0)}$
5. *Let* $\alpha(x) = f(x)^n$. *We then have:*

$$\alpha'(x_0) = nf'(x_0)f(x_0)^{n-1}$$

**Proof.** 1) and 2) are obvious as direct consequences of the definitions of a sum of functions and a product of a function by a real number.

3) We can write:

$$\begin{aligned}\frac{f(x_0+h)g(x_0+h) - f(x_0)g(x_0)}{h} &= g(x_0+h)\frac{f(x_0+h) - f(x_0)}{h} \\ &\quad + f(x_0)\frac{g(x_0+h) - g(x_0)}{h}\end{aligned}$$

If we consider the limit when $h$ tends toward 0, we obtain the previous result.

4) The derivative of a quotient can be computed as the one of a product by denoting $\phi = \frac{1}{g}$. We then have:

$$f'(x_0)\phi(x_0) + f(x_0)\phi'(x_0) = \frac{f'(x_0)}{g(x_0)} + f(x_0)\phi'(x_0)$$

We then just need to demonstrate that $\phi'(x_0) = -\frac{g'(x_0)}{g^2(x_0)}$. We have:

$$\phi'(x_0) = \frac{\frac{1}{g(x_0+h)} - \frac{1}{g(x_0)}}{h} = \frac{1}{g(x_0+h)g(x_0)}\frac{g(x_0) - g(x_0+h)}{h}$$

It follows that:

$$\lim_{h\to 0}\frac{\frac{1}{g(x_0+h)}-\frac{1}{g(x_0)}}{h}=-\frac{g'(x_0)}{g^2(x_0)}$$

5) We use the following relation:

$$\begin{aligned} a^n-b^n &= (a-b)(a^{n-1}+a^{n-2}b...+b^{n-1}) \\ &= (a-b)\sum_{k=0}^{n-1}a^k b^{n-1-k} \end{aligned}$$

Applying this relation using $a=f(x_0+h)$ and $b=f(x_0)$, we obtain:

$$\frac{f(x_0+h)^n-f(x_0)^n}{h}=\frac{f(x_0+h)-f(x_0)}{h}\sum_{k=0}^{n-1}f(x_0+h)^k f(x_0)^{n-1-k}$$

We also know that:

$$\lim_{h\to 0}\sum_{k=0}^{n-1}f(x_0+h)^k f(x_0)^{n-1-k}=nf(x_0)^{n-1}$$

It follows that:

$$\lim_{h\to 0}\frac{f(x_0+h)^n-f(x_0)^n}{h}=f'(x_0)nf(x_0)^{n-1}$$

■

The following corollary is a special case of part (5) of proposition 89. It allows, combined with parts (1) and (2), to calculate the derivative of any polynomial

$$p(x)=\sum_{k=0}^{n}a_k x^k \tag{2.9}$$

**Corollary 90** *The derivative of $f(x)=x^n$ for $n\geq 1$ is $f'(x)=nx^{n-1}$.*

We then get

$$p'(x) = \sum_{k=1}^{n} k a_k x^{k-1} \tag{2.10}$$

The last part of proposition 89 is also a particular case of the following proposition related to the derivation of compound functions.

### 2.3.3 Derivative of compound functions

Compound functions appear naturally in financial models. For example, in option valuation models, some elements are probabilities linked to normal distributions. The formulation of these probabilities is quite complex and, when dealing with management problems, one has to find derivatives of option prices with respect to basic variables like maturity, volatility or the underlying stock price. It is then necessary to calculate derivatives of compound functions.

**Proposition 91** *Let $f$ and $g$ be two functions differentiable respectively at $x$ and $y = f(x)$. The derivative of the compound function $g \circ f$ at $x_0$ is given by:*

$$(g \circ f)' (x_0) = f'(x_0)g' [f(x_0)]$$

Consider for example $h(x) = f(x)^n$, we can write:

$$h(x) = g(f(x))$$

where $g(y) = y^n$. Applying proposition 91 leads to $g'(f(x_0)) = nf(x_0)^{n-1}$; it is the result obtained before. In general we have:

$$[f(x)^n]' = nf'(x)f(x)^{n-1} \tag{2.11}$$

Similarly, if $f$ is strictly monotonic, it is bijective. As such $f$ has an inverse function, noted $f^{-1}$, such that $f^{-1} \circ f = i_D$ where $i_D$ is the identity mapping defined on the domain $D$. Applying the previous proposition gives:

$$\left[f^{-1} \circ f\right]' (x_0) = f'(x_0) \left(f^{-1}\right)' [f(x_0)] = i_D'(x) = 1$$

Therefore, we can deduce the following corollary.

**Corollary 92** *Let $f$ be a function differentiable at $x_0$, strictly monotonic and such that $f'(x_0) \neq 0$. Its inverse $f^{-1}$ has a derivative defined by:*

$$\left(f^{-1}\right)'(y_0) = \frac{1}{f'(x_0)}$$

*with $y_0 = f(x_0)$.*

Remember proposition 82 saying that $f$ and $f^{-1}$ vary in the same direction and that the graph of $f^{-1}$ is symmetric to the one of $f$ with respect to the first bisector. Indeed, the product of the slopes of two symmetric lines (with respect to the first bisector) is equal to 1.

**Example 93** *Let $f(x) = x^2$ be defined on $[0; +\infty[$; its inverse function is $f^{-1}(y) = \sqrt{y}$ defined on $\mathbb{R}_+$. Applying the previous proposition leads to:*

$$(\sqrt{y})' = \frac{1}{2\sqrt{y}}$$

*because $f'(x) = 2x$.*

### 2.3.4 Higher-order derivatives

Financial models deal with higher-order derivatives, especially second-order derivatives, in several fields. In bond portfolio management, what is called the "convexity" of a bond is measured by a second-order derivative. In the world of options, the so-called gamma coefficient is nothing else than a second-order derivative. Finally, in microeconomics, these second-order derivatives are used to define risk aversion coefficients.

**Definition 94** *Let $f$ be a function from $D$ to $\mathbb{R}$, differentiable in the neighborhood of $x_0 \in \overset{\circ}{D}$. The **second-order derivative** of $f$ at $x_0$, denoted $f''(x_0)$ is the limit, if it exists, given by:*

$$f''(x_0) = \lim_{h \to 0} \frac{f'(x_0 + h) - f'(x_0)}{h}$$

*where $f'$ is the derivative of $f$.*

The properties of the first order derivative apply for higher order derivatives. In fact, the above definition shows that the second-order derivative is nothing else than the derivative of the first-order derivative.

**Definition 95** *Let $f$ be a function defined on $D$ with value in $\mathbb{R}$, differentiable $n$ times in the neighborhood of $x_0 \in D^\circ$. We call **derivative of order** $n+1$ of $f$ at $x_0$ and we note $f^{(n+1)}(x_0)$ the limit, if it exists, given by:*

$$f^{(n+1)}(x_0) = \lim_{x \to x_0} \frac{f^{(n)}(x) - f^{(n)}(x_0)}{x - x_0}$$

**Example 96** *The calculations in the two examples (a) and (b) below are a direct consequence of corollary 90*

*a)* $f(x) = 4x^3 - 2x^2 + 1$

$$\begin{aligned} f'(x) &= 12x^2 - 4x \\ f''(x) &= 24x - 4 \\ f^{(3)}(x) &= 24 \end{aligned}$$

*b)* $f(x) = \frac{x^2+1}{2x-1}$

$$\begin{aligned} f'(x) &= \frac{2x(2x-1) - 2(x^2+1)}{(2x-1)^2} = \frac{2x^2 - 2x - 2}{(2x-1)^2} = 2\frac{x^2 - x - 1}{(2x-1)^2} \\ f''(x) &= 2\frac{(2x-1)(2x-1)^2 - 2(2x-1)(2x^2 - 2x - 2)}{(2x-1)^4} \\ &= 2\frac{(2x-1)^2 - 4(x^2 - x - 1)}{(2x-1)^3} \\ &= \frac{10}{(2x-1)^3} \end{aligned}$$

The sign of the second-order derivative helps to determine if the function is convex or concave.

**Proposition 97** *Let $f$ be a twice differentiable function on $D^\circ$. $f$ is concave (convex) iff $f''(x) \leq (\geq) 0$ for all $x$ in $D^\circ$.*

A function is said strictly concave (convex) if $f''$ is strictly negative (positive).

## 2.4 Logarithms and exponential functions

As mentioned in the introduction of the chapter, logarithms and exponential functions are among the most important functions in finance and economics. There are many economic and mathematical reasons for that. As we will see in this section, these functions are infinitely differentiable, they are convex (exponential functions) or concave (logarithms) and are well suited to describe economic phenomena. Logarithms are widely used to represent preferences of economic agents (utility functions) but also to calculate stock returns, especially when markets work in continuous-time. Exponential functions are also used as utility functions (more precisely negative exponentials) but come naturally when capitalizing money over time at a continuous rate of interest.

### 2.4.1 Exponential functions

**Definition 98** *For a given positive number $a$, the function $x \to f(x) = a^x$ defined on $\mathbb{R}$ is called the $a$-exponential function.*

When $x$ is an integer $a^x$ is simply the product $a \times a \times a....$ ($x$ times). When $x$ is a rational number, that is $x = p/q$ where $p$ and $q$ are positive integers, $a^x = \sqrt[q]{a \times a \times a....}$ where the product is repeated $p$ times. If $x$ is a negative integer, $a^x$ is $1/(a \times a \times a....)$ where the denominator is the $x$ times product of $a$.

**Example 99** *Assume that, in the above definition, $x$ measures a duration and $a$ is the appreciation rate of your wealth in a given year ($a = 1+r$ where $r$ is the interest rate). The quantity $a^x$ is the amount you possess after $x$ years, when your initial investment is \$1 at date 0.*

**Proposition 100** *For any positive number $a$, $a$-exponential functions satisfy:*

1. $\forall (x,y) \in \mathbb{R}^2, \;\; a^x a^y = a^{x+y}$
2. $\forall (x,y) \in \mathbb{R}^2, \;\; \frac{a^x}{a^y} = a^{x-y}$
3. $\forall (x,y) \in \mathbb{R}^2, \; (a^x)^y = a^{xy}$
4. $\lim_{x \to -\infty} a^x = 0$
5. $\lim_{x \to +\infty} a^x = +\infty$
6. $a^0 = 1$
7. *The function $x \to a^x$ is strictly increasing and strictly convex.*

### 2.4.2 $e$-exponential functions

Suppose you invest \$1 with an interest rate of 100 %. What does this sentence mean exactly? It means that you will receive \$2 in one year if the compounding frequency of interests is one year. If interests are compounded monthly, you are going to receive $\$\left(1+\frac{1}{12}\right)^{12}$, and if the compounding frequency is daily, the amount received is $\$\left(1+\frac{1}{365}\right)^{365}$. When interests are compounded on a continuous basis, the final amount is defined as:

$$e = \lim_{n \to +\infty} \left(1 + \frac{1}{n}\right)^n \tag{2.12}$$

$e$ is approximately equal to 2.71828.

