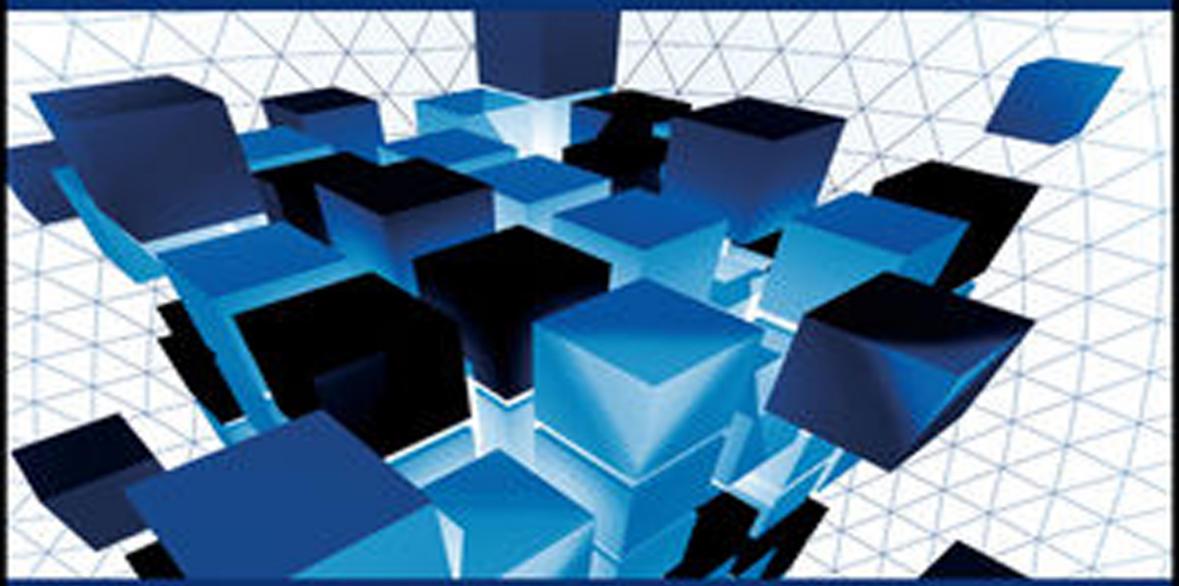
DIGITAL SIGNAL AND IMAGE PROCESSING SERIES

MATRICES AND TENSORS IN SIGNAL PROCESSING SET



Volume 1

# From Algebraic Structures to Tensors

Gérard Favier

ISTE

WILEY

From Algebraic Structures to Tensors

**Matrices and Tensors in Signal Processing Set**

coordinated by
Gérard Favier

Volume 1

# From Algebraic Structures to Tensors

*Edited by*

Gérard Favier

ISTE

WILEY

First published 2019 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.



ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com



Library of Congress Control Number: 2019945792

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-78630-154-3

# Contents

# Preface

This book is part of a collection of four books about matrices and tensors, with applications to signal processing. Although the title of this collection suggests an orientation toward signal processing, the results and methods presented should also be of use to readers of other disciplines.

Writing books on matrices is a real challenge given that so many excellent books on the topic have already been written[1]. How then to stand out from the existing, and which Ariadne's thread to unwind? A way to distinguish oneself was to treat in parallel matrices and tensors. Viewed as extensions of matrices with orders higher than two, the latter have many similarities with matrices, but also important differences in terms of rank, uniqueness of decomposition, as well as potentiality for representing multi-dimensional, multi-modal, and inaccurate data. Moreover, regarding the guiding thread, it consists in presenting structural foundations, then both matrix and tensor decompositions, in addition to related processing methods, finally leading to applications, by means of a presentation as self-contained as possible, and with some originality in the topics being addressed and the way they are treated.

Therefore, in Volume 2, we shall use an index convention generalizing Einstein's summation convention, to write and to demonstrate certain equations involving multi-index quantities, as is the case with matrices and tensors. A chapter will be dedicated to Hadamard, Kronecker, and Khatri–Rao products, which play a very important role in matrix and tensor computations.

After a reminder of main matrix decompositions, including a detailed presentation of the SVD (for singular value decomposition), we shall present different tensor operations, as well as the two main tensor decompositions which will be the basis of both fundamental and applied developments, in the last two volumes. These standard tensor decompositions can be seen as extensions of matrix SVD to tensors of order

1 A list of books, far from exhaustive, is provided in Chapter 1.

higher than two. A few examples of equations for representing signal processing problems will be provided to illustrate the use of such decompositions. A chapter will be devoted to structured matrices. Different properties will be highlighted, and extensions to tensors of order higher than two will be presented. Two other chapters will concern quaternions and quaternionic matrices, on the one hand, and polynomial matrices, on the other hand.

In Volume 3, an overview of several tensor models will be carried out by taking some constraints (structural, linear dependency of factors, sparsity, and non-negativity) into account. Some of these models will be used in Volume 4, for the design of digital communication systems. Tensor trains and tensor networks will also be presented for the representation and analysis of massive data (big data). The algorithmic aspect will be taken into consideration with the presentation of different processing methods.

Volume 4 will mainly focus on tensorial approaches for array processing, wireless digital communications (first point-to-point, then cooperative), modeling and identification of both linear and nonlinear systems, as well as the reconstruction of missing data in data matrices and tensors, the so-called problems of matrix and tensor completion. For these applications, tensor-based models will be more particularly detailed. Monte Carlo simulation results will be provided to illustrate some of the tensorial methods. This will be particularly the case of semi-blind receivers recently developed for wireless communication systems.

Matrices and tensors, and more generally linear algebra and multilinear algebra, are at the same time exciting, extensive, and fundamental topics equally important for teaching and researching as for applications. It is worth noting here that the choices made for the content of the books of this collection have not been guided by educational programs, which explains some gaps compared to standard algebra treaties. The guiding thread has been rather to present the definitions, properties, concepts and results necessary for a good understanding of processing methods and applications considered in these books. In addition to the great diversity of topics, another difficulty resided in the order in which they should be addressed, due to the fact that a lot of topics overlap, certain notions or/and some results being sometimes used before they have been defined or/and demonstrated, which requires the reader to be referred to sections or chapters that follow.

Four particularities should be highlighted. The first relates to the close relationship between some of the topics being addressed, certain methods presented and recent research results, particularly with regard to tensorial approaches for signal processing. The second reflects the will to situate the results stated in their historical context, using some biographical information on certain authors being cited, as well as lists of references comprehensive enough to deepen specific results, and also to extend the biographical sources provided. This has motivated the introductory chapter entitled “Historical elements of matrices and tensors.”

The last two characteristics concern the presentation and illustration of properties and methods under consideration. Some will be provided without demonstration because of their simplicity or availability in numerous books thereabout. Others will be demonstrated, either for pedagogical reasons, since their knowledge should allow for better understanding the results being demonstrated, or because of the difficulty to find them in the literature, or still due to the originality of the proposed demonstrations as it will be the case, for example, of those making use of the index convention. The use of many tables should also be noted with the purpose of recalling key results while presenting them in a synthetic and comparative manner.

Finally, numerous examples will be provided to illustrate certain definitions, properties, decompositions, and methods presented. This will be particularly the case for the fourth book dedicated to applications of tensorial tools, which has been my main source of motivation. After 15 years of works dedicated to research (pioneering for some), aiming to use tensor decompositions for modeling and identifying nonlinear dynamical systems, and for designing wireless communication systems based on new tensor models, it seemed to me useful to share this experience and this scientific path for trying to make tensor tools as accessible as possible and to motivate new applications based on tensor approaches.

This first book, whose content is described below, provides an introduction to matrices and tensors based on the structures of vector spaces and tensor spaces, along with the presentation of fundamental concepts and results. In the first part (Chapters 2 and 3), a refresher of the mathematical bases related to classical algebraic structures is presented, by way of bringing forward a growing complexity of the structures under consideration, ranging from monoids to vector spaces, and to algebras. The notions of norm, inner product, and Hilbert basis are detailed in order to introduce Banach and Hilbert spaces. The Hilbertian approach, which is fundamental for signal processing, is illustrated based on two methods widely employed for signal representation and analysis, as well as for function approximation, namely, Fourier and orthogonal polynomial series.

Chapter 4 is dedicated to matrix algebra. The notions of fundamental subspaces associated with a matrix, rank, determinant, inverse, auto-inverse, generalized inverse, and pseudo-inverse are described therein. Matrix representations of linear and bilinear/sesquilinear maps are established. The effect of a change of basis is studied, leading to the definition of equivalent, similar, and congruent matrices. The notions of eigenvalue and eigenvector are then defined, ending with matrix eigendecomposition, and in some cases, with its diagonalization, which are topics to be covered in Volume 2. The case of certain structured matrices, such as symmetric/hermitian matrices and orthogonal/unitary matrices, is more specifically considered. The interpretation of eigenvalues as extrema of the Rayleigh quotient is presented, before introducing the notion of generalized eigenvalues.

In Chapter 5, we consider partitioned matrices. This type of structure is inherent to matrix products in general, and Kronecker and Khatri–Rao products in particular.

Partitioned matrices corresponding to block-diagonal and block-triangular matrices, as well as to Jordan forms are described. Next, transposition/conjugate transposition, vectorization, addition and multiplication operations, as well as Hadamard and Kronecker products, are presented for partitioned matrices. Elementary operations and associated matrices allowing the partitioned matrices to be decomposed are detailed. These operations are then utilized for block-triangularization, block-diagonalization, and block-inversion of partitioned matrices. The matrix inversion lemma, which is widely used in signal processing, is deduced from block-inversion formulae. This lemma is used to demonstrate a few inversion formulae very often encountered in calculations. Fundamental results on generalized inverse, determinant, and rank of partitioned matrices are presented. The Levinson algorithm is demonstrated using the formula for inverting a partitioned square matrix, recursively with respect to the matrix order. This algorithm, which is one of the most famous in signal processing, allows to efficiently solve the problem of parameter estimation of autoregressive (AR) models and linear predictors, recursively with respect to the order of the AR model and of the predictor, respectively. To illustrate the results of Chapter 3 relatively to orthogonal projection, it is shown that forward and backward linear predictors, optimal in the sense of the MMSE (minimum mean squared error), can be interpreted in terms of orthogonal projectors on subspaces of the Hilbert space of the second-order stationary random signals.

In Chapter 6, hypermatrices and tensors are introduced in close connection with multilinear maps and multilinear forms. Hypermatrix vector spaces are first defined, along with operations such as inner product and contraction of hypermatrices – the particular case of the n-mode hypermatrix-matrix product being considered in more detail. Hypermatrices associated with multilinear forms and maps are highlighted, and symmetric hypermatrices are introduced through the definition of symmetric multilinear forms. Then, tensors of order $N > 2$ are defined in a formal way as elements of a tensor space, i.e., a tensor product of N vector spaces. The effect of changes to the tensor space on the coordinate hypermatrix of a tensor are analyzed. In addition, some attributes of the tensor product are described, with a focus on the so-called universal property. Following this, the notions of a rank based on the canonical polyadic decomposition (CPD) of a tensor are introduced, as well as the ranking of a tensor's eigenvalues and singular values. These highlight the similarities and the differences between matrices, and tensors of order greater than two. Finally, the concept of tensor unfolding is illustrated via the definition of isomorphisms of tensor spaces.

I want to thank my colleagues Sylvie Icart and Vicente Zarzoso for their review of some chapters and Henrique de Morais Goulart, who co-authored Chapter 4.

Gérard FAVIER
August 2019
favier@i3s.unice.fr

# 1

# Historical Elements of Matrices and Tensors

The objective of this introduction is by no means to outline a rigorous and comprehensive historical background of the theory of matrices and tensors. Such a historical record should be the work of a historian of mathematics and would require thorough bibliographical research, including reading the original publications of authors cited to analyze and reconstruct the progress of mathematical thinking throughout years and collaborations. A very interesting illustration of this type of analysis is provided, for example, in the form of a "representation of research networks"[1], over the period 1880–1907, in which are identified the interactions and influences of some mathematicians, such as James Joseph Sylvester (1814–1897), Karl Theodor Weierstrass (1815–1897), Arthur Cayley (1821–1895), Leopold Kronecker (1823–1891), Ferdinand Georg Frobenius (1849–1917), or Eduard Weyr (1852–1903), with respect to the theory of matrices, the theory of numbers (quaternions, hypercomplex numbers), bilinear forms, and algebraic structures.

Our modest goal here is to locate in time the contributions of a few mathematicians and physicists[2] who have laid the foundations for the theory of matrices and tensors, and to whom we will refer later in our presentation. This choice is necessarily very incomplete.

1 F. Brechenmacher, "Les matrices : formes de représentation et pratiques opératoires (1850–1930)", Culture MATH - Expert site ENS Ulm / DESCO, December 20, 2006.

2 For more information on the mathematicians cited in this introduction, refer to the document "Biographies de mathématiciens célèbres", by Johan Mathieu, 2008, and the remarkable site Mac Tutor History of Mathematics Archive (http://www-history.mcs.st-andrews.ac.uk) of the University of St. Andrews, in Scotland, which contains a very large number of biographies of mathematicians.

*From Algebraic Structures to Tensors,*
First Edition. Edited by Gérard Favier.

The first studies of determinants that preceded those of matrices were conducted independently by the Japanese mathematician Seki Kowa (1642–1708) and the German mathematician Gottfried Leibniz (1646–1716), and then by the Scottish mathematician Colin Maclaurin (1698–1746) for solving $2 \times 2$ and $3 \times 3$ systems of linear equations. These works were then generalized by the Swiss mathematician Gabriel Cramer (1704–1752) for the resolution of $n \times n$ systems, leading, in 1750, to the famous formulae that bear his name, whose demonstration is due to Augustin-Louis Cauchy (1789–1857).

In 1772, Théophile Alexandre Vandermonde (1735–1796) defined the notion of determinant, and Pierre-Simon Laplace (1749–1827) formulated the computation of determinants by means of an expansion according to a row or a column, an expansion which will be presented in section 4.11.1. In 1773, Joseph-Louis Lagrange (1736–1813) discovered the link between the calculation of determinants and that of volumes. In 1812, Cauchy used, for the first time, the determinant in the sense that it has today, and he established the formula for the determinant of the product of two rectangular matrices, a formula which was found independently by Jacques Binet (1786–1856), and which is called nowadays the Binet–Cauchy formula.

In 1810, Johann Carl Friedrich Gauss (1777–1855) introduced a notation using a table, similar to matrix notation, to write a $3 \times 3$ system of linear equations, and he proposed the elimination method, known as Gauss elimination through pivoting, to solve it. This method, also known as Gauss–Jordan elimination method, was in fact known to Chinese mathematicians (first century). It was presented in a modern form, by Gauss, when he developed the least squares method, first published by Adrien-Marie Legendre (1752–1833), in 1805.

Several determinants of special matrices are designated by the names of their authors, such as Vandermonde's, Cauchy's, Hilbert's, and Sylvester's determinants. The latter of whom used the word "matrix" for the first time in 1850, to designate a rectangular table of numbers. The presentation of the determinant of an $n$th-order square matrix as an alternating $n$-linear function of its $n$ column vectors is due to Weierstrass and Kronecker, at the end of the 19th century.

The foundations of the theory of matrices were laid in the 19th century around the following topics: determinants for solving systems of linear equations, representation of linear transformations and quadratic forms (a topic which will be addressed in detail in Chapter 4), matrix decompositions and reductions to canonical forms, that is to say, diagonal or triangular forms such as the Jordan (1838–1922) normal form with Jordan blocks on the diagonal, introduced by Weierstrass, the block-triangular form of Schur (1875–1941), or the Frobenius normal form that is a block-diagonal matrix, whose blocks are companion matrices.

A history of the theory of matrices in the 19th century was published by Thomas Hawkins[3] in 1974, highlighting, in particular, the contributions of the British mathematician Arthur Cayley, seen by historians as one of the founders of the theory of matrices. Cayley laid the foundations of the classical theory of determinants[4] in 1843. He then developed matrix computation[5] by defining certain matrix operations as the product of two matrices, the transposition of the product of two matrices, and the inversion of a $3 \times 3$ matrix using cofactors, and by establishing different properties of matrices, including, namely, the famous Cayley–Hamilton theorem which states that every square matrix satisfies its characteristic equation. This result highlighted for the fourth order by William Rowan Hamilton (1805–1865), in 1853, for the calculation of the inverse of a quaternion, was stated in the general case by Cayley in 1857, but the demonstration for any arbitrary order is due to Frobenius, in 1878.

An important part of the theory of matrices concerns the spectral theory, namely, the notions of eigenvalue and characteristic polynomial. Directly related to the integration of systems of linear differential equations, this theory has its origins in physics, and more particularly in celestial mechanics for the study of the orbits of planets, conducted in the 18th century by mathematicians, physicists, and astronomers such as Lagrange and Laplace, then in the 19th century by Cauchy, Weierstrass, Kronecker, and Jordan.

The names of certain matrices and associated determinants are those of the mathematicians who have introduced them. This is the case, for example, for Alexandre Théophile Vandermonde (1735–1796) who gave his name to a matrix whose elements on each row (or each column) form a geometric progression and whose determinant is a polynomial. It is also the case for Carl Jacobi (1804–1851) and Ludwig Otto Hesse (1811–1874), for Jacobian and Hessian matrices, namely, the matrices of first- and second-order partial derivatives of a vector function, whose determinants are called Jacobian and Hessian, respectively. The same is true for the Laplacian matrix or Laplace matrix, which is used to represent a graph. We can also mention Charles Hermite (1822–1901) for Hermitian matrices, related to the so-called Hermitian forms (see section 4.15). Specific matrices such as Fourier (1768–1830) and Hadamard (1865–1963) matrices are directly related to the transforms of the same name. Similarly, Householder (1904–1993) and Givens (1910–1993) matrices are associated with transformations corresponding to reflections and rotations, respectively. The so-called structured matrices, such as Hankel (1839–1873) and Toeplitz (1881–1943) matrices, play a very important role in signal processing.

3 Thomas Hawkins, "The theory of matrices in the 19th century", *Proceedings of the International Congress of Mathematicians, Vancouver*, 1974.

4 Arthur Cayley, "On a theory of determinants", *Cambridge Philosophical Society* 8, l–16, 1843.

5 Arthur Cayley, "A memoir on the theory of matrices", *Philosophical Transactions of the Royal Society of London* 148, 17–37, 1858.

Matrix decompositions are widely used in numerical analysis, especially to solve systems of equations using the method of least squares. This is the case, for example, of EVD (eigenvalue decomposition), SVD (singular value decomposition), LU, QR, UD, Cholesky (1875–1918), and Schur (1875–1941) decompositions, which will be presented in Volume 2.

Just as matrices and matrix computation play a fundamental role in linear algebra, tensors and tensor computation are at the origin of multilinear algebra. It was in the 19th century that tensor analysis first appeared, along with the works of German mathematicians Georg Friedrich Bernhard Riemann[6] (1826–1866) and Elwin Bruno Christoffel (1829–1900) in geometry (non-Euclidean), introducing the index notation and notions of metric, manifold, geodesic, curved space, curvature tensor, which gave rise to what is today called Riemannian geometry and differential geometry.

It was the Italian mathematician Gregorio Ricci-Curbastro (1853–1925) with his student Tullio Levi-Civita (1873–1941) who were the founders of the tensor calculus, then called absolute differential calculus[7], with the introduction of the notion of covariant and contravariant components, which was used by Albert Einstein (1879–1955) in his theory of general relativity, in 1915.

Tensor calculus originates from the study of the invariance of quadratic forms under the effect of a change of coordinates and, more generally, from the theory of invariants initiated by Cayley[8], with the introduction of the notion of hyperdeterminant which generalizes matrix determinants to hypermatrices. Refer to the article by Crilly[9] for an overview of the contribution of Cayley on the invariant theory. This theory was developed by Jordan and Kronecker and involved controversy[10] between these two authors, then continued by David Hilbert (1862–1943), Elie Joseph Cartan (1869–1951), and Hermann Klaus Hugo Weyl (1885–1955), for algebraic forms

6 A detailed analysis of Riemann's contributions to tensor analysis has been made by Ruth Farwell and Christopher Knee, "The missing link: Riemann's Commentatio, differential geometry and tensor analysis", *Historia Mathematica* 17, 223–255, 1990.

7 G. Ricci and T. Levi-Civita, "Méthodes de calcul différentiel absolu et leurs applications", *Mathematische Annalen* 54, 125–201, 1900.

8 A. Cayley, "On the theory of linear transformations", *Cambridge Journal of Mathematics* 4, 193–209, 1845. A. Cayley, "On linear transformations", *Cambridge and Dublin Mathematical Journal* 1, 104–122, 1846.

9 T. Crilly, "The rise of Cayley's invariant theory (1841–1862)", *Historica Mathematica* 13, 241–254, 1986.

10 F. Brechenmacher, "La controverse de 1874 entre Camille Jordan et Leopold Kronecker: Histoire du théorème de Jordan de la décomposition matricielle (1870–1930)", *Revue d'histoire des Mathématiques, Society Math De France* 2, no. 13, 187–257, 2008 (hal-00142790v2).

(or homogeneous polynomials), or for symmetric tensors[11]. A historical review of the theory of invariants was made by Dieudonné and Carrell[12].

This property of invariance vis-à-vis the coordinate system characterizes the laws of physics and, thus, mathematical models of physics. This explains that tensor calculus is one of the fundamental mathematical tools for writing and studying equations that govern physical phenomena. This is the case, for example, in general relativity, in continuum mechanics, for the theory of elastic deformations, in electromagnetism, thermodynamics, and so on.

The word tensor was introduced by the German physicist Woldemar Voigt (1850–1919), in 1899, for the geometric representation of tensions (or pressures) and deformations in a body, in the areas of elasticity and crystallography. Note that the word tensor was introduced independently by the Irish mathematician, physicist and astronomer William Rowan Hamilton (1805–1865), in 1846, to designate the modulus of a quaternion[13].

As we have just seen in this brief historical overview, tensor calculus was used initially in geometry and to describe physical phenomena using tensor fields, facilitating the application of differential operators (gradient, divergence, rotational, and Laplacian) to tensor fields[14].

Thus, we define the electromagnetic tensor (or Maxwell's (1831–1879) tensor) describing the structure of the electromagnetic field, the Cauchy stress tensor, and the deformation tensor (or Green–Lagrange deformation tensor), in continuum mechanics, and the fourth-order curvature tensor (or Riemann–Christoffel tensor) and the third-order torsion tensor (or Cartan tensor[15]) in differential geometry.

---

11 M. Olive, B. Kolev, and N. Auffray, "Espace de tenseurs et théorie classique des invariants", *21ème Congrès Français de Mécanique*, Bordeaux, France, 2013 (hal-00827406).

12 J. A. Dieudonné and J. B. Carrell, *Invariant Theory, Old and New*, Academic Press, 1971.

13 See page 9 in E. Sarrau, *Notions sur la théorie des quaternions*, Gauthiers-Villars, Paris, 1889, http://rcin.org.pl/Content/13490.

14 The notion of tensor field is associated with physical quantities that may depend on both spatial coordinates and time. These variable geometric quantities define differentiable functions on a domain of the physical space. Tensor fields are used in differential geometry, in algebraic geometry, general relativity, and in many other areas of mathematics and physics. The concept of tensor field generalizes that of vector field.

15 E. Cartan, "Sur une généralisation de la notion de courbure de Riemann et les espaces à torsion", *Comptes rendus de l'Académie des Sciences* 174, 593–595, 1922. Elie Joseph Cartan (1869–1951), French mathematician and student of Jules Henri Poincaré (1854–1912) and Charles Hermite (1822–1901) at the Ecole Normale Supérieure. He brought major contributions concerning the theory of Lie groups, differential geometry, Riemannian geometry, orthogonal polynomials, and elliptic functions. He discovered spinors, in 1913, as part of his work on the

After their introduction as computational and representation tools in physics and geometry, tensors have been the subject of mathematical developments related to polyadic decomposition (Hitchcock 1927) aiming to generalize dyadic decompositions, that is to say, matrix decompositions such as SVD.

Then emerged their applications as tools for the analysis of three-dimensional data generalizing matrix analysis to sets of matrices, viewed as arrays of data characterized by three indices. We can mention here the works of pioneers in factor analysis by Cattell[16] and Tucker[17] in psychometrics (Cattell 1944; Tucker 1966), and Harshman[18] in phonetics (Harshman 1970) who have introduced Tucker's and PARAFAC (parallel factors) decompositions. This last one was proposed independently by Carroll and Chang (1970), under the name of canonical decomposition (CANDECOMP), following the publication of an article by Wold (1966), with the objective to generalize the (Eckart and Young 1936) decomposition, that is, SVD, to arrays of order higher than two. This decomposition was then called CP (for CANDECOMP/PARAFAC) by Kiers (2000). For an overview of tensor methods applied to data analysis, the reader should consult the books by Coppi and Bolasco (1989) and Kroonenberg (2008).

From the early 1990s, tensor analysis, also called multi-way analysis, has also been widely used in chemistry, and more specifically in chemometrics (Bro 1997). Refer to, for example, the book by Smilde *et al.* (2004) for a description of various applications in chemistry.