**Definition 101** *The e-**exponential** is the function $f$ defined by:*

$$f(x) = e^x$$

*where* $e = \lim_{n \to +\infty} \left(1 + \frac{1}{n}\right)^n$.

For the above issue of capitalization, $e^x$ is the amount received after one year when \$1 is invested at a continuous-rate $x$. This amount is equal to:

$$\lim_{n \to +\infty} \left(1 + \frac{x}{n}\right)^n$$

But stating $n = mx$, allows us to write:

$$\lim_{n \to +\infty} \left(1 + \frac{x}{n}\right)^n = \lim_{m \to +\infty} \left(1 + \frac{1}{m}\right)^{mx} = \left[\lim_{m \to +\infty} \left(1 + \frac{1}{m}\right)^m\right]^x = e^x$$

**Remark 102** *When "exponential function" is used without other specifications, it means e-exponential function.*

**Proposition 103** *The function $x \rightarrow e^x$ is strictly increasing and strictly convex; moreover, $(e^x)' = e^x$.*

To give the intuition of the result, denote $f_n(x) = \left(1 + \frac{x}{n}\right)^n$. The derivative of $f_n$ is given by:

$$\begin{aligned} f_n'(x) &= n \times \frac{1}{n}\left(1 + \frac{x}{n}\right)^{n-1} \\ &= \left(1 + \frac{x}{n}\right)^n \times \frac{1}{1 + \frac{x}{n}} = \frac{n}{n+x} f_n(x) \end{aligned}$$

When $n$ tends to $+\infty$, the ratio $\frac{n}{n+x}$ tends to 1. It follows that the derivative of $f_n'(x)$ is close to $e^x$ when $n$ is sufficiently large. As $e^x$ is increasing in $x$, the derivative (which is still $e^x$) is also increasing and $e^x$ is convex by proposition 97.

Figure 2.7 shows the graph of $e^x$ for $x$ varying between -1 to 3.

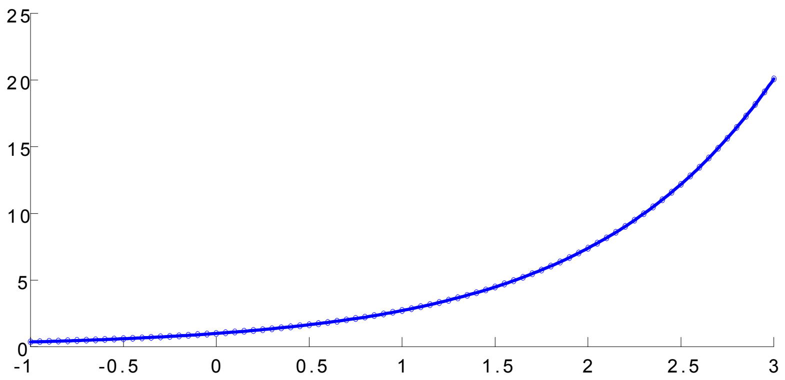


Figure 2.7: Exponential function $f(x) = e^x$

### 2.4.3 Logarithms

In the previous section, we saw that $a$-exponential functions are strictly increasing on $\mathbb{R}$ and take values in $\mathbb{R}_+^*$ ; $a^x$ is then a bijection from $\mathbb{R}$ to $\mathbb{R}_+^*$.

It has an inverse function called $a$-logarithm.

**Definition 104** *The $a$-**logarithm,** denoted $\log_a(.)$ is the function defined by:*

$$y \in \mathbb{R}_+^* \to z = \log_a(y) \text{ so that } a^z = y$$

It follows that $\log_a(a^z) = z$ ; the property 100 (of the exponential functions) induces the following one for $a$-logarithms.

**Proposition 105** *The a-logarithm has the following properties:*

1. $\forall (x,y) \in \mathbb{R}_+^*,\ \log_a(xy) = \log_a(x) + \log_a(y)$
2. $\forall (x,y) \in \mathbb{R}_+^*,\ \log_a(\frac{x}{y}) = \log_a(x) - \log_a(y)$
3. $\forall (x,y) \in \mathbb{R}_+^*,\ \log_a(x^y) = y \log_a(x)$
4. $\lim_{x \to 0^+} \log_a(x) = -\infty$
5. $\lim_{x \to +\infty} \log_a(x) = +\infty$
6. $\log_a(1) = 0$
7. *The function $x \to \log_a(x)$ is strictly increasing and strictly concave.*

When $a = e$, we simply note $\ln(x)$ for $\log_e(x)$. This logarithm is called Neper logarithm (referring to the mathematician John Napier)....or simply natural logarithm. We also get for logarithms a proposition similar to proposition 97

**Proposition 106** *The function $x \to f(x) = \ln(x)$, defined on $\mathbb{R}_+^*$ is strictly increasing, strictly concave, and its derivative is given by $f'(x) = \frac{1}{x}$.*

Remember that the derivative of $e^x$ is $e^x$ and that $\ln(x)$ is the inverse function of $e^x$. Applying corollary 92 gives immediately the derivative of $\ln(x)$.

Figure 2.8 shows the graph of $\ln(x)$ for $x$ varying between 0.1 to 4.

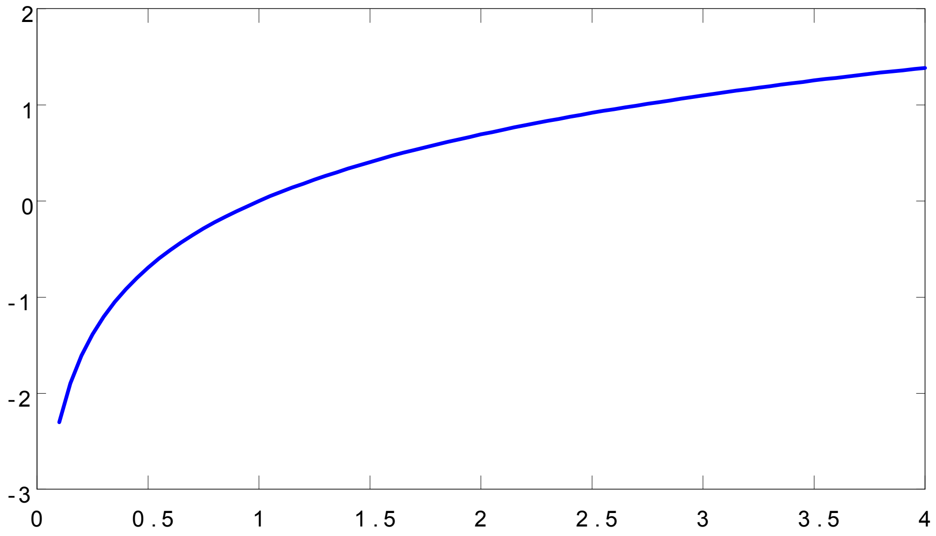


Figure 2.8: Logarithm function $f(x) = \ln(x)$

## 2.5 Polynomial approximations and Taylor formula

Logarithms and exponential functions presented in the preceding section are "complex" functions, in the sense that they cannot be written as polynomials. Of course, most mathematical functions, even those that are infinitely differentiable, are not polynomials. However, all regular functions are close to polynomials. They can be approximated by polynomials; the accuracy of the approximation depends on the degree of the polynomial and on the choice of coefficients. This is the essential role of the Taylor expansion (or Taylotr formula) presented hereafter.

### 2.5.1 Rolle's theorem

**Proposition 107** ***Rolle's theorem***

*Let $f$ be a function defined on an interval $[a;b]$, taking values in $\mathbb{R}$ so that $f(a) = f(b)$, and whose derivative $f'$ is continuous. It then exists $c \in ]a;b[$ satisfying $f'(c) = 0$.*

This important proposition is very intuitive. If $f(a) = f(b)$, two situations are possible. Either $f(x) = f(a)$ for any $x \in [a;b]$ and $f'(x) = 0$ since the function is constant. Or there exists $x \in ]a;b[$ such that $f(x) \neq f(a)$. Assume without loss of generality that $f(x) > f(a)$. It means that the derivative of $f$ is strictly positive somewhere between $a$ and $x$. But to reach $b$ with $f(b) = f(a) < f(x)$, $f$ must decrease somewhere between $x$ and $b$...and the derivative is negative at at least one point. But the derivative being continuous, it must be equal to 0 at one point in the interval $]a;b[$ by the intermediate value theorem (proposition 81).

The geometric interpretation of this proposition is the following. Consider the line joining $(a, f(a))$ and $(b, f(b))$. By the assumption of Rolle's theorem, this line is horizontal with a zero slope, equal to $f'(c)$.

**Example 108** *Let $f$ be defined by:*

$$f(x) = x^2 - 3x + 2$$

*$f(x) = 0$ for $x = 1$ and $x = 2$. The derivative of $f$ is equal to $f'(x) = 2x - 3$, therefore it is equal to 0 at $x = \frac{3}{2}$, point located in the interval $[1;2]$ at the ends of which the function $f$ is equal to 0.*

*Figure 2.9 shows the graph of $f(x)$ with the null derivative at $x = 3/2$. As $f(1) = f(2) = 0$, the horizontal line joining these points is parallel to the tangent $(y = -0.25)$ at $x = 3/2$.*

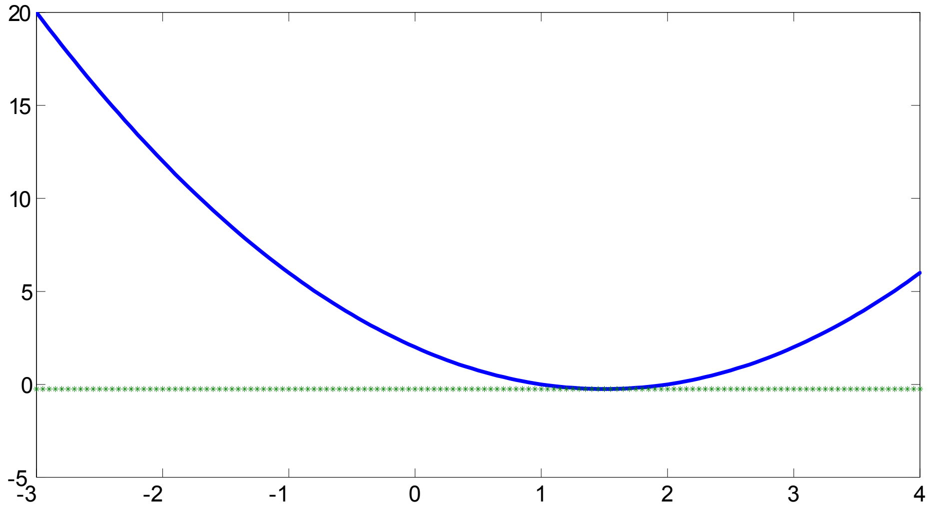


Figure 2.9: $f(x) = x^2 - 3x + 2$

### 2.5.2 Linear approximation

**Proposition 109** *Let $f$ be a differentiable function defined on an interval $[a;b]$, taking values in $\mathbb{R}$. There exists $c \in [a;b]$ such that:*

$$f(b) - f(a) = (b-a) f'(c)$$

Indeed, let us consider the function $g$ defined by:

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b-a}(x-a)$$

This function satisfies $g(a) = g(b) = 0$. Rolle's theorem implies that there exists $c$ so that $g'(c) = 0$, that is:

$$g'(c) = f'(c) - \frac{f(b) - f(a)}{b-a} = 0$$

which proves the result of proposition 109.