In parallel, at the end of the 1980s, statistic "objects," such as moments and cumulants of order higher than two, have naturally emerged as tensors (McCullagh 1987). Tensor-based applications were then developed in signal processing for solving the problem of blind source separation using cumulants (Cardoso 1990, 1991; Cardoso and Comon 1990). The book by Comon and Jutten (2010) outlines an overview of methods for blind source separation.

In the early 2000s, tensors were used for modeling digital communication systems (Sidiropoulos *et al.* 2000a), for array processing (Sidiropoulos *et al.* 2000b), for multi-dimensional harmonics recovery (Haardt *et al.* 2008; Jiang *et al.* 2001; Sidiropoulos 2001), and for image processing, more specifically for face recognition

representations of groups. Like tensor calculus, spinor calculus plays a major role in quantum physics. His name is associated with Albert Einstein (1879–1955) for the classical theory of gravitation that relies on the model of general relativity.

16 Raymond Cattell (1905–1998), Anglo-American psychologist who used factorial analysis for the study of personality with applications to psychotherapy.

17 Ledyard Tucker (1910–2004), American mathematician, expert in statistics and psychology, and more particularly known for tensor decomposition which bears his name.

18 Richard Harshman (1943–2008), an expert in psychometrics and father of three-dimensional PARAFAC analysis which is the most widely used tensor decomposition in applications.

(Vasilescu and Terzopoulos 2002). The field of wireless communication systems has then given rise to a large number of tensor models (da Costa *et al.* 2018; de Almeida and Favier 2013; de Almeida *et al.* 2008; Favier *et al.* 2012a; Favier and de Almeida 2014b; Favier *et al.* 2016). These models will be covered in a chapter of Volume 3. Tensors have also been used for modeling and parameter estimation of dynamic systems both linear (Fernandes *et al.* 2008, 2009a) and nonlinear, such as Volterra systems (Favier and Bouilloc 2009a, 2009b, 2010) or Wiener-Hammerstein systems (Favier and Kibangou 2009a, 2009b; Favier *et al.* 2012b; Kibangou and Favier 2008, 2009, 2010), and for modeling and estimating nonlinear communication channels (Bouilloc and Favier 2012; Fernandes *et al.* 2009b, 2011; Kibangou and Favier 2007). These different tensor-based applications in signal processing will be addressed in Volume 4.

Many applications of tensors also concern speech processing (Nion *et al.* 2010), MIMO radar (Nion and Sidiropoulos 2010), and biomedical signal processing, particularly for electroencephalography (EEG) (Cong *et al.* 2015; de Vos *et al.* 2007; Hunyadi *et al.* 2016), and electrocardiography (ECG) signals (Padhy *et al.* 2018), magnetic resonance imaging (MRI) (Schultz *et al.* 2014), or hyperspectral imaging (Bourennane *et al.* 2010; Velasco-Forero and Angulo 2013), among many others. Today, tensors viewed as multi-index tables are used in many areas of application for the representation, mining, analysis, and fusion of multi-dimensional and multi-modal data (Acar and Yener 2009; Cichocki 2013; Lahat *et al.* 2015; Morup 2011).

A very large number of books address linear algebra and matrix calculus, for example: Gantmacher (1959), Greub (1967), Bellman (1970), Strang (1980), Horn and Johnson (1985, 1991), Lancaster and Tismenetsky (1985), Noble and Daniel (1988), Barnett (1990), Rotella and Borne (1995), Golub and Van Loan (1996), Lütkepohl (1996), Cullen (1997), Zhang (1999), Meyer (2000), Lascaux and Théodor (2000), Serre (2002), Abadir and Magnus (2005), Bernstein (2005), Gourdon (2009), Grifone (2011), and Aubry (2012).

For multilinear algebra and tensor calculus, there are much less reference books, for example: Greub (1978), McCullagh (1987), Coppi and Bolasco (1989), Smilde *et al.* (2004), Kroonenberg (2008), Cichocki *et al.* (2009), and Hackbusch (2012). For an introduction to multilinear algebra and tensors, see Ph.D. theses by de Lathauwer (1997) and Bro (1998). The following synthesis articles can also be consulted: (Bro 1997; Cichocki *et al.* 2015; Comon 2014; Favier and de Almeida 2014a; Kolda and Bader 2009; Lu *et al.* 2011; Papalexakis *et al.* 2016; Sidiropoulos *et al.* 2017).

# 2

# Algebraic Structures

## 2.1. A few historical elements

We make here a brief historical note concerning algebraic structures. The notion of structure plays a fundamental role in mathematics. In a treatise entitled *Eléments de mathématique*, comprising 11 books, Nicolas Bourbaki[1] distinguishes three main types of structures: algebraic structures, ordered structures that equip sets with an order relation, and topological structures equipping sets with a topology that allows the definition of topological concepts such as open sets, neighborhood, convergence, and continuity. Some structures are mixed, that is, they combine several of the three basic structures. That is the case, for instance, of Banach and Hilbert spaces which combine the vector space structure with the notions of norm and inner product, that is, a topology.

Algebraic structures endow sets with laws of composition governing operations between elements of a same set or between elements of two distinct sets. These composition laws known as internal and external laws, respectively, exhibit certain properties such as associativity, commutativity, and distributivity, with the existence (or not) of a symmetric for each element, and of a neutral element. Algebraic structures make it possible to characterize, in particular, sets of numbers, polynomials, matrices, and functions. The study of these structures (groups, rings, fields, vector spaces, etc.) and their relationships is the primary purpose of general algebra, also called abstract algebra. A reminder of the basic algebraic structures will be carried out in this chapter.

The vector spaces gave rise to linear algebra for the resolution of systems of linear equations and the study of linear maps (also called linear mappings, or linear transformations). Linear algebra is closely related to the theory of matrices and matrix algebra, of which an introduction will be made in Chapter 4.

1 Nicolas Bourbaki is the pseudonym of a group of French mathematicians formed in 1935.

*From Algebraic Structures to Tensors,*
First Edition. Edited by Gérard Favier.

Multilinear algebra extends linear algebra to the study of multilinear maps, through the notions of tensor space and tensor, which will be introduced in Chapter 6.

Although the resolution of (first- and second-degree) equations can be traced to the Babylonians[2] (about 2000 BC, according to Babylonian tables), then to the Greeks (300 BC), to the Chinese (200 BC), and to the Indians (6th century), algebra as a discipline emerged in the Arab-Muslim world, during the 8th century. It gained momentum in the West, in the 16th century, with the resolution of algebraic (or polynomial) equations, first with the works of Italian mathematicians Tartaglia (1500–1557) and Jérôme Cardan (1501–1576) for cubic equations, whose first resolution formula is attributed to Scipione del Ferro (1465–1526) and Lodovico Ferrari (1522–1565) for quartic equations. The work of François Viète (1540–1603) then René Descartes (1596–1650) can also be mentioned, for the introduction of the notation making use of letters to designate unknowns in equations, and the use of superscripts to designate powers.

A fundamental structure, linked to the notion of symmetry, is that of the group, which gave rise to the theory of groups, issued from the theory of algebraic equations and the study of arithmetic properties of algebraic numbers, at the end of the 18th century, and of geometry, at the beginning of the 19th century. We may cite, for example, Joseph-Louis Lagrange (1736–1813), Niels Abel (1802–1829), and Evariste Galois (1811–1832), for the study of algebraic equations, the works of Carl Friedrich Gauss (1777–1855) on the arithmetic theory of quadratic forms, and those of Felix Klein (1849–1925) and Hermann Weyl (1885–1955) in non-Euclidean geometry. We can also mention the works of Marie Ennemond Camille Jordan (1838–1922) on the general linear group, that is, the group of invertible square matrices, and on the Galois theory. In 1870, he published a treatise on the theory of groups, including the reduced form of a matrix, known as Jordan form, for which he received the Poncelet prize of the Academy of Sciences.

Groups involve a single binary operation.

The algebraic structure of ring was proposed by David Hilbert (1862–1943) and Emmy Noether (1882–1935), while that of field was introduced independently by Leopold Kronecker (1823–1891) and Richard Dedekind (1831–1916). In 1893, Heinrich Weber (1842–1913) presented the first axiomatization[3] of commutative fields, completed in 1910 by Ernst Steinitz (1871–1928). Field extensions led to

2 Arnaud Beauville, “Histoire des équations algébriques”, https://math.unice.fr/beauvill/pubs/Equations.pdf and http://www.maths-et-tiques.fr/index.php/Histoire-des-maths/Nombres/Histoire-de-l-algebre.

3 Used for developing a scientific theory, the axiomatic method is based on a set of propositions called axioms. Its founders are Greek mathematicians, which include Euclid and Archimedes (c. 300 BC) as the most famous, for their work in geometry (Euclidean) and arithmetic. It was

the Galois theory, with the initial aim to solve algebraic equations. In 1843, a first example of non-commutative field was introduced by William Rowan Hamilton (1805–1865), with quaternions.

Rings and fields are algebraic structures involving two binary operations, generally called addition and multiplication.

The underlying structure to the study of linear systems, and more generally to linear algebra, is that of vector space (v.s.) introduced by Hermann Grassmann (1809–1877), then axiomatically formalized by Giuseppe Peano, with the introduction of the notion of $\mathbb{R}$-vector space, at the end of the 19th century. German mathematicians David Hilbert (1862–1943), Otto Toeplitz (1881–1940), Hilbert's student, and Stefan Banach (1892–1945) were the ones who extended vector spaces to spaces of infinite dimension, called Hilbert spaces and Banach spaces (or normed vector spaces (n.v.s.)).

The study of systems of linear equations and linear transformations, which is closely linked to that of matrices, led to the concepts of linear independence, basis, dimension, rank, determinant, and eigenvalues, which will be considered in this chapter as well as in Chapters 4 and 6. In Chapter 3, we shall see that by equipping v.s. with a norm and an inner product, n.v.s. can be obtained on the one hand, and pre-Hilbertian spaces on the other. The concept of distance allows for the definition of metric spaces. Norms and distances are used for studying the convergence of sequences in the context of Banach and Hilbert spaces, of infinite dimension, which will also be addressed in the next chapter. The extension of linear spaces to multilinear spaces will be considered in this chapter and in Chapter 6 through the introduction of the tensor product, with generalization of matrices to hypermatrices and tensors of order higher than two.

## 2.2. Chapter summary

The objective of this chapter is to carry out an overview of the main algebraic structures, while recalling definitions and results that will be useful for other chapters. First, we recall some results related to sets and maps, and we then present the definitions and properties of internal and external composition laws on a set. Various algebraic structures are then detailed: groups, rings, fields, modules, v.s., and algebras. The notions of substructures and quotient structures are also defined.

---

at the end of the 19th century that the axiomatic method experienced a growing interest with the works of Richard Dedekind (1831–1916), Georg Cantor (1845–1918), and Giuseppe Peano (1858–1932), for the construction of the sets of integers and real numbers, as well as those of David Hilbert for his axiomatization of Euclidean geometry.

The v.s. structure is considered in more detail. Different examples are given, including v.s. of linear maps and multilinear maps. The concepts of vector subspace, linear independence, basis, dimension, direct sum of subspaces, and quotient space are recalled, before summarizing the different structures under consideration in a table.

The notion of homomorphism, also called morphism, is then introduced, and morphisms of groups, rings, vector spaces, and algebras are described. The case of morphisms of v.s., that is, of linear maps, is addressed in more depth. The notions of isomorphism, endomorphism, and dual basis are defined. The canonical factorization of linear maps based on the notion of quotient v.s., and the rank theorem, which is a fundamental result in linear algebra, are presented.

## 2.3. Sets

### 2.3.1. *Definitions*

A set $\mathcal{A}$ is a collection of elements $\{a_1, a_2, \cdots\}$. It is said that $a_i$ is an element of the set $\mathcal{A}$, or $a_i$ belongs to $\mathcal{A}$, and it is written $a_i \in \mathcal{A}$, or $\mathcal{A} \ni a_i$.

A subset $\mathcal{B}$ of a set $\mathcal{A}$ is a set whose elements also belong to $\mathcal{A}$. It is said that $\mathcal{B}$ is included in $\mathcal{A}$, and it is written $\mathcal{B} \subseteq \mathcal{A}$ or $\mathcal{A} \supseteq \mathcal{B}$

$$\mathcal{B} \subseteq \mathcal{A} \;\Leftrightarrow\; \forall x \in \mathcal{B} \;\Rightarrow\; x \in \mathcal{A}.$$

If $\mathcal{B} \subseteq \mathcal{A}$ and $\mathcal{B} \neq \mathcal{A}$, then it is said that $\mathcal{B}$ is a proper subset of $\mathcal{A}$, and we write $\mathcal{B} \subset \mathcal{A}$.

The empty set, denoted by $\emptyset$, is by definition the set that contains no elements. We have $\emptyset \subseteq \mathcal{A}, \forall \mathcal{A}$.

A finite set $E$ is a set that has a finite number of elements. This number $N$ is called the cardinality of $E$, and it is often denoted by $|E|$ or Card$(E)$. There are $2^N$ distinct subsets.

An infinite set $E$ is said to be countable when there exists a bijection between $E$ and the set of natural numbers ($\mathbb{N}$) or integers ($\mathbb{Z}$). This definition is due to Cantor[4]. This means that the elements of a countable set can be indexed as $x_n$, with $n \in \mathbb{N}$ or $\mathbb{Z}$. This is the case of sampled signals, namely, discrete-time signals, where $n$ is the sampling index ($t = nT_e$ , where $T_e$ is the sampling period), while sets of analog signals (i.e. continuous-time signals) $x(t)$ are not countable with respect to the time variable $t \in \mathbb{R}$.

4 Georg Cantor (1845–1918), Russian mathematician who is at the origin of the theory of sets. He is known for the theorem that bears his name, relative to set cardinality, as well as for his contributions to the theory of numbers.

### 2.3.2. *Sets of numbers*

In Table 2.1, we present a few sets of numbers[5] that satisfy the following inclusion relations: $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{P} \subset \mathbb{R} \subset \mathbb{C} \subset \mathbb{Q} \subset \mathbb{O}$. We denote by $\mathbb{R}^* = \mathbb{R}\backslash\{0\}$ the set of

| Sets | Definitions |
|---|---|
| $\mathbb{N}$ | Natural numbers including 0 |
| $\mathbb{Z}$ | Integers |
| $\mathbb{P}$ | Rational numbers |
| $\mathbb{R}$ | Real numbers |
| $\mathbb{R}^+$ | Positive real numbers |
| $\mathbb{R}^-$ | Negative real numbers |
| $\mathbb{C}$ | Complex numbers |
| $\mathbb{Q}$ | Quaternions |
| $\mathbb{O}$ | Octonions |

**Table 2.1.** *Sets of numbers*

non-zero real numbers. Similarly for $\mathbb{N}^*$, $\mathbb{Z}^*$, $\mathbb{P}^*$, and $\mathbb{C}^*$. In Volume 2, a chapter will be dedicated to complex numbers, quaternions, and octonions[6], with the purpose of highlighting the matrix representations of these numbers, and introducing quaternionic and octonionic matrices.

### 2.3.3. *Cartesian product of sets*

The Cartesian product[7] of $N$ sets $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_N$, denoted $\mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_N$, or still $\underset{n=1}{\overset{N}{\times}} \mathcal{A}_n$, is the set of all ordered $N$-tuples $(x_1, x_2, \cdots, x_N)$, where $x_n \in \mathcal{A}_n, n \in \langle N \rangle = \{1, 2, \cdots, N\}$:

$$\underset{n=1}{\overset{N}{\times}} \mathcal{A}_n = \{(x_1, x_2, \cdots, x_N) : x_n \in \mathcal{A}_n, n \in \langle N \rangle\}.$$

5 For the set of rational numbers, the notation $\mathbb{P}$ is a substitute for the usual notation $\mathbb{Q}$, which will be used to designate the set of quaternions instead of $\mathbb{H}$, often used to refer to Hamilton, discoverer of quaternions.

6 Quaternions and octonions, which can be considered as generalizations of complex numbers, themselves extending real numbers, are part of hypercomplex numbers.

7 The notion of Cartesian product is due to René Descartes (1596–1650), French philosopher, mathematician, and physicist, and author of philosophical works including the treatise entitled *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences* (Discourse on the Method for Rightly Conducting the Reason, and Seeking Truth in the Sciences), which contains the famous quote "I think, therefore I am" (originally in Latin "Cogito, ergo sum"). He introduced the Cartesian product to represent the Euclidian plane and three-dimensional space, in the context of analytic geometry, also called Cartesian geometry, using coordinate systems.

| Operations | Definitions |
|---|---|
| Equality | $\mathcal{A} = \mathcal{B}$ if and only if $\mathcal{A} \subseteq \mathcal{B}$ and $\mathcal{B} \subseteq \mathcal{A}$. |
| Transitivity | if $\mathcal{A} \subset \mathcal{B}$ and $\mathcal{B} \subset \mathcal{C}$, then $\mathcal{A} \subset \mathcal{C}$. |
| Union (or sum) | $\mathcal{A} \cup \mathcal{B} = \{x : x \in \mathcal{A} \text{ or } x \in \mathcal{B}\}$. |
| Intersection (or product) | $\mathcal{A} \cap \mathcal{B} = \{x : x \in \mathcal{A} \text{ and } x \in \mathcal{B}\}$. |
| Complementation | $\mathcal{A} \subset \Omega \Rightarrow \overline{\mathcal{A}} = \{x \in \Omega : x \notin \mathcal{A}\}$. |
| Reduction (or difference) | $\mathcal{A} - \mathcal{B} = \mathcal{A} \cap \overline{\mathcal{B}}$. |
| Exclusive or | $\mathcal{A} \oplus \mathcal{B} = (\mathcal{A} - \mathcal{B}) \cup (\mathcal{B} - \mathcal{A})$. |

**Table 2.2.** *Set operations*

For example, we define the Cartesian product of $N$ sets of indices $\mathcal{J}_n = \{1, \cdots, I_n\}$, as $\mathcal{J} = \overset{N}{\underset{n=1}{\times}} \mathcal{J}_n$. The elements of $\mathcal{J}$ are the ordered $N$-tuples of indices $(i_1, \cdots, i_N)$, with $i_n \in \mathcal{J}_n$. Later in the book, $\overset{N}{\underset{n=1}{\times}} I_n$ will be used to highlight the dimensions.

When $\mathcal{A}_n = \mathcal{A}, \forall n \in \langle N \rangle$, the Cartesian product will be written as $\overset{N}{\underset{n=1}{\times}} \mathcal{A}_n = \mathcal{A}^N$.

If the sets are vector spaces, we then have a Cartesian product of vector spaces which is a fundamental notion underlying, in particular, the definition of multilinear maps, and therefore, as it will be seen in section 6.6, that of tensor spaces.

### 2.3.4. *Set operations*

In Table 2.2, we summarize the main set operations[8].

Union and intersection are commutative, associative, and distributive:

– Commutativity: $\mathcal{A} \cup \mathcal{B} = \mathcal{B} \cup \mathcal{A}$ ; $\mathcal{A} \cap \mathcal{B} = \mathcal{B} \cap \mathcal{A}$.

– Associativity: $(\mathcal{A} \cup \mathcal{B}) \cup \mathcal{C} = \mathcal{A} \cup (\mathcal{B} \cup \mathcal{C})$ ; $(\mathcal{A} \cap \mathcal{B}) \cap \mathcal{C} = \mathcal{A} \cap (\mathcal{B} \cap \mathcal{C})$.

– Distributivity: $\mathcal{A} \cup (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{C})$ ; $\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{C})$.

Exclusive or (also called exclusive disjonction), noted $\mathcal{A} \oplus \mathcal{B}$, is the set of all elements of $\mathcal{A}$ or $\mathcal{B}$ which do not belong to the two sets at once.

8 Set union and intersection are also denoted by $\mathcal{A} + \mathcal{B}$ and $\mathcal{A}\,\mathcal{B}$, respectively.

$N$ sets $\mathcal{A}_1, \cdots, \mathcal{A}_N$ are said to be mutually exclusive if they are pairwise disjoint, that is, if $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset\,,\ \forall i, j \neq i$.

The following properties hold for $\mathcal{A} \subset \Omega$ and $\mathcal{B} \subset \Omega$:

– $\mathcal{A} \cup \overline{\mathcal{A}} = \Omega\,;\, \mathcal{A} \cap \overline{\mathcal{A}} = \emptyset\,;\, \mathcal{A} = \overline{\overline{\mathcal{A}}}$.

– $\overline{\Omega} = \emptyset\,;\, \overline{\emptyset} = \Omega$.

– If $\mathcal{B} \subset \mathcal{A}$ then $\overline{\mathcal{B}} \supset \overline{\mathcal{A}}$.

### 2.3.5. *De Morgan's laws*

De Morgan's laws, also called rules[9], are properties related to the complement of a union or an intersection of subsets of the same set. Thereby, for two subsets $\mathcal{A}$ and $\mathcal{B}$, it follows that:

$$\overline{\mathcal{A} \cup \mathcal{B}} = \overline{\mathcal{A}} \cap \overline{\mathcal{B}}\ ;\ \overline{\mathcal{A} \cap \mathcal{B}} = \overline{\mathcal{A}} \cup \overline{\mathcal{B}}$$

and in general for $N$ subsets:

$$\overline{\bigcup_{n=1}^{N} \mathcal{A}_n} = \bigcap_{n=1}^{N} \overline{\mathcal{A}_n}\ ;\ \overline{\bigcap_{n=1}^{N} \mathcal{A}_n} = \bigcup_{n=1}^{N} \overline{\mathcal{A}_n}.$$

The equalities above are logical equivalences, and the symbol of equality can be replaced by the symbol of equivalence ($\Leftrightarrow$).

De Morgan's laws express the fact that the complement of unions and intersections of sets can be obtained by replacing all sets by their complements, unions by intersections, and intersections by unions. Therefore, for example:

$$\overline{\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C})} \Leftrightarrow \overline{\mathcal{A}} \cup (\overline{\mathcal{B}} \cap \overline{\mathcal{C}}),$$

or equivalently:

$$\overline{\mathcal{A}(\mathcal{B} + \mathcal{C})} \Leftrightarrow \overline{\mathcal{A}} + \overline{\mathcal{B}}\,\overline{\mathcal{C}}.$$

### 2.3.6. *Characteristic functions*

For a given subset $F$ of $E$, the characteristic function, or indicator function, is the function $\chi_F : E \to \{0, 1\}$ such that:

$$E \ni x \mapsto \chi_F(x) = \begin{cases} 1 & \text{if } x \in F \\ 0 & \text{if } x \notin F \end{cases}.$$

9 Augustus de Morgan (1806–1871), British mathematician who is the founder of modern logic with George Boole (1815–1864).

### 2.3.7. *Partitions*

A $N$-partition of $\Omega$ is a collection of $N$ disjoint subsets $\mathcal{A}_n$, $n \in \langle N \rangle$, of $\Omega$ whose union is equal to $\Omega$:

$$\mathcal{A}_n \subset \Omega \ , \ \bigcup_{n=1}^{N} \mathcal{A}_n = \Omega \ , \ \mathcal{A}_n \cap \mathcal{A}_i = \emptyset \ \ \forall n \text{ and } i \neq n.$$

### 2.3.8. *$\sigma$-algebras or $\sigma$-fields*

Let $\Omega$ be a non-empty set. A $\sigma$-algebra (or $\sigma$-field) on $\Omega$ is a collection $A$ of subsets of $\Omega$ satisfying the following properties:

– $A$ is not empty.

– $A$ is closed under complement, namely, $\forall \mathcal{A}_n \in A$, then $\overline{\mathcal{A}}_n \in A$.

– $A$ is closed under countable unions, namely, if $\mathcal{A}_n \in A, \forall n \in \mathbb{N}$, then $\bigcup_{n \in \mathbb{N}} \mathcal{A}_n \in A$. A union is said to be countable because the set of subsets $\mathcal{A}_n$ is countable.

The pair $(\Omega, A)$ is called a measurable space, and the subsets $\mathcal{A}_n$ are called measurable sets. By equipping the measurable space $(\Omega, A)$ of a measure $\mu : A \rightarrow [0, +\infty]$, the triplet $(\Omega, A, \mu)$ is called a measure space.

In probability theory, $\Omega$ is the universal set, that is, the set of all possible experimental outcomes of a random trial, also called elementary events. Defining an event $\mathcal{A}_n$ as a set of elementary events, a collection (or field) $A$ of events is called a $\sigma$-field, and the pair $(\Omega, A)$ is a measurable space. When this space is endowed with a probability measure $P$, the triplet $(\Omega, A, P)$ defines a probability space, where $P$ satisfies for any element $\mathcal{A}_n$ of $A$: $P(\emptyset) = 0$; $P(\Omega) = 1$; $0 \leq P(A_n) \leq 1$ and $P(\emptyset) = 0$, meaning that the empty set is an impossible event, whereas $P(\Omega) = 1$ means that $\Omega$ is a sure event, that is, an event which always occurs.