Denoting $b = a + h$ and assuming $h$ small leads to:

$$f(a+h) = f(a) + hf'(c)$$

with $c \in [a; a+h]$.

This equality means that we can approximate $f(a+h)$ by $f(a) + hf'(a)$ when $h$ is sufficiently small. In the neighborhood of $a$, $f$ can be approximated by a linear function.

### 2.5.3 Taylor's formula

The straight generalization of the above linear approximation is the so-called Taylor formula allowing to approximate a sufficiently regular function by a $n$-degree polynomial.

**Proposition 110** ***Taylor's formula***

*Let $f$ be a function defined on an interval $[a;b]$, taking values in $\mathbb{R}$, and having continuous derivatives up to order $n+1$. There exists $c \in ]a;b[$ so that:*

$$\begin{aligned} f(b) &= f(a) + (b-a) f'(a) + \frac{(b-a)^2}{2} f''(a) + ... \\ & \frac{(b-a)^n}{n!} f^{(n)}(a) + \frac{(b-a)^{n+1}}{(n+1)!} f^{(n+1)}(c) \end{aligned}$$

Taylor formula is often used with $a = 0$ and $b = x$ where $x$ is a number close to 0. In that case, we can write $c \in ]a;b[$ as $\alpha x$ where $\alpha$ is a number between 0 and 1. This particular case of the Taylor formula is often called the **Mac-Laurin formula**.

**Proposition 111** *Let $f$ be a function defined on an interval $I$ containing 0, taking values in $\mathbb{R}$, and having continuous derivatives up to order $n+1$.*

*There exists* $\alpha \in ]0;1[$ *so that:*

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2}f''(0) + \frac{x^n}{n!}f^{(n)}(0) + \frac{x^{n+1}}{(n+1)!}f^{(n+1)}(\alpha x) \quad (2.13)$$

*where* $x \in I$.

The last term of the RHS of equation (2.13) can be written as:

$$\frac{x^{n+1}}{(n+1)!}f^{(n)}(\alpha x) = x^n \times \frac{xf^{(n)}(\alpha x)}{(n+1)!}$$

If $\frac{xf^{(n)}(\alpha x)}{(n+1)!}$ tends to 0 when $x \to 0$, we can write:

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2}f''(0) + \frac{x^n}{n!}f^{(n)}(0) + x^n\varepsilon(x)$$

with $\lim_{x\to 0}\varepsilon(x) = 0$.

**Definition 112** *A function* $f$ *has a* ***Taylor expansion*** *of order* $n$ *in the neighborhood of 0 if there exist real numbers* $a_0, ..., a_n$ *and a function* $\varepsilon$ *such that:*

$$f(x) = a_0 + a_1x + ...a_nx^n + x^n\varepsilon(x)$$

*where the function* $\varepsilon$ *satisfies* $\lim_{x\to 0}\varepsilon(x) = 0$.

**Definition 113** *A function* $g$ *defined on an open set* $D \subset \mathbb{R}$ *has a Taylor expansion of order* $n$ *in the neighborhood of* $x_0 \in D$ *if there exist* $b_0, ..., b_n$ *and a function* $\varepsilon$ *such that:*

$$g(x) = b_0 + b_1(x - x_0) + ... + b_n(x - x_0)^n + (x - x_0)^n\varepsilon(x)$$

*where the function* $\varepsilon$ *satisfies* $\lim_{x\to x_0}\varepsilon(x) = 0$.

The remainder of the Taylor expansion, that is the term $(x - x_0)^n\varepsilon(x)$ is often written $o((x - x_0)^n)$ because $\lim_{x\to x_0}\varepsilon(x) = 0$.

**Proposition 114** *If a function $g$ has a Taylor expansion of order $n$ in the neighborhood of $x_0 \in D$, it is unique.*

# Chapter 3

# Integrals

Integrals of functions may be approacheded from at least two directions. First, the analytical approach consists in determining, for a given known function $f$, a function $F$ whose derivative $F'$ is equal to $f$. We could say that $F$ is the "anti-derivative" of $f$.

Second, the geometric approach defines the integral of $f$ between $a$ and $b$ ( if $f$ is defined and positive on a given interval $[a;b]$) as the area between the horizontal axis and the curve representing $f$ between $a$ and $b$.

The last interpretation of the integral is the average value of the function $f$ over the interval $[a;b]$ . In probability theory, the expectation of a random variable $X$ is a weighted average of the possible values of $X$. It is nothing less than an integral. For example, the expected return of a given stock over a future period of time can be calculated as an integral.

## 3.1 Integral of a step function

Everybody knows how to calculate the area of a rectangle: it is enough to multiply the length by the width to get the result. It is the reason why we start to present integrals by means of the most simple functions, namely the step functions. All functions in this section are defined on an interval $[a;b]$

and take values in $\mathbb{R}$.

### 3.1.1 Step functions: definition

**Definition 115** *A subdivision of an interval $I = [a; b]$ is a set of real numbers $S_x = (x_0, ..., x_p)$ such that:*

$$a = x_0 < x_1 < .... < x_p = b$$

$\max_{i=1,\ ...,p} |x_i - x_{i-1}|$ *is called the step of the* ***subdivision****.*

**Definition 116** *A function $f$ is a* ***step function*** *if there exists a subdivision $S_x = \{x_0, ..., x_p\}$ such that $f$ is constant on any interval $]x_{i-1}; x_i[$ for $i = 1, ..., p$. For any $x \in ]x_{i-1}; x_i[$, the value of $f(x)$ on $]x_{i-1}; x_i[$ is denoted $y_i$.*

This definition does not specify the values of $f$ at $x_i$. In most economic or financial problems, $f(x_i) = y_i$ or $f(x_i) = y_{i-1}$, depending on the right or left continuity of $f$. However, in this chapter these values are not important. More generally, when dealing with integrals of functions, the value of the function at a given point does not matter. We will be more precise on that in the following paragraphs.

Figure 3.1 gives an example of a step function defined by $f(x) = Int(x/5)+1$ where $Int(y)$ is the integer part of number $y$. $x/5$ is lower than 1 when $x < 5$ consequently $f(x) = 1$ for $x < 5$, etc.

When a step function $f$ is positive, it is easy to calculate the area delimited by the horizontal axis and the graph of $f$. This area is a sum of areas of rectangles. On figure 3.1 it is easy to calculate the area under the curve as:

$$5 \times 1 + 5 \times 2 + 5 \times 3 + 5 \times 4 = 50 \tag{3.1}$$

The length of the horizontal axis being 20, it is also easy to see that the average value of $f$ on the interval $[0; 20]$ is computed as $50/20 = 2.5$. In fact

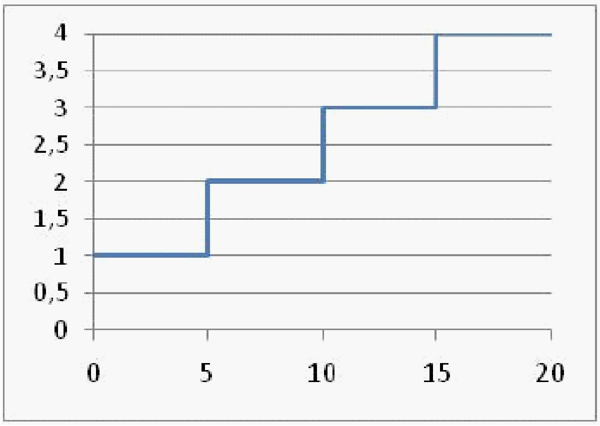


Figure 3.1: $f(x) = Int(x/5) + 1$

2.5 is the arithmetic average of the four values $1, 2, 3, 4$ because the lengths of the intervals over which $f$ is constant are always equal to 5. All the intervals of the subdivision have the same length (weight) in the calculation of the weighted average.

The following proposition summarizes some intuitive properties of step functions.

**Proposition 117** *a) If $f$ and $g$ are two step functions defined on $I = [a; b]$, $f + g$ is a step function on $I$.*

*b) If $f$ is a step function defined on $I = [a; b]$ and $\alpha$ is a real number $\alpha f$ is a step function on $I$.*

*c) If $f$ is a step function defined on $I = [a; b]$, the absolute value $|f|$ is a step function.*

### 3.1.2 Integral of a step function

**Definition 118** *Let $f$ denote a step function defined on $[a; b]$, valued $y_i$ on the interval $]x_{i-1}; x_i[$; the* **integral of $f$** *on $[a; b]$ is denoted $\int_a^b f(x)dx$ and* *defined by:*

$$\int_a^b f(x)dx = \sum_{i=1}^{p} y_i \left(x_i - x_{i-1}\right)$$

The quantity $\int_a^b f(x)dx$ does not depend on the fact that intervals in the subdivision are closed, open or semi-closed(open). Even if $f$ is valued $z_j \neq y_j$ at $x_j$, $\int_a^b f(x)dx$ remains unchanged, simply because an interval $[x_j; x_j]$ has a null length. In the definition, the function $f$ is valued $y_i$ over each interval $]x_{i-1}; x_i[$ ; when the $y_i s$ are positive, each element of the sum $\int_a^b f(x)dx$ is the area of a rectangle with length sizes $y_i$ and width size $(x_i - x_{i-1})$.

**Proposition 119** *Let $f$ denote a step function defined on $[a; b]$. The integral $\int_a^b f(x)dx$ does not depend on the subdivision.*

To be clear about what is said in the proposition, remember the definition of the step function. It is constant over intervals. It means that such a function already defines a "natural" subdivision of the horizontal axis as on figure 3.1. The proposition means that as long as the values taken by the step function are defined, the way $[a; b]$ is sliced does not matter, simply because the step function is constant on each interval $]x_{i-1}; x_i[$. Assume for example that you divide this interval in two sub-intervals$]x_{i-1}; x^*[\bigcup [x^*; x_i[$; it does not change the global area because

$$y_i(x_i - x^*) + y_i(x^* - x_{i-1}) = y_i(x_i - x_{i-1}) \tag{3.2}$$

The following proposition summarizes properties of integrals of step functions but most of these properties remain valid for more general functions. However, the intuition behind these results is clearer for step functions.

**Proposition 120** *Let $f$ and $g$ denote two step functions defined on $I = [a; b]$.*

*a) The integral of the sum* $f+g$ *is equal to the sum of the integrals of* $f$ *and* $g$ *:*

$$\int_a^b (f+g)(x)dx = \int_a^b f(x)dx + \int_a^b g(x)dx$$

*b)* $\forall \alpha \in \mathbb{R}$*, the integral of* $\alpha f$ *is equal to* $\alpha$ *times the integral of* $f$ *:*

$$\int_a^b (\alpha f)(x)dx = \alpha \int_a^b f(x)dx$$

*c) For any* $c \in I$*, we have:*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

*d) If* $f \geq g$ *then* $\int_a^b f(x)dx \geq \int_a^b g(x)dx$

## 3.2 General case

We are now ready to generalize the above definitions to larger families of functions, for example continuous functions (which remain bounded over intervals like $[a;b]$ by proposition 83).

### 3.2.1 Integrable functions on $[a;b]$

Let $f$ denote a continuous function on $[a;b]$ and $a = x_0 < ... < x_p = b$ a subdivision of $[a;b]$. Let $m_i$ and $M_i$ be defined as follows:

$$\begin{aligned} m_i &= \inf \{f(x), x \in [x_{i-1}; x_i[\} \\ M_i &= \sup \{f(x), x \in [x_{i-1}; x_i[\} \end{aligned}$$

$m_i(M_i)$ is the minimum (maximum) value of $f$ on $[x_{i-1}; x_i[$.

Denote $s_p$ and $S_p$ the following quantities:

$$\begin{aligned} s_p &= \sum_{i=1}^{p} m_i (x_i - x_{i-1}) \\ S_p &= \sum_{i=1}^{p} M_i (x_i - x_{i-1}) \end{aligned}$$

$s$ and $S$ are indexed by $p$ which refers to the step of the subdivision equal to $\max_{i=1}^{p} (x_i - x_{i-1})$.

Figure 3.2 illustrates the way we can define the integral of a continuous function by means of $s_p$ and $S_p$. The area below the graph of the function $f(x) = \sqrt{x}$ is bounded by the areas below the graphs of two step functions $g$ (thin line correponding to $S_p$) and $h$ (bold line corresponding to $s_p$). In fact

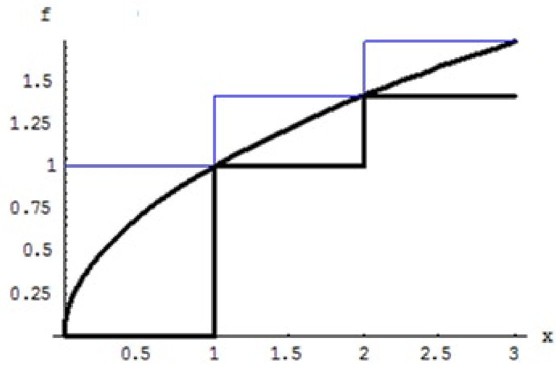


Figure 3.2: Upper and lower bounds of the integral of $f(x) = \sqrt{x}$

we have:

$$\begin{aligned}\int_0^3 h(x)dx &= 1+\sqrt{2} = s_p \\ \int_0^3 g(x)dx &= 1+\sqrt{2}+\sqrt{3} = S_p\end{aligned}$$

We therefore know that $1+\sqrt{2} \leq \int_0^3 f(x)dx \leq 1+\sqrt{2}+\sqrt{3}$.

Considering a refinement of the subdivision (increasing the number of intervals $p$ and decreasing the length of intervals $x_i - x_{i-1}$) leads to the convergence of $s_p$ and $S_p$ (which are adjacent sequences) to the same value when $p \to +\infty$. This common value is the integral of $f$ between $a = 0$ and $b = 3$ and is denoted $\int_0^3 f(x)dx$.

**Definition 121** *A function $f$ defined on $[a;b]$ is Riemann-integrable if the sequences $s_p$ and $S_p$ are adjacent*[1]*. The common limit of these sequences is called the integral of $f$ over $[a;b]$ and denoted $\int_a^b f(x)dx$.*

$s_p$ and $S_p$ are in fact the integrals of two step functions taking respectively the following values, $\{m_i, i = 1, ..., p\}$ for $s_p$ and $\{M_i, i = 1, ..., p\}$ for $S_p$.

[1]see chapter 1.

This remark therefore allows to define integrable functions in a more general way. In fact, looking more closely to the construction method by means of adjacent sequences reveals that continuity of $f$ is not a necessary condition to define integrals. The important point is to be able to "insert" $f$ between two sequences of step functions.

**Definition 122** *A function $f$ defined on $[a;b]$ is integrable if for any $\varepsilon > 0$ there exist two step functions $g$ and $G$ defined on $[a;b]$ such that:*

$$g \leq f \leq G \text{ and } \int_a^b (G(x) - g(x))\, dx < \varepsilon$$

Denote $\mathcal{E}_m$ ($\mathcal{E}_M$) the set of step functions lower (greater) than $f$ and define $I_m$ and $I_M$ by:

$$I_m = \sup_{g \in \mathcal{E}_m} \int_a^b g(x)dx \text{ and } I_M = \inf_{G \in \mathcal{E}_M} \int_a^b G(x)dx$$

These bounds exist because any step function in $\mathcal{E}_m$ is bounded above by a function in $\mathcal{E}_M$ and any element of $\mathcal{E}_M$ is bounded below by an element of $\mathcal{E}_m$. Definition 122 then have the following meaning: $f$ integrable $\Longleftrightarrow$ $I_m = I_M$.

**Definition 123** *The integral of an integrable function $f$ over $[a;b]$ is equal to $I_m = I_M$.*

This definition is very useful in numerical methods. It allows to approximate the value of the integral when the formulation of $f$ is too complicated.

### 3.2.2 Properties of integrals

**Proposition 124** *Any continuous function on $[a;b]$ is Riemann-integrable.*

Proposition 124 is intuitive if one thinks to positive functions[2]. In fact, a continuous function on a closed interval $[a;b]$ is bounded. The integral $\int_a^b f(x)dx$ is then bounded above by $M(b-a)$ if $M$ denotes the maximum of $f$ over $[a;b]$. It is of course bounded below by 0 if $f$ is positive.

The following proposition is the extension of proposition 120 to continuous functions (not necessary step functions).

**Proposition 125** *Let $f$ and $g$ denote two continuous functions defined on $I = [a;b]$.*

*a) For any pair $(\alpha, \beta)$ of real numbers:*

$$\int_a^b (\alpha f + \beta g)(x)dx = \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx$$

*b) For any $c \in I$, we have:*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

*c) If $f \geq g$ then $\int_a^b f(x)dx \geq \int_a^b g(x)dx$*

In part II of the book, devoted to linear algebra, we will see that part (a) of the above proposition means that the set of integrable functions is a vector space.

The intermediate value theorem is a good illustration of the way integrals are built.

**Proposition 126** *1) Let $f$ denote a continuous function defined on $[a;b]$ and $m$, $M$ satisfying $m \leq f(x) \leq M$ for any $x$. There exists $C \in [m;M]$ such that:*

$$C(b-a) = \int_a^b f(x)dx$$

[2] When functions are not everywhere positive, they can always be written as a difference of two positive functions.

*2) If $g$ is integrable, bounded, and keeps the same sign over $[a;b]$, there exists $B \in [m;M]$ such that:*

$$\int_a^b f(x)g(x)dx = B\int_a^b g(x)dx$$

**Proof.** 1) Applying (c) of proposition 125 leads to:

$$m \leq f \leq M \Rightarrow \int_a^b mdx \leq \int_a^b f(x)dx \leq \int_a^b Mdx$$

We then write:

$$m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$$

We can therefore find $C \in [m;M]$ satisfying:

$$C(b-a) = \int_a^b f(x)dx$$

2) Applying once again (c) of proposition 125 leads to:

$$mg \leq fg \leq Mg \Rightarrow \int_a^b mg(x)dx \leq \int_a^b f(x)g(x)dx \leq \int_a^b Mg(x)dx$$

From this inequality we deduce immediately:

$$m\int_a^b g(x)dx \leq \int_a^b f(x)g(x)dx \leq M\int_a^b g(x)dx$$

It follows that there exists $B$ located between $m$ and $M$ and satisfying:

$$\int_a^b f(x)g(x)dx = B\int_a^b g(x)dx$$

■

The interpretation of proposition 126 is easy (if one remembers the step function of figure 3.1). In this simple example, we had $a = 0, b = 20$ and $\int_a^b f(x)dx = 50$. Consequently, $B = 2.5$; it is exactly the average value of $f$ over the interval $[0; 20]$. In this case, the function $f$ can be viewed as a system of weights to calculate the weighted average value of $g$.

**Corollary 127** *If $f$ is continuous on $[a; b]$, there exists $c \in [a; b]$ so that:*

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

As the constant $C$ in the proof of part (1) of proposition 126 is in $[m; M]$ and because $f$ is continuous, there exists $c$ such that $C = f(c)$. The result of the corollary immediately follows from part (1) of proposition 126.

**Proposition 128** *Let $f$ be an integrable function on $[a; b]$ so that $|f|$ is integrable. We have the following inequality:*

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|\, dx$$

This inequality is still valid if $a$ and $b$ are replaced by $c$ and $d$, two numbers in the interval $[a; b]$. It is easy to prove this proposition by applying part (c) of proposition 125 and by noting that $|f| \geq f$ and $|f| \geq -f$. We can then write:

$$\int_a^b |f(x)|\, dx \geq \max\left( \int_a^b f(x)dx; -\int_a^b f(x)dx \right) = \left| \int_a^b f(x)dx \right|$$

**Proposition 129** ***Schwarz inequality***

*Let $f$ and $g$ be two functions defined on $[a; b]$ so that $f^2$ and $g^2$ are integrable. We then have:*

$$\left[ \int_a^b f(t)g(t)dt \right]^2 \leq \int_a^b f^2(t)dt \int_a^b g^2(t)dt$$

**Proof.** We provide the proof of this proposition because the "trick" used in it is useful in a number of circumstances.

For any real number $\lambda$, we can write:

$$\int_a^b (f(t) + \lambda g(t))^2 \, dt = \lambda^2 \int_a^b g^2(t)dt + 2\lambda \int_a^b f(t)g(t)dt + \int_a^b f^2(t)dt$$

This equality comes from the elementary properties of integrals. As a function of $\lambda$, the expression on the right-hand side is a second-degree polynomial which is always positive due to the formulation of the left-hand side. Consequently, the reduced discriminant $\Delta'$ is negative. But this discriminant is equal to:

$$\Delta' = \left[\int_a^b f(t)g(t)dt\right]^2 - \int_a^b g^2(t)dt \int_a^b f^2(t)dt \leq 0$$

It proves the Schwarz inequality. A by-product of this result is the following: when two functions are square-integrable, the product of the two functions is integrable. This property is also very useful in probability theory[3]. ■

[3] Roger. P (2010a), *Probability for Finance, chapter 3.*

### 3.2.3 Primitive integrals

**Definition 130** *Let $f$ and $F$ be two functions defined on $[a;b]$ and taking values in $\mathbb{R}$: $F$ is a primitive integral of $f$ if $F$ is differentiable at any point in $]a;b[$ and $F'(x) = f(x)$ for $x \in \,]a;b[$.*

**Proposition 131** *If $f$ is continuous on $I = [a;b]$, the function $x \to F(x) = \int_a^x f(t)dt$ is a primitive integral of $f$. The function $F$ is thus differentiable and satisfies $F' = f$.*

*If $F$ is a primitive integral of $f$, we have:*

$$\forall (x,y) \in I \times I, \int_x^y f(t)dt = F(y) - F(x)$$

In proposition 131, the essential relationship between $F$ and $f$ is $F' = f$. It turns out that a usual notation for $F$ is $\int f(x)dx$ without specifying the bounds of the integral. In the remainder of the book, $\int f(x)dx$ is a function satisfying $\left(\int f(x)dx\right)' = f$

**Proposition 132** *Let $f$ be an integrable function defined on $I = [a;b]$. If $F$ and $G$ are two primitive integrals of $f$ there exists a constant $c$ so that $F(x) - G(x) = c$ for all $x$ in $]a;b[$.*

This result can be easily illustrated. Consider $F(x) = x^2$ and $G(x) = x^2 + 5$; the derivatives of the two functions are $F'(x) = G'(x) = 2x$. It means that $F$ and $G$ can be primitive integrals for $f(x) = 2x$ because they differ only by a constant (and the derivative of a constant is 0).