### 2.3.9. *Equivalence relations*

Let $E$ be a non-empty set. An equivalence relation on $E$, denoted by $\sim$, is a binary relation that is reflexive, symmetric, and transitive:

– Reflexivity: $\forall a \in E \ , \ a \sim a.$

– Symmetry: $\forall (a, b) \in E^2 \ , \ a \sim b \ \Rightarrow \ b \sim a.$

– Transitivity: $\forall (a, b, c) \in E^3 \ , \ a \sim b \ \text{ and } \ b \sim c \ \Rightarrow \ a \sim c.$

The elements equivalent to an element $a$ form a set called the equivalence class of $a$, denoted by $c_a \subset E$. The set of all equivalence classes associated with the equivalence relation $\sim$ forms a partition of $E$, denoted by $E/\sim$ and called quotient set or quotient space of $E$.

### 2.3.10. *Order relations*

An order relation on a set $E$, usually denoted by $<$, is a binary relation which is reflexive, antisymmetric, and transitive:

– Reflexivity: $\forall a \in E\ ,\ a < a$.

– Antisymmetry: $\forall (a, b) \in E^2\ ,\ (a < b \text{ and } b < a) \Rightarrow a = b$.

– Transitivity: $\forall (a, b, c) \in E^3\ ,\ a < b \text{ and } b < c \Rightarrow a < c$.

A set $E$ endowed with an order relation is said to be ordered. Order relations allow the notions of partially/totally ordered sets, and upper bound/lower bound to be defined, with the possibility to introduce inequalities.

EXAMPLE 2.1.– The relation $a < b$, where $<$ is the symbol less than, is not an order relation on $\mathbb{R}$ because it is neither reflexive nor antisymmetric. On the other hand, $a \leq b$ is an order relation on $\mathbb{R}$.

NOTE 2.2.– In $\mathbb{C}$, two complex numbers cannot be compared. However, it is possible to define a lexicograpic order such that: $z_1 = a_1 + ib_1 < z_2 = a_2 + ib_2$, with $i^2 = -1$, if and only if either $a_1 < a_2$, or $a_1 = a_2$ and $b_1 < b_2$.

## 2.4. Maps and composition of maps

### 2.4.1. *Definitions*

Given two sets $E$ and $F$, a map (or mapping) $f$ from $E$ to $F$ assigns every element $x$ of $E$ to one and only one element of $F$, denoted $f(x)$, namely:

$$f : E \to F,\ x \mapsto y = f(x).$$

It is said that $y$ is the image of $x$ and $x$ is an antecedent of $y$ by $f$. The set $f(E)$ is called the image of $E$, denoted $\text{Im}(f)$ and defined by:

$$\text{Im}(f) = f(E) = \{y \in F :\ \exists x \in E\ ,\ y = f(x)\}.$$

When $0 \in F$, the kernel of $f$, denoted by $\text{Ker}(f)$ or $\mathcal{N}(f)$, is the set defined by:

$$\text{Ker}(f) = \{x \in E\ :\ f(x) = 0\}.$$

NOTE 2.3.– The difference between map and function lies in the fact that a function is not necessarily defined on the whole set $E$, that is, some elements of $E$ may not have any image.

### 2.4.2. *Properties*

Injectivity, surjectivity, and bijectivity properties of maps are summarized below.

– $f$ is said to be injective if and only if $\forall x, u \in E,\ f(x) = f(u) \Rightarrow x = u$. This means that at every element of $F$ corresponds at most one element of $E$.

– $f$ is said to be surjective if and only if $\forall y \in F, \exists x \in E$ such that $f(x) = y$, that is, $\mathrm{Im}(f) = F$. This means that at every element of $F$ corresponds at most one element of $E$.

– $f$ is said to be bijective if and only if it is both injective and surjective. It is also said that it is a bijection. $f$ is then invertible, because for all $y \in F$, there exists a unique $x \in E$ such that $f(x) = y$. The inverse of $f$, called the inverse map and denoted by $f^{-1}$, is also a bijection.

### 2.4.3. *Composition of maps*

The composition of two maps $f$ from *E* to *F* and $g$ from *F* to *G*, denoted by $g \circ f$ or $g(f)$, is a map from *E* to *G*, with $f$ applied before $g$, defined by:

$$g \circ f :\ E \to G\,,\ x \mapsto g[f(x)] = (g \circ f)(x).$$

The symbol $\circ$ defines a law of composition which is associative, meaning that, for $E \xrightarrow{f} F \xrightarrow{g} G \xrightarrow{h} H$, we have:

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

## 2.5. Algebraic structures

In this section, we shall make a brief overview of a few algebraic structures: semi-groups, monoids, groups, Abelian groups, semi-rings, rings, ideals, fields, modules, vector spaces, and algebras. These structures can be distinguished by the binary operations that enable combining the elements of a set and by the properties that satisfy these operations. We first define what is a law of composition, also called an algebraic law, and we then describe the properties that such law can satisfy.

### 2.5.1. *Laws of composition*

#### 2.5.1.1. *Definitions*

Let $E$ be a non-empty set. Two types of laws of composition can be distinguished: internal laws between two elements of $E$ and external laws between an element of $E$ and an element of another non-empty set, denoted by $\mathbb{K}$ and called an operator set or domain. In the following, $\mathbb{K}$ will be mostly equal to $\mathbb{R}$ or $\mathbb{C}$.

An internal composition law $*$ on $E$ is a map $E^2 \to E$ that establishes a correspondence between the element $(x, y)$ of $E^2$ and a unique element of $E$, denoted by $x * y$. We denote by $(E, *)$ the set $E$ endowed with the operation $*$, also called binary operation, or internal law. Such a set is called a magma or groupoid.

An external law of composition to the left of $E$, denoted by $\circ$, is a map $\mathbb{K} \times E \to E$ such that:

$$\mathbb{K} \times E \ni (\lambda, x) \mapsto \lambda \circ x \in E.$$

Similarly, $\circ$ is an external law of composition to the right of $E$ if:

$$E \times \mathbb{K} \ni (x, \lambda) \mapsto x \circ \lambda \in E.$$

It is also said that $\circ$ is an action of $\mathbb{K}$ on $E$. It is often the case that $\lambda x$ is used instead of $\lambda \circ x$.

#### 2.5.1.2. *Properties of internal laws*

We present hereafter the main properties satisfied by an internal law.

– The operation $*$ is said to be commutative if: $\forall (x, y) \in E^2$, we have:

$$x * y = y * x.$$

– The operation $*$ is called associative if: $\forall (x, y, z) \in E^3$, we have:

$$(x * y) * z = x * (y * z) = x * y * z.$$

– The operation $*$ admits a neutral (also called unit or identity) element $e \in E$ if: $\forall x \in E$, we have:

$$x * e = e * x = x.$$

If the operation $*$ has a neutral element in $E$, it is then unique.

If the composition law is multiplicative, it is often denoted by $\cdot$ or $\times$ and the neutral element is usually denoted by $1_E$ or simply $1$. When it is additive, it is denoted by $+$, and the neutral element is denoted by $0_E$ or $0$.

– The element $x \in E$ has a right-symmetric (left-symmetric) if: $\exists\, x' \in E$ such that:

$$x * x' = e \quad (x' * x = e).$$

When $x * x' = x' * x = e$, then $x'$ is said to be a symmetric of $x$.

If the composition law is multiplicative, the symmetric of $x$ is called the inverse of $x$, and denoted by $x^{-1}$. When it is additive, the symmetric of $x$ is called the opposite of $x$, and denoted by $-x$.

– Given the set $E$ equipped with two internal composition laws $*$ and $+$.

It is said that $*$ is left-distributive over $+$ if: $\forall(x, y, z) \in E^3$, we have:

$$x * (y + z) = (x * y) + (x * z).$$

It is said that $*$ is right-distributive over $+$ if: $\forall(x, y, z) \in E^3$, we have:

$$(x + y) * z = (x * z) + (y * z).$$

It is said that $*$ is distributive over $+$ if it is left- and right-distributive.

EXAMPLE 2.4.– Addition $(+)$ and multiplication $(\cdot)$ are commutative and associative binary operations on the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{P}$, $\mathbb{R}$, and $\mathbb{C}$, but multiplication is not commutative on the set $\mathbb{Q}$ of quaternions, and on the set $\mathbb{K}^{n\times n}$ of square matrices of order $n$.

### 2.5.1.3. *Closure and induced laws*

The set $E$ is said to be closed under the operation $*$ (or stable for $*$) if for any pair $(x, y) \in E^2$, we have $x * y \in E$. The closure property means that the composition $x * y$ of any two elements $x$ and $y$ in $E$ gives an element of $E$.

$F$ being a non-empty subset of $(E, *)$, it is said that $F$ is stable for $*$ (or closed under $*$) if:

$$\forall(x, y) \in F^2,\ x * y \in F.$$

It is then said that $*$ is an induced law on $F$ by the law of $E$, or $F$ inherits the same structure as $E$.

### 2.5.1.4. *Illustration with the convolution operation*

In signal processing, the convolution operation, denoted by $*$, is defined by means of the equations in Table 2.3, where $(x_n, y_n, h_n)$ and $(x(t), y(t), h(t))$ are the input and output signals, and the impulse response of a linear time invariant (LTI) system, discrete and continuous, respectively.

| Systems | Convolutions |
|---|---|
| Discrete | $y = h * x \Leftrightarrow y_n = (h * x)(n) = \sum_{k=-\infty}^{\infty} h_k x_{n-k}$ |
| Continuous | $y = h * x \Leftrightarrow y(t) = (h * x)(t) = \int_{-\infty}^{\infty} h(s)x(t-s)ds$ |

**Table 2.3.** *Convolution operation*

This operation satisfies the following properties:

– Commutativity: $h * x = x * h$.

– Associativity: $g * (h * x) = (g * h) * x$.

– Distributivity: $h * (x + u) = h * x + h * u\,;\, (g + h) * x = g * x + h * x$.

The commutativity property can be verified by making the changes of variable $m = n - k$ and $q = t - s$ in the definition of convolution:

$$(h * x)(n) = \sum_{k=-\infty}^{\infty} h_k x_{n-k} = \sum_{m=-\infty}^{\infty} x_m h_{n-m} = (x * h)(n)$$

$$(h * x)(t) = \int_{s=-\infty}^{\infty} h(s)x(t-s)\,ds = \int_{q=-\infty}^{\infty} x(q)h(t-q)\,dq = (x * h)(t).$$

The associativity property reflects the fact that the output of a system composed of two LTI systems in cascade, with impulse responses $h$ and $g$, is equivalent to the output of an LTI system whose impulse response is given by the convolution $g*h$ of the impulse responses of the two LTI systems.

Indeed, defining $u = h * x$, the output $y_n = \big(g * (h * x)\big)(n)$ can be developed as:

$$y_n = (g * u)(n) = \sum_k g_k\, u_{n-k} = \sum_k g_k \big( \sum_l h_l\, x_{n-k-l} \big)$$

and the change of variable $q = k + l$ gives:

$$y_n = \sum_q \big( \sum_k g_k h_{q-k} \big)\, x_{n-q} = \sum_q (g * h)(q)\, x_{n-q} = \big((g * h) * x\big)(n)$$

which demonstrates the associativity property: $g * (h * x) = (g * h) * x$.

NOTE 2.5.– A connection in cascade of two LTI systems is such that the output of the first system $(u = h * x)$ is the input of the second one whose output is given by $y = g * u$. The commutative and associative properties of the convolution operation imply that $g * h = h * g$, meaning that the output of the global system, i.e. the output of the system inter-connected in cascade, is not changed if the order of connecting the LTI systems is reversed. However, in practice, the linearity of a physical system being conditioned on the amplitude of its input which must be less than a certain threshold, the connection in cascade of two LTI systems must be such that the output of the first system is within the range of values for which the second LTI system remains linear.

The convolution also satisfies the properties of linearity and translation-invariance (or shift-invariance):

– Linearity: $h * (\lambda x + u) = \lambda(h * x) + h * u \ , \ \forall \lambda \in \mathbb{K}$.

– Translation: $h * [x(n - \tau)] = (h * x)(n - \tau) = y(n - \tau) \ , \ \forall \tau \in \mathbb{N}^*$.

The property of linearity is also called principle of superposition in the sense where a superposition of inputs $\lambda x$ and $u$ is associated with the superposition of the corresponding outputs $\lambda(h * x)$ and $h * u$.

Translation-invariance, related to the non-dependency of $h$ with respect to time ($n$ or $t$), means that a time delay $\tau$ of the input corresponds to the same time delay for the output, which explains the designation by linear time-invariant system. In the case of analog systems, replace $n$ by $t$ in the property of translation-invariance, with $\tau \in \mathbb{R}^+$.

Since the summation and integration are infinite in the definition equations, the existence of the convolution product is related to the convergence of the sequence $y_n$ in the discrete case, and of the function $y(t)$, in the continuous case. This is verified, for example, when $(h_n, x_n)$ and $(h(t), x(t))$ are absolutely summable digital signals on $\mathbb{Z}$, and absolutely integrable analog signals on $\mathbb{R}$, respectively. In the next chapter, we shall see that this amounts to saying that the signals belong to the Hilbert spaces $l^1(\mathbb{Z}, \mathbb{K})$ and $L^1(\mathbb{R}, \mathbb{K})$ defined as:

$$l^1(\mathbb{Z}, \mathbb{K}) = \left\{h : \mathbb{Z} \ni n \mapsto h_n \in \mathbb{K} \, , \sum_{n \in \mathbb{Z}} |h_n| < \infty\right\} \quad [2.1]$$

$$L^1(\mathbb{R}, \mathbb{K}) = \left\{h : \mathbb{R} \ni t \mapsto h(t) \in \mathbb{K} \, , \int_{t \in \mathbb{R}} |h(t)|dt < \infty\right\}. \quad [2.2]$$

When the support of the impulse response $h_n$ is finite, the discrete system is called a finite impulse response (FIR) system. Otherwise, it is called an infinite impulse response (IIR) system. FIR and IIR systems play a very important role in signal processing, both for modeling linear systems and for linear filtering of signals.

### 2.5.2. *Definition of algebraic structures*

The properties previously listed define calculus rules, called axioms. In Table 2.4, we summarize the definitions of different algebraic structures, indicating the internal and external operations involved and their properties for each structure. The operations designated by $(*, +, +_M, +_E, \times_E, .)$ represent internal operations, that is, operations between elements of a same set ($E$, $A$, $\mathbb{K}$, or $M$), while $(\circ, \times)$ represent external operations, called scalar multiplications, namely, operations between elements of $E$ and of a field $\mathbb{K}$ or a ring $A$.

| Structures | Properties |
|---|---|
| **Semi-group** $(E, *)$ | $*$ is associative |
| **Monoid** $(E, *)$ | Semi-group with an identity element |
| **Group** $(E, *)$ | Monoid in which each element has a symmetric |
| **Abelian group** $(E, *)$ **or Commutative group** | Group with $*$ commutative |
| **Semiring** $(A, +, \cdot)$ | $(A, +)$ is an Abelian semi-group<br>$(A, \cdot)$ is a semi-group<br>$\cdot$ is distributive over $+$ |
| **Ring** $(A, +, \cdot)$ | $(A, +)$ is an Abelian group<br>$(A, \cdot)$ is a semi-group<br>$\cdot$ is distributive over $+$ |
| **Field** $(\mathbb{K}, +, \cdot)$ | $(\mathbb{K}, +, \cdot)$ is a unitary ring<br>$(\mathbb{K}^*, \cdot)$ is a group with the multiplicative identity $1_{\mathbb{K}}$ |
| **Module** $(M, +_M, \times)$<br>over a ring $(A, +, \cdot)$<br>$+_M$ is an internal law<br>$\times$ is an external law | $(M, +_M)$ is an Abelian additive group<br>$(A, +, \cdot)$ is a unitary ring with the multiplicative identity $1_A$<br>The external operation $\times$ is associative<br>The external operation $\times$ is distributive over $+$ and $+_M$<br>The identity element $1_A$ for the operation $\cdot$ is also the identity element for $\times$ |
| **Vector space** $(E, +_E, \circ)$<br>over a field $(\mathbb{K}, +, \cdot)$<br>$+_E$ is an internal law<br>$\circ$ is an external law | $(E, +_E)$ is an Abelian additive group<br>The external operation $\circ$ is associative<br>The external operation $\circ$ is distributive over $+$ and $+_E$<br>The identity element $1_{\mathbb{K}}$ for the operation $\cdot$ is also the identity element for $\circ$ |
| **Algebra** $(E, +_E, \times_E, \circ)$<br>over a field $(\mathbb{K}, +, \cdot)$<br>$(+_E, \times_E)$ are internal laws<br>$\circ$ is an external law | $(E, +_E, \circ)$ is a vector space over $\mathbb{K}$<br>$(E, +_E, \times_E)$ is a ring<br>The internal operation $\times_E$ is distributive over the external law $\circ$ |

**Table 2.4.** *Algebraic structures*

It should be noted that semi-groups, monoids, and groups involve a single internal operation only, denoted by $*$, while semirings, rings, and fields involve two internal operations denoted by $(+, \cdot)$. In the case of modules and vector spaces, there is an internal operation and an external operation designated by $(+_M, \times)$ and $(+_E, \circ)$, respectively. Finally, in the case of algebras, there are two internal operations $(+_E\ , \times_E)$ and an external operation $(\circ)$.

### 2.5.3. *Substructures*

Let $E$ be a set equipped with a structure such as a group, ring, field, or v.s. It is said that a subset $F$ of $E$ is endowed with a substructure such as a subgroup, subring, subfield, or vector subspace (also called subspace), if the following conditions are met:

– $F$ is stable for the laws of $E$.

– $F$ contains the identity elements of $E$.

– Laws induced on $F$ by the laws of $E$ satisfy the definition axioms of the structure of $E$, thus inducing the same structure for $F$.

### 2.5.4. *Quotient structures*

Given a set $E$ with a specific structure, it is said that a subset of $E$ is equipped with a quotient structure for an equivalence relation if its elements are equivalence classes on $E$, with operations satisfying the same axioms as the structure of $E$. Therefore, a quotient structure consists of elements of $E$ sharing a property defined by an equivalence relation.

So, we can define quotient groups, quotient rings, and quotient spaces. In section 2.5.6.7, we give as an example the quotient ring $\mathbb{Z}/n\mathbb{Z}$ of integers congruent modulo $n$, and in section 2.5.15, we define the notion of quotient vector space, which is used in section 2.6.3.2 for the canonical factorization of linear maps.

In the following, we detail the following structures: groups, rings, fields, modules, vector spaces, and algebras.

### 2.5.5. *Groups*

#### 2.5.5.1. *Definition*

A set $(E, *)$ is a group if it is closed under the $*$ operation, and if it satisfies the following axioms:

– The $*$ operation is associative: $x * (y * z) = (x * y) * z \ \ , \forall x, y, z \in E$.

– There exists a neutral element $e \in E$ such that: $x * e = e * x = x \ \ , \forall x \in E$.

– Any element $x \in E$ has a symmetric $x' \in E$ such that: $x * x' = x' * x = e$.

A group is thus a monoid having in addition the property of existence of a symmetric for each element. The neutral element and the symmetric of any element of a group are unique. If the group $E$ is finite, Card$(E)$ is called the order of $E$. Note that a set equipped with an associative binary operation is called a semi-group.

### 2.5.5.2. *Commutative or Abelian groups*

A group is commutative or Abelian (in honor of the mathematician Abel[10]) if the $*$ operation is commutative, that is, $\forall x, y \in E, \quad x * y = y * x$.

### 2.5.5.3. *Additive groups*

In the case of an Abelian group, the internal composition law is, in general, denoted by +, and it is then referred to as additive group. The previous axioms then become:

– Commutativity property:

$$x + y = y + x \ , \ \forall x, y \in E.$$

– Additive associativity property:

$$x + (y + z) = (x + y) + z = x + y + z \ , \ \forall x, y, z \in E.$$

– The additive neutral element (called zero element), denoted by $0_E$ or simply 0, is such that:

$$x + 0_E = 0_E + x = x \ , \ \forall x \in E.$$

– The symmetric of $x$, denoted by $(-x)$ and called opposite or additive inverse, is such that:

$$x + (-x) = (-x) + x = 0_E.$$

### 2.5.5.4. *Examples*

As examples, we provide hereafter sets of numbers and the permutation group.

– *Numbers:* $(\mathbb{Z}, +), (\mathbb{P}, +), (\mathbb{R}, +)$, and $(\mathbb{C}, +)$ are additive Abelian groups, including 0 as neutral element. $(\mathbb{P}, .), (\mathbb{R}, .)$, and $(\mathbb{C}, .)$ are not groups because the element 0 is not invertible. On the other hand, $(\mathbb{P}^*, .), (\mathbb{R}^*, .)$, and $(\mathbb{C}^*, .)$ are multiplicative Abelian groups, with 1 as identity element.

$(\mathbb{N}, +)$ and $(\mathbb{N}, .)$ are not groups because the opposite and the inverse of a natural number are not natural numbers.

---

10 Niels Henrik Abel (1802–1829), Norwegian mathematician who obtained several important results on the convergence of numerical series and generalized integrals, and on the resolution of quintic equations using elliptic functions. In 1830, he received with Carl Jacobi (1804–1851) the Grand Prix de l'Académie des Sciences (Paris) for their contributions to the theory of elliptic functions. In his memory, since 2003, the Norwegian Academy of Science and Letters awards the Abel Prize, equivalent to the Nobel Prize for mathematicians.

– *Permutation group of a finite set:* Given a finite set $E$ having a cardinality of $N$, a permutation of $E$ is a bijection from $E$ to itself.

A permutation is often denoted by $\pi(\cdot)$. For example, for $E = \langle N \rangle = \{1, \cdots, N\}$, the set of the first $N$ natural numbers, we have:

$$\pi : \langle N \rangle \to \langle N \rangle \ , \ \langle N \rangle \ni i \mapsto \pi(i) = j \in \langle N \rangle.$$

If $i < j$, the pair $\{i, j\}$ is called an inversion by $\pi$ if $\pi(i) - \pi(j) > 0$, therefore of opposite sign to $i - j$. The signature (or sign) of a permutation $\pi$, denoted by $\sigma(\pi)$, is defined as the parity of the number of inversions associated with pairs $\{i, j\}$ with $i < j$, that is, equal to +1 if this number is even, and equal to $-1$ if it is odd. The signature is given by the following formula:

$$\sigma(\pi) = \prod_{1 \leq i < j \leq N} \frac{\pi(j) - \pi(i)}{j - i}.$$

So, a permutation $\pi$ is even (respectively, odd) if $\sigma(\pi) = 1$ (respectively, $\sigma(\pi) = -1$).

The set of all permutations of $E$, endowed with the composition law $\circ$ corresponding to the composition of maps, is a group, called symmetric group of order $N$, and denoted by $\mathcal{S}_N$, of cardinality $N!$.

Indeed, the composition of two permutations is a permutation. Consequently, the composition law $\circ$ is an internal composition law on $\mathcal{S}_N$. This operation is associative, the neutral element is the permutation that transforms each element of $E$ to itself, and every permutation admits an inverse permutation. The set $(\mathcal{S}_N, \circ)$ is thus a group.

The composition of two permutations being not commutative, the group of permutations is not commutative.

EXAMPLE 2.6.– For the set $E = \{i, j, k\}$ of three indices and the permutations: $\pi_1 : (i, j, k) \mapsto (j, k, i)$ and $\pi_2 : (i, j, k) \mapsto (k, j, i)$, we have: $\pi_1 \circ \pi_2 : (i, j, k) \mapsto (i, k, j)$ and $\pi_2 \circ \pi_1 : (i, j, k) \mapsto (j, i, k)$, and thus, $\pi_1 \circ \pi_2 \neq \pi_2 \circ \pi_1$.

#### 2.5.5.5. *Subgroups*

Let $F$ be a non-empty subset of a group $(E, *)$. $F$ is a subgroup of $E$ if and only if:

– $F$ is stable for $*$, that is: $\forall (x, y) \in F^2, \ x * y \in F$.

– $F$ equipped with the induced law $*$ has a group structure, which means that:

$$\forall x \in F, \ x' \in F, \text{ where } x' \text{ is the symmetric of } x.$$

When the law $*$ is written multiplicatively, the symmetric of $y$ being noted $y^{-1}$, a subgroup may be defined as $\forall (x, y) \in F^2 \ , \ xy^{-1} \in F$.

We can show that the neutral element of $F$ is the same as the one of $E$.

## 2.5.6. *Rings*

### 2.5.6.1. *Definition*

A ring is a set $A$ endowed with two internal composition laws, often written as $(+, \cdot)$, and called addition and multiplication, such that the axioms described in Table 2.5 are satisfied for all $\lambda, \mu, \nu \in A$, or more specifically:

– $(A, +)$ is an additive Abelian group.

– Multiplication $(\cdot)$ is associative, and therefore, $(A, \cdot)$ is a semi-group.

– Multiplication $(\cdot)$ is distributive over the addition $(+)$.

| Axioms | Properties |
|---|---|
| $\lambda + \mu \in A$ | closure under the $+$ operation |
| $\lambda + \mu = \mu + \lambda$ | $+$ is commutative |
| $\lambda + (\mu + \nu) = (\lambda + \mu) + \nu$ | $+$ is associative |
| $\lambda + 0 = 0 + \lambda = \lambda$ | 0 is the neutral element for $+$ |
| $\lambda + (-\lambda) = (-\lambda) + \lambda = 0$ | $(-\lambda)$ is the opposite of $\lambda$ |
| $\lambda.\mu \in A$ | closure under the $\cdot$ operation |
| $\lambda.(\mu.\nu) = (\lambda.\mu).\nu$ | $\cdot$ is associative |
| $\lambda.(\mu + \nu) = (\lambda.\mu) + (\lambda.\nu)$ | $\cdot$ is left-distributive over $+$ |
| $(\lambda + \mu).\nu = (\lambda.\nu) + (\mu.\nu)$ | $\cdot$ is right-distributive over $+$ |

**Table 2.5.** *Definition axioms for a ring*

It should be noted that in general, the sign $\cdot$ is omitted, and $\lambda \cdot \mu$ is replaced by $\lambda\mu$. On the other hand, the neutral element for the law $+$ is often denoted by $0_A$ or $0$.