### 3.2.4 Primitive integrals of usual functions

**Proposition 133** *Let $f$ be a polynomial defined on $\mathbb{R}$ by:*

$$f(x) = \sum_{k=0}^{n} a_k x^k$$

*We then have:*

$$\int f(x)dx = \sum_{k=0}^{n} \frac{a_k}{k+1} x^{k+1}$$

We just need to note that the derivative of $\frac{x^{k+1}}{k+1}$ is $x^k$ in order to prove this proposition.

**Proposition 134** *1) The natural logarithm, denoted* $\ln()$*, is defined on* $\mathbb{R}_+^*$ *and satisfies* $\ln(1) = 0$*. It is the primitive integral of* $f(x) = \frac{1}{x}$*. We then have, for all* $x \in [1; +\infty[$*:*

$$\ln(x) = \int_1^x \frac{1}{t} dt$$

*2) The exponential is its own primitive integral. In other words, we have:*

$$\int_0^x \exp(t)dt = \exp(x) - \exp(0) = \exp(x) - 1$$

As mentioned before, we can also use the notation based on integrals without bounds to specify these results:

$$\ln(x) = \int \frac{1}{x} dx \text{ and } \exp(x) = \int \exp(t)dt$$

## 3.3 Computations

### 3.3.1 Integration by parts

Remember (or go back to chapter 2) that the derivative of the product of two functions is defined as follows:

$$[h(t)g(t)]' = h'(t)g(t) + h(t)g'(t) \tag{3.3}$$

When solving integration problems, it is often useful to decompose the function to be integrated in a product like $g'(t)f(t)$ and to use integration by parts.

Integrating both sides of equation (3.3) leads to:

$$\int_a^x [g(t)h(t)]'\, dt = \int_a^x g(t)h'(t)dt + \int_a^x g'(t)h(t)dt \tag{3.4}$$

We finally obtain:

$$\int_a^x g'(t)h(t)dt = [g(t)h(t)]_a^x - \int_a^x g(t)h'(t)dt \tag{3.5}$$

where $[g(t)h(t)]_a^x = \int_a^x [g(t)h(t)]'\, dt$ is equal to the difference $g(x)h(x) - g(a)h(a)$.

**Example 135** *Let $f$ denote the Neper logarithm $f(t) = \ln(t)$ and $F$ a primitive integral of $f$ such that:*

$$F(x) - F(\varepsilon) = \int_\varepsilon^x \ln(t)dt$$

*where $x$ and $\varepsilon$ are strictly positive numbers. It is not obvious to guess what is the function whose derivative is $\ln(x)$. But we know that the derivative of $\ln(x)$ is $\frac{1}{x}$. Integration by parts will be the right technique to find the primitive integral of $\ln(x)$. Let us define:*

$$\begin{aligned} h(x) &= \ln(x) \\ g'(x) &= 1 \end{aligned}$$

*so that $\int_\varepsilon^x \ln(t)dt = \int_\varepsilon^x h(t)g'(t)dt$. Applying formula 3.5 and using $h'(t) = \frac{1}{t}$ and $g(t) = t$ leads to:*

$$\begin{aligned} \int_\varepsilon^x \ln(t)dt &= \int_\varepsilon^x g'(t)h(t)dt = [g(t)h(t)]_\varepsilon^x - \int_\varepsilon^x g(t)h'(t)dt \\ &= [t\ln(t)]_\varepsilon^x - \int_\varepsilon^x \frac{1}{t} \times tdt \\ &= x\ln(x) - x - (\varepsilon\ln(\varepsilon) - \varepsilon) \end{aligned}$$

*We know that* $\lim_{\varepsilon \to 0} \varepsilon \ln(\varepsilon) = 0$ *but* $\ln()$ *is only defined on the set of strictly positive numbers. It is then necessary to define a new function* $t \to \phi(t) = t\ln(t)$ *with the assumption* $\phi(0) = 0$*. Doing so allows to obtain the final result:*

$$F(x) = \int_0^x \ln(t)dt = x\ln(x) - x$$

### 3.3.2 Change of variable

In definition 118, assume that $x_i$ can be written $g(y_i)$ where $g$ is a strictly increasing and differentiable function defined on an interval $[c; d]$, taking values in $[a; b]$ and satisfying $g(c) = a$ and $g(d) = b$. We could therefore write:

$$\sum_{i=1}^{p} f(x_{i-1})(x_i - x_{i-1}) = \sum_{i=1}^{p} f(g(y_{i-1}))(g(y_i) - g(y_{i-1}))$$

When $p$ tends to infinity, we can use the approximation:

$$g(y_i) - g(y_{i-1}) = g'(y_{i-1})(y_i - y_{i-1}) \tag{3.6}$$

In other words:

$$\int_a^b f(x)dx = \lim_{p \to +\infty} \sum_{i=1}^{p} f(x_{i-1})(x_i - x_{i-1}) = \lim_{p \to +\infty} \sum_{i=1}^{p} f(g(y_{i-1}))g'(y_{i-1})(y_i - y_{i-1})$$

The right-hand side can also be written as the following integral:

$$\int_c^d f[g(y)]\, g'(y)dy \tag{3.7}$$

In solving problems, the route is generally taken the other way. When you do not know how to calculate an integral, you are going to look for a formulation like expression 3.7. You will then be able to write it as $\int_a^b f(x)dx$, that is an integral you can compute.

One of the essential points in the change of variable technique is that the interval $[c;d]$ must be transformed in the interval $[a;b]$. It assumes implicitly that $g$ is bijective, with $c = g^{-1}(a)$ and $d = g^{-1}(b)$. In most practical problems $g$ is either strictly increasing or strictly decreasing, and of course differentiable.

Finally, when using $g(y) = x$, we need to replace $f(x)$ by $f[g(y)]$ but also $dx$ by $g'(y)dy$.

**Example 136** *In financial models, returns are often assumed to be driven by a gaussian distribution. The density of this probability distribution (when standardized) is* $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. *The following integral is the average value of such a variable when it falls between 0 and 1.*

$$I = \frac{1}{\sqrt{2\pi}} \int_0^1 x \exp(-x^2/2)dx$$

*Let* $y = g(x) = -x^2/2$. *Noticing that* $g'(x) = -x$, *we can write:*

$$I = -\frac{1}{\sqrt{2\pi}} \int_0^1 g'(x) \exp(g(x))dx$$

*It appears that* $g'(x) \exp(g(x))$ *is the derivative of the compound function* $\exp(g(x))$. *It follows:*

$$\begin{aligned}\frac{1}{\sqrt{2\pi}} \int_0^1 x \exp(-x^2/2)dx &= -\frac{1}{\sqrt{2\pi}} \left[\exp(g(x))\right]_0^1 \\ &= -\frac{1}{\sqrt{2\pi}} \left[\exp(y)\right]_0^{-1/2}\end{aligned}$$

$g(x)$ *has been replaced by* $y$ *in this expression. The values of* $y$ *are used to specify the bounds of the integral. When* $x = 0$ *we have* $y = 0$ *and when* $x = 1$, $y = -\frac{1}{2}$. *Finally we get:*

$$\frac{1}{\sqrt{2\pi}} \int_0^1 x \exp(-x^2/2)dx = \frac{1}{\sqrt{2\pi}} \left(1 - \exp(-\frac{1}{2})\right)$$

The two computation methods, integration by parts and change of variable, are very useful to solve practical problems but there does not exist a sure way to know which of the two will be efficient. Training is the best way to improve your skills in integral computation.

## 3.4 Improper integrals

Up to now, we computed integrals of bounded functions on a closed interval $[a; b]$ with $-\infty < a < b < +\infty$. We excluded two situations:

1) One of the boundaries (or both of them) of this interval is infinite.

2) The function $f$ is defined on the semi-open interval $[a; b[$ (or $]a; b]$) and $\lim_{x\to b^-} f(x) = \pm\infty$ (or $\lim_{x\to a^+} f(x) = \pm\infty$).

Taking into account these two situations is possible under some conditions and the corresponding integrals are called improper integrals.

### 3.4.1 Unbounded domains

**Definition 137** *1) Let $f$ be a function defined on $[a; +\infty[$. The* ***improper*** *integral of $f$ is the quantity denoted $\int_a^{+\infty} f(x)dx$ and defined by:*

$$\int_a^{+\infty} f(x)dx = \lim_{b\to+\infty} \int_a^b f(x)dx$$

*2) Let $g$ be a function defined on the $]-\infty; b]$. The improper integral of $g$ is the quantity denoted $\int_{-\infty}^b g(x)dx$ and defined by:*

$$\int_{-\infty}^b g(x)dx = \lim_{a\to-\infty} \int_a^b f(x)dx$$

*3) Let $h$ be a function defined on $\mathbb{R}$. The improper integral of $h$ is the* *quantity denoted* $\int_{-\infty}^{+\infty} h(x)dx$ *and defined by:*

$$\int_{-\infty}^{+\infty} h(x)dx = \int_{-\infty}^{a} h(x)dx + \int_{a}^{+\infty} h(x)dx$$

*where* $a$ *is any real number.*

Let us first assume that $f$ is positive: if we know a primitive integral $F$ of $f$, then the integral $\int_a^{+\infty} f(x)dx$ is convergent if $\lim_{x\to+\infty} F(x)$ is finite. When we do not know any primitive integral of $f$, we need to find a function $g$, which we know a primitive integral of, that represents an upper bound of the function $f$. If $g$ is integrable, then $f$ will also be integrable.

**Proposition 138** *Let* $f$ *and* $g$ *be two positive functions defined on* $[a;+\infty[$ *and satisfying* $f \leq g$*. If the integral of* $g$ *is convergent then the integral of* $f$ *is also convergent:*

$$\int_a^{+\infty} f(x)dx \leq \int_a^{+\infty} g(x)dx$$

*Similarly, if the integral of* $f$ *is divergent then the integral of* $g$ *is also divergent.*

### 3.4.2 Unbounded functions

The second situation to be dealt with is the case of unbounded functions. For example, the function $f(x) = 1/x$ tends to infinity when $x$ tends to 0. The question is to know whether these functions can delimit a finite area and consequently be integrable.

**Definition 139** *1) Let* $f$ *be a function defined on the semi-open interval* $[a;c[$ *and satisfying:*

$$\lim_{x\to c^-} f(x) = +\infty \ (or \ -\infty)$$

*The improper integral of* $f$ *on* $[a;c[$ *is the limit defined by:*

$$\int_a^c f(x)dx = \lim_{b \to c^-} \int_a^b f(x)dx$$

*2) Let* $g$ *be a function defined on the semi-open interval* $]c;b]$ *and satisfying:*

$$\lim_{x \to c^+} g(x) = +\infty \ (or \ -\infty)$$

*The improper integral of* $g$ *on* $]c;b]$ *is the limit defined by:*

$$\int_c^b f(x)dx = \lim_{a \to c^+} \int_a^b f(x)dx$$

If the limits exist, $f$ is said integrable on $[a;c[$ and $g$ integrable on $]c;b]$.

# Chapter 4

# Matrices

This chapter is devoted to the presentation of matrices, their definition and their essential properties. Matrices are essential to develop later on the notions of linear mappings and vector spaces. Moreover, matrices arise naturally when solving multidimensional optimization problems. Nowadays, the calculation rules presented in this chapter are almost never used in a "paper and pencil" approach but in general by means of computer programs or spreadsheets. Especially in finance, people deal with large data sets shaped as matrices, with a given number of rows and columns. The typical dataset is a table of prices or returns. Each line corresponds to one day and each column to one stock. Therefore, matrix calculations are done in spreadsheets and it is essential to know how to manage these calculations. This is the reason why it is necessary to understand, at the theoretical level, how to "combine" matrices through addition, multiplication, inversion, transposition, and so on.

## 4.1 Definitions

**Definition 140** *Let $n$ and $p$ two positive integers. A* ***matrix*** *$A$ of dimensions $(n, p)$, is a table of real numbers with $n$ lines and $p$ columns written as* *follows:*

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & & & \\ . & & & \\ . & & & \\ a_{n1} & \dots & \dots & a_{np} \end{bmatrix}$$

*We also write* $A = (a_{i,j}, i = 1, ..., n; j = 1, ..., p)$ *or simply* $(a_{ij})$ *when there is no confusion about dimensions. To refer to dimensions,* $A$ *is called a* $(n, p)$ *matrix.*

As mentioned in the introduction, the typical matrix in finance contains prices or returns. $a_{ij}$ is then for example the closing price of stock $j$ at date $i$ or the return on stock $j$ over day $i$.

**Definition 141** *1) Let* $A$ *be a* $(n, p)$ *matrix.* $A$ *is called* ***square matrix*** *when* $n = p$.

*2) The coefficients* $a_{ii}, i = 1, ..., n$, *are called the diagonal elements of* $A$. $A$ *matrix* $A$ *is* ***diagonal*** *if all the non-diagonal terms are null, that is if* $a_{ij} = 0$ *for* $i \neq j$.

*3)* $A$ *is a* ***lower triangular matrix*** *if all the terms located above the diagonal are equal to zero, that is if* $a_{ij} = 0$ *for* $i < j$.

*4)* $A$ *is an* ***upper triangular matrix*** *if all the terms located below the diagonal are null, that is if* $a_{ij} = 0$ *for* $i < j$

*5) A matrix* $A$ *is a* ***null matrix*** *if all elements* $a_{ij}$ *are equal to 0.*

If $A$ is a $(n, p)$ matrix, we note $A_{(i)}$ the matrix containing only the $i$-th line of $A$ ; similarly $A^{(j)}$ is the matrix corresponding to the $j$-th column of $A$. $A_{(i)}$ is then a $(1, p)$ matrix and $A^{(j)}$ is a $(n, 1)$ matrix. In the example of stock prices, $A_{(i)}$ would be the set of closing prices on day $i$ and $A^{(j)}$ would be the time-series of stock-$j$ prices.

Triangular matrices appear in financial models when dealing with payoffs of bonds. Assume you hold coupon bearing bonds denoted $A_{(1)}, A_{(2)}, A_{(3)}$ with maturities 1, 2 and 3 years. In a three-year horizon model, $A_{(1)} = (a_{11}, 0, 0)$, $A_{(2)} = (a_{21}, a_{22}, 0)$ and $A_{(3)} = (a_{31}, a_{32}, a_{33})$ where $a_{jj}$ is the reimbursement price of bond $j$ including the last coupon. $a_{ij}$ is the coupon paid by bond $i$ at date $j$. Of course, $a_{ij} = 0$ when $i < j$ since a bond does not pay anything after the maturity date. It turns out that the matrix $A$ is triangular.

**Example 142** *Suppose that you want to build the term structure of interest rates with an horizon of $T = 3$ years. This means that you want to determine the date-0 value of a dollar paid at date 1, at date 2, at date 3, etc. Assume that three bonds are traded with three different maturities (from one to three years) and a common face value equal to \$100. The matrix A below summarizes the future payoffs of the three bonds. The coupon rates are respectively 5%, 4% and 6%.*

$$A = \begin{bmatrix} 105 & 0 & 0 \\ 4 & 104 & 0 \\ 6 & 6 & 106 \end{bmatrix} \tag{4.1}$$

*Assume that the date-0 prices are respectively 100, 99 and 101.9. We observe that we could represent the problem by a function from $\mathbb{R}^3$ to $\mathbb{R}_+$ where*

$$f(105, 0, 0) = 100 \tag{4.2}$$
$$f(4, 104, 0) = 99 \tag{4.3}$$
$$f(6, 6, 106) = 101.9 \tag{4.4}$$

*The following sections of this chapter will allow us to solve this problem and part II of the book illustrates the properties of $f$ which is a linear mapping when there are no arbitrage opportunities on the bond market.*

## 4.2 Elementary algebra on matrices

### 4.2.1 Transposition

Note $\mathcal{M}_{n,p}$ the set of $(n,p)$ matrices.

**Definition 143** *Let $A \in \mathcal{M}_{n,p}$. The **transposed matrix** of $A$ (or simply the **transpose** of $A$) denoted $A'$ (or $A^T$) is the matrix obtained by switching lines and columns of $A$.*

As $A \in \mathcal{M}_{n,p}$, we have $A^T \in \mathcal{M}_{p,n}$.

**Example 144** *Let $A$ be defined by:*

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

*The transpose of $A$ is given by:*

$$A^T = A' = \begin{bmatrix} 2 & 6 & 1 \\ 4 & 3 & 2 \end{bmatrix}$$

*$A$ has three rows and two columns. As mentioned above, the transpose $A'$ has two rows and three columns.*

### 4.2.2 Sum of matrices

**Definition 145** *Let* $A = (a_{i,j})$ *and* $B = (b_{i,j})$ *be two matrices of* $\mathcal{M}_{n,p}$*. The matrix* $A+B$*, sum of* $A$ *and* $B$*, is defined by* $A + B = (a_{ij} + b_{ij})$*, that is:*

$$
\begin{aligned}
A+B &= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & & & \\ . & & & \\ . & & & \\ a_{n1} & \dots & \dots & a_{np} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & & & \\ . & & & \\ . & & & \\ b_{n1} & \dots & \dots & b_{np} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1p}+b_{1p} \\ a_{21}+b_{21} & & & \\ . & & & \\ . & & & \\ a_{n1}+b_{n1} & \dots & \dots & a_{np}+b_{np} \end{bmatrix}
\end{aligned}
$$

**Example 146** *Let* $A$ *and* $B$ *be defined by:*

$$
A = \begin{bmatrix} 2 & 4 \\ 6 & 3 \\ 1 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} -1 & 6 \\ 2 & 1 \\ -3 & 0 \end{bmatrix}
$$

$A+B$ *is given by:*

$$
A+B = \begin{bmatrix} 2-1 & 4+6 \\ 6+2 & 3+1 \\ 1-3 & 2+0 \end{bmatrix} = \begin{bmatrix} 1 & 10 \\ 8 & 4 \\ -2 & 2 \end{bmatrix}
$$

This example and the definitions also illustrate that two matrices can be added only if their dimensions coincide.

**Proposition 147** *Let* $A, B$ *and* $C$ *denote three matrices of* $\mathcal{M}_{n,p}$ *and denote* $O_{n,p}$*, the* $(n,p)$*-matrix containing only zeros. We then have:*

1) $A+B=B+A$
2) $A+O_{n,p}=A$
c) $A+(B+C)=(A+B)+C$

### 4.2.3 Multiplication of a matrix by a real number

**Definition 148** *Let $A=(a_{i,j})$ and $\beta\in\mathbb{R}$. We note $\beta A$ the matrix obtained by multiplying each term of $A$ by the real number $\beta$, that is:*

$$\beta A=\begin{bmatrix}\beta a_{11} & \beta a_{12} & \dots & \beta a_{1p}\\ \beta a_{21} & & & \\ . & & & \\ . & & & \\ \beta a_{n1} & \dots & \dots & \beta a_{np}\end{bmatrix}$$

**Remark 149** *If $\beta=-1$, we get the matrix $\beta A=-A$. Consequently, subtracting a matrix $B$ from a matrix $A$ is equivalent to add $B$ to the matrix $A$ multiplied by $-1$.*

**Proposition 150** *Let $A$ and $B$ be two matrices of $\mathcal{M}_{n,p}$ and $(\alpha, \beta)$ a couple of real numbers. We have:*

*1)* $(\alpha + \beta) A = \alpha A + \beta A$

*2)* $\alpha (A + B) = \alpha A + \alpha B$

*3)* $\alpha (\beta A) = (\alpha\beta) A$

*4)* $0.A = O_{n,p}$

### 4.2.4 Product of matrices

We first start by simple matrices with either one row or one column.

**Definition 151** *Let $A \in \mathcal{M}_{1,n}$ and $B \in \mathcal{M}_{n,1}$. We call product of $A$ and $B$, and we note $AB$, the matrix of $\mathcal{M}_{1,1}$ defined by:*

$$AB = a_{11}b_{11} + a_{12}b_{21} + ... + a_{1n}b_{n1}$$

**Example 152** *Let $A = \begin{bmatrix} 1 & 3 & 1 \end{bmatrix}$ and $B^T = \begin{bmatrix} 3 & 1 & 2 \end{bmatrix}$. The product $AB$ writes:*

$$AB = \begin{bmatrix} 1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = 3 \times 1 + 1 \times 3 + 2 \times 1 = 8$$

*We multiplied a matrix $A(1,3)$ by a matrix $B(3,1)$ to obtain a matrix $(1,1)$ which is simply a real number.*

In the chapter devoted to vector spaces in part II, matrices with one line or one column will be called "vectors". In this framework, the above product is called the inner product of $A$ and $B$.