FACT 2.7.– A semi-ring differs from a ring by the fact that a neutral element is not required for addition; in other words, $(A, +)$ is an Abelian semi-group instead of an Abelian group.

### 2.5.6.2. *Commutative unitary and integral rings*

Hereafter, we define the unitary commutative ring structure and the notion of integral ring.

– A ring $(A, +, \cdot)$ is commutative if the $\cdot$ operation is commutative, that is, if $p\,q = q\,p$, for all $(p, q) \in A^2$.

– The major difference between addition and multiplication operations is the fact that it is not assumed that any element of $A$ has a multiplicative inverse. Consequently, the set $(A, \cdot)$ is a semi-group.

A ring with a multiplicative identity element (called unit element and denoted by $1_A$ or 1) for multiplication is called a unitary (or unital) ring. We have $1 \cdot p = p \cdot 1 = p$ for all $p \in A$.

In some books, the definition of a ring includes the existence of a multiplicative identity. $(A, \cdot)$ is then a monoid.

– In the case of a unitary commutative ring $A$, an element $p \in A$ is called invertible if: $\exists\, q \in A$ such that $p\,q = q\,p = 1$. Usually, the (multiplicative) inverse of $p$ is denoted by $p^{-1}$. It is unique.

– If $p$ and $q$ are two elements of a commutative ring such that $p\,q = 0$, with $p \neq 0$ and $q \neq 0$, $p$ and $q$ are called zero divisors. An element that is not a zero divisor is called regular.

A ring is said to be integral if it has no zero divisors except 0, that is, we have $p\,q = 0 \;\Rightarrow\; p = 0$ or $q = 0$. A non-invertible element $p \neq 0$ of an integral ring is called irreducible. An integral ring is also called an integral domain.

– An element $p \in A$ is said to be nilpotent if there exists $n \in \mathbb{N}^*$ such that $p^n = 0$.

#### 2.5.6.3. *Characteristic of a ring*

Let $A$ be a unitary ring, with multiplicative identity $1_A$. The characteristic of $A$ is the smallest natural number $n \in \mathbb{N}^*$ such that $n.1_A = \underbrace{1_A + \cdots + 1_A}_{n \text{ terms}} = 0_A$. It is then said that $A$ has characteristic $n$. If no such integer exists, $A$ is said to have characteristic zero.

PROPOSITION 2.8.– *The characteristic of an integral unitary ring is either zero or a prime number.*

PROOF.– Let us assume that $n > 0$ and non-prime. It can then be written as $n = p\,q$, with $0 < p < n$ and $0 < q < n$. It can be deduced that $0_A = n.1_A = (p\,q).1_A = (p.1_A)(q.1_A)$. Since $A$ is integral, it implies that $p.1_A = 0$ or $q.1_A = 0$, which contradicts the assumption that $n$ is the smallest positive integer such that $n.1_A = 0$, and subsequently $n$ is prime. □

### 2.5.6.4. *Examples*

As examples of rings, we provide sets of numbers and the set of polynomials in one variable.

– *Numbers:* $(\mathbb{Z},+,\cdot)$, $(\mathbb{P},+,\cdot)$, $(\mathbb{R},+,\cdot)$, and $(\mathbb{C},+,\cdot)$ are commutative unitary and integral rings.

– *Polynomials:* The set of polynomials in one variable (or indeterminate) $z$, with coefficients in a commutative ring $A$, is a commutative unitary ring, denoted by $A[z]$, whose elements are written as $p(z)=\sum\limits_n a_n z^n$, with the following internal operations:

$$(\sum_n a_n z^n)+(\sum_n b_n z^n)=\sum_n (a_n+b_n)z^n$$

$$(\sum_n a_n z^n).(\sum_m b_m z^m)=\sum_p c_p z^p\ ,\ c_p=\sum_k a_k b_{p-k}.$$

The zero polynomial is the identity element for addition $(p(z)=0\,,\forall z)$, and the polynomial 1 is the identity element for multiplication $(p(z)=1\,,\forall z)$.

It should be noted that the coefficient $c_p$ is given by the convolution of the sequences of coefficients $\{a_n\}$ and $\{b_n\}$, as defined in Table 2.3.

Defining the degree of $p(z)$, denoted by $\deg(p)$, as the smallest natural number $N_p \in \mathbb{N}$ such that $a_n$ is zero for any $n > N_p$, and writing the polynomial $p(z)$ of degree $N_p$ by means of its coefficients as $p=(a_n)_{n\in\langle N_p\rangle}$, we have for $q=(b_n)_{n\in\langle N_q\rangle}$:

$$p+q=(c_n)_{n\in\langle N\rangle} \text{ with } N\leq \max(N_p,N_q)$$

$$pq=(c_n)_{n\in\langle N\rangle} \text{ with } N\leq N_p+N_q.$$

If $N_p\neq N_q$, then $\deg(p+q)=\max(N_p,N_q)$. In addition, if the ring $A$ is integral, we have $\deg(pq)=N_p+N_q$.

### 2.5.6.5. *Newton's binomial theorem*

If $x$ and $y$ are two elements of a ring (like the sets of real or complex numbers, polynomials, square matrices of the same size, etc.), Newton's binomial theorem[11],

11 Isaac Newton (1642–1727), British physicist, mathematician, astronomer, philosopher, theologian who founded classical mechanics, also known as Newtonian mechanics, with his theory about the motion of bodies. His best-known discovery concerns the universal law of gravitation, which would have been inspired by the fall of an apple on his head, a legend based on the account that his doctor made. Newton is also known for the binomial theorem, and in numerical analysis for the infinitesimal calculus of which he is the discoverer, and the Newton–Raphson method, originally developed for finding an approximation for the roots of a polynomial.

also called the binomial theorem, is given for any natural number $n$, by:

$$(x+y)^n = \sum_{i=1}^{n} C_n^i x^{n-i} y^i \text{ with } C_n^i = \frac{n!}{(n-i)!\,i!}$$

where the numbers $C_n^i$, also denoted by $\binom{n}{i}$, are called binomial coefficients.

#### 2.5.6.6. *Subrings*

A subring of a unitary ring $(A, +, \cdot)$ is a non-empty subset $B$ of $A$ such that:

– $B$ is stable for addition and multiplication.

– $(B, +)$ is a subgroup of $(A, +)$.

– $1_A \in B$, which implies that $B$ is itself unitary.

EXAMPLE 2.9.– $\mathbb{Z}$ and $\mathbb{P}$ are subrings of $\mathbb{R}$.

$\mathbb{R}$ viewed as the set of numbers of the form $a + 0\,i$, with $a \in \mathbb{R}$ and $i^2 = -1$, is a subring of $\mathbb{C}$, the set of complex numbers of the form $a + b\,i$, with $a, b \in \mathbb{R}$.

#### 2.5.6.7. *Quotient rings*

The set $\mathbb{Z}/n\mathbb{Z} = \{k + n\mathbb{Z}, 0 \leq k \leq n-1\}$ is the quotient ring of $\mathbb{Z}$ consisting of integers congruent modulo $n \in \mathbb{N}^*$, and called the ring of integers modulo $n$, such that the difference between two arbitrary integers of this set is divisible by $n$. The equivalence relation corresponds here to the congruence on the integers. The set $\mathbb{Z}/n\mathbb{Z}$ equipped with the addition and multiplication operations such that:

$$(k + n\mathbb{Z}) + (p + n\mathbb{Z}) = (k + p) + n\mathbb{Z}$$
$$(k + n\mathbb{Z})(p + n\mathbb{Z}) = kp + n\mathbb{Z}$$

with the classes $n\mathbb{Z}$ and $1+n\mathbb{Z}$ as identity elements have a commutative ring structure for any $n \in \mathbb{N}^*$. The elements of $\mathbb{Z}/n\mathbb{Z}$ are called equivalence classes (or also congruence classes modulo $n$), each class $c_k, k \in [0, n-1]$, containing all the elements of $\mathbb{Z}$ having the same remainder on division by $n$.

For example, the class $c_3$ of $\mathbb{Z}/5\mathbb{Z}$ contains elements 3, 8, 13, etc.

NOTE 2.10.– Consider two integers $p$ and $q$ congruent modulo $n$, i.e. having the same remainder $r$ when divided by $n$, such that $p = \alpha\, n + r$ and $q = \beta\, n + r$, with $0 \leq r < n$. The congruence relation can be written as $p = k\, n + q$, with $k = \alpha - \beta$. It is linked with Euclidean division which consists in the division of two integers to produce a quotient and a remainder smaller than the divisor.

### 2.5.6.8. *Ideals*

Let $I$ be a subset of a ring $(A, +, \cdot)$. It is said that $I$ is an ideal of $A$ if it is an additive subgroup of $(A, +)$ such that the product of any element of $I$ by any element of $A$ is still an element of $I$, or equivalently, $I$ is closed under the multiplication by elements of $A$.

EXAMPLE 2.11.– In the commutative ring $\mathbb{Z}$, any set of the form $n\mathbb{Z}$, where $n$ is a non-zero natural number, is an ideal of $\mathbb{Z}$.

More precisely, in the case of a non-commutative unitary ring, a distinction can be made between right ideals, left ideals, and bi-ideals (i.e. both left and right ideals), also called two-sided ideals:

– A left (respectively, right; bi-) ideal $I$ of $A$ is an additive subgroup of $(A, +)$ such that: $\forall (x, y) \in A \times I$, we have $x\,y \in I$ (respectively, $y\,x \in I$; $x\,y$ and $y\,x \in I$). When $A$ is commutative, any right or left ideal is a bi-ideal.

– A left (respectively, right; bi-) ideal is said to be principal if it is generated by a single element $x \in A$, and we write this ideal: $(x) = A \cdot x = \{q\,x,\, q \in A\}$ (respectively, $x \cdot A$; $A \cdot x \cdot A$). This ideal is the set of multiples of $x$ in $A$. Consequently, for all $x, y \in A$, we have (x divides y) $\Leftrightarrow (y) \subset (x)$.

– It is said that $A$ is a principal ring if it is integral and all its ideals are principal.

– An intersection of ideals being an ideal, one can define the ideal generated by a set $\{x_1, \cdots, x_N\}$ of elements of $A$ as the set $(x_1, \cdots, x_N) = \left\{ \sum_{n=1}^{N} q_n\, x_n, q_n \in A,\, n \in \langle N \rangle \right\}$. It is then said that the ideal is of finite type because it is generated by a finite number of elements.

– Let $A$ be a ring with identity element 0. Then, $\{0\}$ and $A$ are ideals of $A$, called trivial ideals. Any non-trivial ideal is called a proper ideal.

The principal ring structure is used in arithmetic for defining the notions of GCD (greatest common divisor) and LCM (least common multiple), and therefore of decomposition into prime factors, and of Bézout's identity[12].

12 Étienne Bézout (1730–1783), French mathematician, famous for the books that he wrote for the teaching of mathematics, having experienced a great success, for many years, in France as well as in England and the United States. His research focused mainly on the theory of algebraic equations. He is more particularly known for the theorem and identity which bear his name, for the resolution of diophantine equations.

## 2.5.7. *Fields*

### 2.5.7.1. *Definition*

A field $(\mathbb{K}, +, \cdot)$ endowed with two internal operations (+ and $\cdot$), called addition and multiplication, is a unitary ring, namely, such that the axioms listed in Table 2.5 are satisfied for all $\lambda, \mu, \nu \in \mathbb{K}$, with the existence of a multiplicative identity element denoted by $1_{\mathbb{K}}$. Moreover, a field is such that all non-zero elements are invertible: $\forall \lambda \neq 0_{\mathbb{K}}$, $\exists \mu$ such that $\lambda\mu = \mu\lambda = 1_{\mathbb{K}}$, and therefore, $(\mathbb{K}^*, \cdot)$ is a group. On the other hand, $(\mathbb{K}, \cdot)$ cannot be a group because the identity element $0_{\mathbb{K}}$ such that $\lambda 0_{\mathbb{K}} = 0_{\mathbb{K}}\lambda = 0_{\mathbb{K}}, \forall \lambda \in \mathbb{K}$, does not admit any inverse $0_{\mathbb{K}}^{-1}$ such that $0_{\mathbb{K}}0_{\mathbb{K}}^{-1} = 0_{\mathbb{K}}^{-1}0_{\mathbb{K}} = 1_{\mathbb{K}}$. The inverse of $\lambda$ is generally denoted by $\lambda^{-1}$.