**Definition 153** *Let $A = (a_{ij}) \in \mathcal{M}_{n,p}$ and $B = (b_{i1}) \in \mathcal{M}_{p,1}$. The product* *of $A$ and $B$, denoted $AB$, is the matrix in $\mathcal{M}_{n,1}$ defined by:*

$$AB = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1p} \\ a_{21} & & & \\ . & & & \\ . & & & \\ a_{n1} & ... & ... & a_{np} \end{bmatrix} \begin{bmatrix} b_{11} \\ ... \\ b_{p1} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + ... + a_{1p}b_{p1} \\ a_{21}b_{11} + a_{22}b_{21} + ... + a_{2p}b_{p1} \\ ... \\ a_{n1}b_{11} + a_{n2}b_{21} + ... + a_{np}b_{p1} \end{bmatrix}$$

$AB$ has $n$ lines (number of lines of $A$) and one column (number of columns of $B$). The $i$-th line of $AB$ is the product of the matrix-line $A_{(i)}$ (the $i$-th line of $A$) by the matrix $B$.

**Example 154** *Let $A$ and $B$ be two matrices defined by:*

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \\ 4 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

*The product $AB$ writes:*

$$AB = \begin{bmatrix} 1 & 2 \\ 0 & 2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \times 1 + 1 \times 2 \\ 4 \times 0 + 1 \times 2 \\ 4 \times 4 + 1 \times 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 17 \end{bmatrix}$$

*Typically, suppose that columns of $A$ represent future payoffs of financial securities in different (3) states of nature and $B$ contains the quantities of these securities held by a given investor. In this simple framework, the product $AB$ provides the possible payoffs of the investor's portfolio in the different states of nature.*

**Definition 155** *Let $A \in \mathcal{M}_{n,p}$ and $B \in \mathcal{M}_{p,m}$. The product $AB$, denoted $C$,* *is a* $(n,m)$ *matrix the generic term of which being written:*

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}$$

The product $AB$ of two matrices $A$ and $B$ is possible only if the number of columns of $A$ is equal to the number of lines of $B$.

**Transpose of a product of matrices**

**Proposition 156** *Let* $A$ *and* $B$ *two matrices so that the product* $AB$ *is possible. We have:*

$$(AB)^T = B^T A^T$$

This proposition also illustrates the fact that, even if $A$ and $B$ are square matrices, the product is not commutative, that is there is no reason to have $AB = BA$.

**Example 157** *Consider the two matrices of the preceding example*

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \\ 4 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

*The product* $B^T A^T$ *is given by:*

$$\begin{aligned} B^T A^T &= \begin{bmatrix} 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 4 \\ 2 & 2 & 1 \end{bmatrix} \\ &= [4 \times 1 + 1 \times 2; 4 \times 0 + 1 \times 2; 4 \times 4 + 1 \times 1] \\ &= \begin{bmatrix} 6 & 2 & 17 \end{bmatrix} = (AB)^T \end{aligned}$$

### 4.2.5 Square matrices and inverse matrices

To simplify, we denote $\mathcal{M}_n$ (instead of $\mathcal{M}_{n,n}$) the set of square $(n, n)$ matrices.

**Definition 158** *1) The identity matrix in $\mathcal{M}_n$ is the $(n,n)$ matrix whose only non-zero elements are the diagonal elements equal to 1. This matrix is noted $I_n$.*

*2) A matrix $A \in \mathcal{M}_n$ is said* ***invertible*** *if there exists a matrix $B \in \mathcal{M}_n$ so that $AB = BA = I_n$. $B$ is then called the inverse matrix of $A$ and noted $A^{-1}$.*

**Example 159** *The so-called Arrow-Debreu securities are financial assets paying \$1 in a given state of nature and 0 in all the other states. It means that in a world with $n$ states of nature, the payoffs of an Arrow-Debreu security write as a column of $I_n$.*

*Suppose that the columns of $A$ in the above definition are the payoffs of securities traded on the financial market. The inverse matrix $A^{-1}$ defines the quantities of securities that should be bought to duplicate the Arrow-Debreu securities. For example let $A$ be defined as follows:*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 2 \\ 4 & 1 & 1 \end{bmatrix} \tag{4.5}$$

*The inverse matrix is*

$$A = \begin{bmatrix} 0 & -\frac{1}{7} & \frac{2}{7} \\ -1 & \frac{11}{7} & -\frac{1}{7} \\ 1 & -1 & 0 \end{bmatrix} \tag{4.6}$$

*Look now at the first column of $A^{-1}$ which is $(0;-1;1)'$. It means that selling one unit of the second asset and buying one unit of the third generates the payoff $(1;0;0)'$. The reader can easily check it is the case.*

$$-\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{4.7}$$

*The two other columns of $A^{-1}$ give the portfolios duplicating the two other Arrow-Debreu securities.*

**Proposition 160** *Let $B, C$ two elements of $\mathcal{M}_{n,p}$ :*

*a) If $A \in \mathcal{M}_n$ is invertible, $AB = AC \Rightarrow B = C$.*

*b) If $D \in \mathcal{M}_p$ is invertible, $BD = CD \Rightarrow B = C$.*

*c) If two matrices $H$ and $V$ of $\mathcal{M}_n$ are invertible, the product $HV$ is invertible and $(HV)^{-1} = V^{-1}H^{-1}$.*

**Proof.** The proof of this proposition is simple but useful.

a) $AB = AC \Rightarrow A^{-1}AB = A^{-1}AC \Rightarrow B = C$ as, by definition of $A^{-1}$ we have $A^{-1}A = I_n$.

b) $BD = CD \Rightarrow BDD^{-1} = CDD^{-1} \Rightarrow B = C$ for the same reason as in (a).

c) On the one hand:

$$HVV^{-1}H^{-1} = H\left(VV^{-1}\right)H^{-1} = HI_nH^{-1} = HH^{-1} = I_n$$

and on the other hand:

$$V^{-1}H^{-1}HV = V^{-1}\left(H^{-1}H\right)V = V^{-1}I_nV = V^{-1}V = I_n$$

These two calculations prove the result, that is $(HV)^{-1} = V^{-1}H^{-1}$. ■

### 4.2.6 Elementary matrices

Elementary matrices are elements of $\mathcal{M}_n$ that allow to perform some specific transformations on elements of $\mathcal{M}_n$, for example:

- permute two lines or two columns of a given matrix $A$ ;

- multiply a line or a column of $A$ by a real number $c$ ;

- Add a line (column) of $A$ to a line (column) of $A$ multiplied by a constant.

In the next subsection we use $(3,3)$ matrices to simplify the presentation but the rules we describe apply to $(n,n)$ matrices.

#### Permutation of lines or columns

Let $\Pi_{12}$ be the matrix defined by:

$$\Pi_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This matrix is obtained by swapping the two first columns (or the two first lines) of the identity matrix $I_3$.

$$\begin{aligned} \Pi_{12}A &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\ A\Pi_{12} &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a_{12} & a_{11} & a_{13} \\ a_{22} & a_{21} & a_{23} \\ a_{32} & a_{31} & a_{33} \end{pmatrix} \end{aligned}$$

$\Pi_{12}A$ is deduced from the matrix $A$ by switching lines 1 and 2. $A\Pi_{12}$ is deduced from $A$ by switching columns 1 and 2.

In the following proposition we note $\Pi_{ij}$ the matrix deduced from the identity matrix by switching columns $i$ and $j$.

**Proposition 161** *The premultiplication by* $\Pi_{ij}$ *of a matrix* $A$ *exchanges lines* $i$ *and* $j$ *of the matrix* $A$.

*The postmultiplication by* $\Pi_{ij}$ *of a matrix* $A$ *exchanges columns* $i$ *and* $j$ *of the matrix* $A$.

### Multiplying a line or a column by a constant

Let us consider the matrix $I_1(c)$ defined by:

$$I_1(c) = \begin{pmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This matrix is deduced from the identity matrix by multiplying the first diagonal term by a constant $c$.

$$\begin{aligned} I_1(c)A &= \begin{pmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} & ca_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\ AI_1(c) &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} ca_{11} & a_{12} & a_{13} \\ ca_{21} & a_{22} & a_{23} \\ ca_{31} & a_{32} & a_{33} \end{pmatrix} \end{aligned}$$

The result obtained with $I_1(c)$ is generalized with $I_k(c)$, the matrix deduced from the identity matrix by multiplying the $k$-th diagonal term by a constant $c$.

**Proposition 162** *The premultiplication of* $A$ *by* $I_k(c)$ *multiplies the* $k$*-th line of* $A$ *by* $c$. *The postmultiplication of* $A$ *by* $I_k(c)$ *multiplies the* $k$*-th column of* $A$ *by* $c$.

**Combining two lines or two columns**

Let us consider the matrix $I_{13}(c)$ defined by:

$$I_{13}(c) = \begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & c \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

As shown in the right-hand side of the above equality, $I_{13}(c)$ is deduced from the identity matrix by adding a constant $c$ to the term located on the first line and the third column.

$$I_{13}(c)A = \begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} + ca_{31} & a_{12} + ca_{32} & a_{13} + ca_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$AI_{13}(c) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} + ca_{11} \\ a_{21} & a_{22} & a_{23} + ca_{21} \\ a_{31} & a_{32} & a_{33} + ca_{31} \end{pmatrix}$$

The result obtained with $I_{13}(c)$ is generalized with $I_{ik}(c)$ in the proposition below.

**Proposition 163** *The premultiplication by $I_{ik}(c)$ of a matrix $A$ adds $c$ times the $k$-th line to the $i$-th line of $A$. The postmultiplication by $I_{ik}(c)$ of a matrix $A$ adds $c$ times the $i$-th column to the $k$-th column of $A$.*

### 4.2.7 Matrix concatenation

**Definition 164** *1) Denote $A$ a matrix with $n$ rows and $p$ columns and $B$ a matrix with $n$ rows and $m$ columns. The concatenate matrix of $A$ and $B$ is* *the $(n, p+m)$-matrix $C$ (also denoted $[A\,|B]$) defined as follows:*

$$C = [A\,|B] = \begin{bmatrix} a_{11} & \dots & a_{1p} & b_{11} & \dots & b_{1m} \\ \dots & & & \dots & & \\ \dots & & & \dots & & \\ a_{n1} & & a_{np} & b_{n1} & & b_{nm} \end{bmatrix}$$

*2) Denote $A$ a matrix with $n$ rows and $p$ columns and $B$ a matrix with $k$ rows and $p$ columns. The concatenate matrix of $A$ and $B$ is the $(n+k, p)$-matrix $D$ (also denoted $\left[\frac{A}{B}\right]$) defined as follows:*

$$D = \left[\frac{A}{B}\right] = \begin{bmatrix} a_{11} & \dots & a_{1p} \\ \dots & & \dots \\ \dots & & \dots \\ a_{n1} & \dots & a_{np} \\ b_{11} & \dots & b_{1p} \\ \dots & & \dots \\ b_{k1} & \dots & b_{kp} \end{bmatrix}$$

Concatenation is often used in solving systems of linear equations, as illustrated in the next section. Of course, the notation $\left[\frac{A}{B}\right]$ for matrices has no relationship with divisions.

## 4.3 Linear equations

### 4.3.1 Introductory example

Consider a financial market where $p$ stocks are traded. In the near future (say tomorrow) there are $n$ possible economic situations. The future price of each stock depends on the economic situation that occurs tomorrow. Suppose an investor wants to consume $c_i$ if the $i$-th situation occurs. We note $\pi_{ij}$ the tomorrow price of stock $j$ in economic state $i$. Which portfolio should the investor buy today to meet her future consumption needs? In other words, which quantities of the $p$ assets should be bought to be able to consume $c_i$ in state $i$?

Denote $x' = (x_1, x_2, ..., x_p)$ these quantities where the prime denotes transposition. If state $i$ occurs, the portfolio value will be:

$$x_1\pi_{i1} + x_2\pi_{i2} + ... + x_p\pi_{ip} = \sum_{j=1}^{p} x_j\pi_{ij}$$

According to the requirements of the investor, the following $n$ equations have to be solved:

$$\begin{aligned} x_1\pi_{11} + x_2\pi_{12} + ...x_p\pi_{1p} &= c_1 \\ x_1\pi_{21} + x_2\pi_{22} + ...x_p\pi_{2p} &= c_2 \\ ..................... & \\ ..................... & \\ x_1\pi_{n1} + x_2\pi_{n2} + ...x_p\pi_{np} &= c_n \end{aligned} \tag{4.8}$$

Denote $M$ the $(n, p)$-matrix of future prices:

$$M = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1p} \\ \pi_{21} & \cdots & \cdots & \pi_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \pi_{n1} & \cdots & \cdots & \pi_{np} \end{bmatrix}$$

and $c$ the $(n, 1)$ matrix of consumptions, $c_i$ being the $i$-th element. The set of equations 4.8 can be summarized in matrix form:

$$Mx = c$$

Remember the dimension rule to multiply matrices. $Mx = c$ is consistent because $M$ has $n$ rows and $p$ columns when $x$ has $p$ rows and one column.

Consequently $Mx$ has $n$ rows and one column, that is the dimensions of $c$.

$Mx = c$ is a system of linear equations because the power of unknowns is equal to 1 (the quantities $x_j$).

The main problems to be solved are the following:

**1) Has a given linear system $Mx = c$ a solution?**

**2) If there is a solution, is it unique?**

**3) If there are solutions, how can we compute them?**

### 4.3.2 A typology of linear systems

In this section we consider the general case of a linear system $Ax = b$, where $A$ is a $(n, m)$ matrix and $b$ a $(n, 1)$ matrix. Can we find a $(m, 1)$ matrix $x$ solving:

$$Ax = b \tag{4.9}$$

**Definition 165** *The linear system $Ax = b$ is:*

- ***square*** *if $A$ is a square matrix (element of $\mathcal{M}_n$).*
- ***a Cramer system*** *if it is a square system with $A$ invertible.*
- ***homogeneous*** *if $b$ is a null matrix*
- ***impossible*** *if there are no solutions.*
- ***indeterminate*** *if there are more than one solution.*

Using inverse matrices allows to get immediately the following proposition.

**Proposition 166** *a) Any homogeneous system of linear equations has a solution.*

*b) Any Cramer system has a unique solution.*

**Proof.** First, it is obvious to see that $x = 0$ is a solution of any homogeneous system.

Second, in a Cramer system, matrix $A$ is invertible; therefore let $x = A^{-1}b$. Remember that $AA^{-1} = I_n$ and that $I_n b = b$. Therefore $x$ solves the system simply because $AA^{-1}b = b$. Moreover $x = A^{-1}b$ is the unique solution. ■

### 4.3.3 Computing solutions

#### A simple example

Assume $n = 2$ and $p = 2$ in the introductory example (section 4.3.1). Two stocks are traded and only two economic situations may occur tomorrow. The possible future prices and desired consumptions are given by:

$$M = \begin{bmatrix} 3 & 1 \\ 4 & 2 \end{bmatrix} \qquad c = \begin{bmatrix} 12 \\ 18 \end{bmatrix}$$

The investor faces the following system:

$$\begin{aligned} 3x_1 + x_2 &= 12 \\ 4x_1 + 2x_2 &= 18 \end{aligned}$$

A simple method to solve such a system is to substitute a given unknown in one of the two equations by a function of the other unknown (obtained using the other equation). For example, from the first equation we obtain:

$$x_2 = 12 - 3x_1 \tag{4.10}$$

Replacing $x_2$ by $12 - 3x_1$ in the second equation leads to:

$$4x_1 + 2 \times (12 - 3x_1) = 18$$

It then follows:

$$-2x_1 + 24 = 18$$

and finally $x_1 = 3$. Replacing $x_1$ by 3 in equation (4.10), gives the complete solution:

$$x_2 = 12 - 3 \times 3 = 3$$

The substitution method is simple for systems with two, and perhaps three, equations. However, it becomes completely unmanageable when many equations and unknowns appear in a large system. Another more systematic method has then to be used.

### The Gauss method

The Gauss method provides an algorithm to solve systems of linear equations. It can be viewed as a generalization of the substitution method illustrated in the above example.

Consider a system of $n$ equations (numbered from 1 to $n$) to be solved. The sequence of steps below leads to a solution when there is one.

- If $a_{11} \neq 0$, replace the first equation by the equivalent one obtained by dividing each term by $a_{11}$. It writes:

$$x_1 + \frac{a_{12}}{a_{11}} x_2 + ... + \frac{a_{1p}}{a_{11}} x_p = \frac{b_1}{a_{11}} \tag{4.11}$$

- To each equation $i$, $i = 2, ..., n$, substract $a_{i1}$ times equation (4.11). For example, equation 2 becomes:

$$a_{21}x_1 + a_{22}x_2 + ... + a_{2p}x_p - a_{21}\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + ... + \frac{a_{1p}}{a_{11}}x_p\right) = b_2 - \frac{a_{21}b_1}{a_{11}}$$

After simplification, we get:

$$\left(a_{22} - \frac{a_{21}a_{12}}{a_{11}}\right)x_2 + \left(a_{23} - \frac{a_{21}a_{13}}{a_{11}}\right)x_3 + ... + \left(a_{2p} - \frac{a_{21}a_{1p}}{a_{11}}\right)x_p = b_2 - \frac{a_{21}b_1}{a_{11}} \tag{4.12}$$

It appears that $x_1$ is not any more in the equation. Repeat the same transformation on the other equations in order to get a system where the first variable $x_1$ only appears in equation (4.11).

- Proceed to the following changes:

$$\begin{aligned} a'_{ij} &= a_{ij} - \frac{a_{i1}a_{1j}}{a_{11}} \\ b'_i &= b_i - \frac{a_{i1}b_1}{a_{11}} \end{aligned}$$

for $i = 2, ..., n$ and $j = 2, ..., p$.

- If $a'_{22} \neq 0$, divide the right and left hand sides of equation (4.12) by $a'_{22}$ to get:

$$x_2 + \frac{a'_{23}}{a'_{22}} x_3 + ... + \frac{a'_{2p}}{a'_{22}} x_p = \frac{b'_2}{a'_{22}} \tag{4.13}$$

- To equations $i$, $i = 3, ..., n$, substract $a'_{i2}$ times equation (4.13). For example, equation 3 becomes:

$$a'_{32}x_2 + ... + a'_{3p}x_p - a'_{32}\left(x_2 + \frac{a'_{23}}{a'_{22}}x_3 + ... + \frac{a'_{2p}}{a'_{22}}x_p\right) = b'_3 - \frac{a'_{32}b'_2}{a'_{22}}$$

The computing process is pursued in order to eliminate one more variable in each equation. When only one variable remains we get easily its value and go back to calculate the values of the other variables.

Finally, different situations may be encountered:

a) What happens when one of the diagonal coefficients ($a_{11}$ at the first stage, $a'_{22}$ at the second, etc.) is equal to 0 ?

If this happens, it is sufficient to change the order of equations. If all such coefficients are zero, the system is indeterminate.

b) If, at the end of the process there remain equations without unknowns, two cases have to considered:

1) if the two members of such equations are different, the system has no solution.

2) if the two members are equal, this equation can be deleted.

c) If several variables remain in the last equation, the number of solutions is infinite.

d) If only one variable remains in the last equation, the value of this variable can be computed and replaced in the other equations to get iteratively the values of all unknowns.

**Example 167** *Consider the matrix $A$ and the vector $b$ defined by:*

$$A = \begin{bmatrix} 1 & 2 & 4 \\ -1 & 3 & 2 \\ 3 & 0 & 2 \\ 2 & 5 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} 2 \\ 1 \\ 6 \\ 4 \end{bmatrix}$$

*The system of linear equations is then:*

$$\begin{aligned} x_1 + 2x_2 + 4x_3 &= 2 \\ -x_1 + 3x_2 + 2x_3 &= 1 \\ 3x_1 + \qquad 2x_3 &= 6 \\ 2x_1 + 5x_2 + x_3 &= 4 \end{aligned}$$

*To apply the Gauss algorithm, it is convenient to concatenate $A$ and $b$ in the following way.*

$$C = [A\,|b] = \begin{bmatrix} 1 & 2 & 4 & 2 \\ -1 & 3 & 2 & 1 \\ 3 & 0 & 2 & 6 \\ 2 & 5 & 1 & 4 \end{bmatrix}$$

*The first coefficient is equal to 1 so the first transform does not change anything. As $a_{21} = -1$, we add the first line to the second. As $a_{31} = 3$, we substract 3 times the first line to the third one and finally we substract twice the first line to the last one.*

*The transformed matrix we denote $C^1$ is equal to:*

$$C^1 = \begin{bmatrix} 1 & 2 & 4 & 2 \\ 0 & 5 & 6 & 3 \\ 0 & -6 & -10 & 0 \\ 0 & 1 & -7 & 0 \end{bmatrix}$$

*This transformation is equivalent to multiply $A$ by the elementary matrices $I_{21}(1)$, $I_{31}(-3)$ and $I_{41}(-2)$, defined as follows (see section 4.2.6) :*

$$I_{21}(1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad I_{31}(-3) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad I_{41}(-2) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}$$

*The same process is repeated on $C^1$ and leads to $C^2$, the pivotal term in this case being $c^1_{22} = 5$.*

$$C^2 = \begin{bmatrix} 1 & 2 & 4 & 2 \\ 0 & 1 & 6/5 & 3/5 \\ 0 & 0 & -14/5 & 18/5 \\ 0 & 0 & -41/5 & -3/5 \end{bmatrix}$$

*The last step is realized by taking the pivotal term $c^2_{33} = -14/5$. We get:*

$$C^3 = \begin{bmatrix} 1 & 2 & 4 & 2 \\ 0 & 1 & 6/5 & 3/5 \\ 0 & 0 & 1 & -9/7 \\ 0 & 0 & 0 & -78/7 \end{bmatrix}$$

*The last line of $C^3$ corresponds to the equation:*

$$0 = -\frac{78}{7}$$

*which of course has no solution. The linear system is therefore impossible to solve.*

# Index