It should be noted that:

– There exists an identity element $1_{\mathbb{K}}$ for multiplication, which is not the case for a (non-unitary) ring.

– A field $(\mathbb{K}, +, .)$ is said to be commutative if the multiplication is commutative.

– Given that a field is an integral unitary ring, its characteristic is 0 or a prime number.

### 2.5.7.2. *Examples*

The following examples of fields are sets of numbers and the set of rational functions.

– *Numbers:* The sets $(\mathbb{P}, +, .), (\mathbb{R}, +, .), (\mathbb{C}, +, .)$ are commutative fields. Note that the set $(\mathbb{Z}, +, .)$ is not a field because its only invertible elements are 1 and $-1$.

– *Rational functions:* Analogously to the field $\mathbb{P}$ of rational numbers which is built from the ring $\mathbb{Z}$ of integers, every element of $\mathbb{P}$ being a fraction of two elements of $\mathbb{Z}$, the set of rational functions (in $z$) over $\mathbb{K}$, denoted by $\mathbb{K}(z)$, is built from the polynomial ring $A[z]$, each element of $\mathbb{K}(z)$ being a fraction of two elements of $A[z]$, i.e. two polynomials in $z$. Equipping the set $\mathbb{K}(z)$ with the internal addition and multiplication operations defined as:

$$\frac{p}{q} + \frac{r}{s} = \frac{ps + qr}{qs} \quad , \quad \frac{p}{q}.\frac{r}{s} = \frac{pr}{qs} \text{ with } q \neq 0 \text{ and } s \neq 0$$

for all rational functions $\frac{p}{q}, \frac{r}{s} \in \mathbb{K}(z)$, then $\mathbb{K}(z)$ has a field structure over $\mathbb{K}$.

#### 2.5.7.3. *Subfields*

A subset $\mathbb{L}$ of a field $\mathbb{K}$ is a subfield of $\mathbb{K}$ if:

– $\mathbb{L}$ is stable for addition and multiplication.

– $1_{\mathbb{K}} \in \mathbb{L}$.

– The induced laws on $\mathbb{L}$ by the laws of $\mathbb{K}$ induce a field structure for $\mathbb{L}$.

### 2.5.8. *Modules*

#### 2.5.8.1. *Definition*

A (left) module over a unitary ring $(A, +, .)$, called left $A$-module, is an Abelian additive group $(M, +_M)$ equipped with an external operation $(\times)$ acting on the left of $M$: $A \times M \to M$ , $A \times M \ni (a, x) \mapsto a \times x \in M$, which satisfies the following axioms for all $(a, b) \in A^2$ and all $(x, y) \in M^2$:

$$\begin{aligned} a \times (b \times x) &= (a.b) \times x \\ 1_A \times x &= x \\ (a + b) \times x &= a \times x +_M b \times x \\ a \times (x +_M y) &= a \times x +_M a \times y. \end{aligned}$$

Similarly, a right $A$-module is defined so that the external multiplication $(\times)$ acts on the right of $M$: $M \times A \to M$.

#### 2.5.8.2. *Submodules*

Let $M$ be a left $A$-module and $N \subset M$. It is said that $N$ is a left submodule of $M$ if and only if $N$ is a subgroup of $M$ stable for the external operation (scalar multiplication), that is, if for all $a \in A$ and all $y \in N$, we have $a \times y \in N$.

### 2.5.9. *Vector spaces*

#### 2.5.9.1. *Definition*

The assumption of linearity plays an important role in automatic control whose purpose is to model, analyze, identify, and control dynamic systems, as well as in signal processing to represent, analyze, and filter signals. On the one hand, this assumption leads to the very important class of linear dynamic systems, represented by way of input–output models, such as linear algebraic or differential equations, or linear state-space models. On the other hand, it facilitates the development of processing methods. It should be noted that the property of linearity is satisfied by several frequency domain transforms such as Fourier and Laplace transforms.

A problem is said to be linear if given two solutions $x$ and $y$, then for any real or complex number $\lambda$, $x + y$ and $\lambda x$ are also solutions. This assumption, which is also known as the superposition principle, is verified in first approximation by many physical phenomena and systems, which justifies its practical use. The mathematical framework of linear systems is the notion of v.s., also called linear space, which is based on the definition of two algebraic operations satisfying eight axioms.

In this chapter, we will pay particular attention to linear maps that are very useful in practice because of their simplicity, and to multilinear maps that will play an important role in Chapter 6, for the definition of hypermatrices and tensors.

Let $E$ be a set and $(\mathbb{K}, +, .)$ a field, whose elements are, respectively, called vectors and scalars. One defines an internal operation denoted by $+_E$, named vector addition and corresponding to an operation between elements of $E$, and an external operation denoted by $\circ$, called scalar multiplication, which is an operation between elements of $E$ and $\mathbb{K}$. The set $E$ is a v.s. over $\mathbb{K}$, denoted by $E(\mathbb{K})$, if the axioms described in Table 2.6 are satisfied for all $x, y, z \in E$, and all $\lambda, \mu \in \mathbb{K}$. A v.s. over $\mathbb{K}$ is called a $\mathbb{K}$-vector space ($\mathbb{K}$-v.s.). When $\mathbb{K} = \mathbb{R}$ (or $\mathbb{C}$), $E$ is referred to as a real (or complex) vector space.

| **Axioms** | **Properties** |
|---|---|
| $x +_E y \in E$ | closure under the $+_E$ operation |
| $x +_E y = y +_E x$ | $+_E$ is commutative |
| $x +_E (y +_E z) = (x +_E y) +_E z$ | $+_E$ is associative |
| $x +_E 0_E = 0_E +_E x = x$ | $0_E$ is the identity element for $+_E$ |
| $x +_E (-x) = (-x) +_E x = 0_E$ | $(-x)$ is the opposite of $x$ |
| $\lambda \circ x \in E$ | closure under the $\circ$ operation |
| $\lambda \circ (\mu \circ x) = (\lambda.\mu) \circ x$ | $\circ$ is associative |
| $1_{\mathbb{K}} \circ x = x$ | $1_{\mathbb{K}}$ is the identity element for $\circ$ |
| $(\lambda + \mu) \circ x = \lambda \circ x +_E \mu \circ x$ | $\circ$ is distributive over the scalar addition $+$ |
| $\lambda \circ (x +_E y) = \lambda \circ x +_E \lambda \circ y$ | $\circ$ is distributive over the vector addition $+_E$ |

**Table 2.6.** *Axiomatic definition of a vector space*

Based on Table 2.6, the following comments can be made:

– In addition to the closure property of $E$ under the operations $+_E$ and $\circ$, the v.s. structure is defined based on eight axioms: the first four are related to the internal law $+_E$, the following two to the external law $\circ$, whereas the last two characterize the compatibility of the two laws based on the distributivity property.

– The set $(E, +_E)$ is a commutative group. This property is associated with the first five axioms of Table 2.6. On the other hand, the last five axioms are identical to those defining a module (see section 2.5.8). It can be concluded that a v.s. is a module over a field $\mathbb{K}$, in other words a $K$-module.

– Based on the definition axioms, we have for all $x \in E$ and $\lambda \in \mathbb{K}$:

$$0_{\mathbb{K}} \circ x = \lambda \circ 0_E = 0_E, \qquad [2.3]$$

and:

$$\lambda \circ x = 0_E \;\Rightarrow\; \lambda = 0_{\mathbb{K}} \text{ or } x = 0_E. \qquad [2.4]$$

These equalities [2.3] and this implication [2.4] make that in general $0_{\mathbb{K}}$ and $0_E$ are represented by the same symbol $0$. In the following, the symbol $\circ$ of the external operation will be omitted, one will write $\lambda x$ instead of $\lambda \circ x$, and the symbol $+_E$ will be replaced by $+$ to simplify the writing of equations.

– The definition of a v.s. requires to define the addition of two vectors, but not the product of two vectors. The introduction of a multiplicative internal law results in the algebra structure (see section 2.5.16).

– By definition, a $\mathbb{K}$-v.s. $E$ contains all linear combinations of its own vectors, which means that if $x_1, \cdots, x_P$ are vectors of $E$, then $\sum_{p=1}^{P} \lambda_p x_p$ is also in $E$, for all $\lambda_p \in \mathbb{K}$. That explains the other designation as a linear space. So, a v.s. can be viewed as a set of vectors, endowed with two operations (vector addition and scalar multiplication) allowing for the definition of linear combinations of vectors of the set.

### 2.5.9.2. *Examples*

In this section, we describe several examples of vector spaces corresponding to sets of Euclidean vectors, numbers, polynomials, and functions. The v.s. of linear maps and multilinear maps will be detailed in the following two sections.

– Set of Euclidian geometry[13] vectors of the plane $(\mathbb{R}^2)$ and of the three-dimensional space $(\mathbb{R}^3)$, equipped with the usual operations on vectors. These examples are at the origin of the name "vector space."

13 Euclid, born around 325 BC, is the most prominent Greek mathematician of Antiquity, author of his treatise on geometry and number theory, named *The Elements*. He is more particularly known for the so-called Euclidean geometry originally based on Euclid's list of axioms, and Euclid's algorithm for finding the GCD of two natural numbers.

– $\mathbb{C}$ = set of complex numbers, with $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, equipped with the following two operations for $p, q \in \mathbb{C}$, $a, b, c, d \in \mathbb{R}$, $\lambda \in \mathbb{K}$, $i^2 = -1$:

$$p + q = (a + b\,i) + (c + d\,i) = (a + c) + (b + d)\,i$$
$$\lambda p = \lambda(a + b\,i) = \lambda a + \lambda b\,i$$

– $\mathbb{Q}$ = set of (real) quaternions; $\mathbb{K} = \mathbb{R}$. Addition and scalar multiplication operations are defined for $p, q \in \mathbb{Q}$, $\lambda \in \mathbb{R}$ as:

$$\begin{aligned} p + q &= (p_0 + p_1 i + p_2 j + p_3 k) + (q_0 + q_1 i + q_2 j + q_3 k) \\ &= p_0 + q_0 + (p_1 + q_1)i + (p_2 + q_2)j + (p_3 + q_3)k \\ \lambda q &= \lambda(q_0 + q_1 i + q_2 j + q_3 k) = \lambda q_0 + \lambda q_1 i + \lambda q_2 j + \lambda q_3 k \end{aligned}$$

with $p_n, q_n \in \mathbb{R}$, $n \in \{0, 1, 2, 3\}$, and $i^2 = j^2 = k^2 = ijk = -1$.

– $\mathbb{K}^N$ = set of all ordered $N$-tuples $\mathbf{x} = (x_1, \cdots, x_N)$ whose components are real ($\mathbb{K} = \mathbb{R}$) or complex ($\mathbb{K} = \mathbb{C}$) numbers, which can be written as column vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$, with $x_n \in \mathbb{K}, n \in \langle N \rangle$, and generalizing the sets $\mathbb{R}^2$ and $\mathbb{R}^3$ previously introduced. Addition and scalar multiplication operations are defined for $\mathbf{x}, \mathbf{y} \in \mathbb{K}^N, \lambda \in \mathbb{K}$ as:

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_N + y_N \end{bmatrix}, \quad \lambda \mathbf{x} = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_N \end{bmatrix}.$$

– The Cartesian product $\overset{N}{\underset{n=1}{\times}} E_n$ of $N$ $\mathbb{K}$-v.s. $E_n$ is a $\mathbb{K}$-v.s. for the operations defined above, with $x_n, y_n \in E_n, n \in \langle N \rangle$.

– $\mathbb{K}_N[z]$ = set of polynomials in one variable $(z)$, of degree $\leq N$, whose coefficients belong to the field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, of the form $p(z) = \sum_{n=0}^{N} a_n z^n$, with the following operations ($\lambda \in \mathbb{K}$):

$$(\sum_n a_n z^n) + (\sum_n b_n z^n) = \sum_n (a_n + b_n) z^n$$
$$\lambda(\sum_n a_n z^n) = \sum_n (\lambda a_n) z^n.$$

– $\mathbb{K}_N[z_1, \cdots, z_M]$ = set of homogeneous polynomials in $M$ variables $(z_1, \cdots, z_M)$, of degree $N$, of the form $p(z_1, \cdots, z_M) = \sum_{k_1, \cdots, k_M} a_{k_1, \cdots, k_M} z_1^{k_1} \cdots z_M^{k_M}$, with $\sum_{m=1}^{M} k_m = N$, that is, such that all non-zero monomials have the same degree $N$.

– $\mathcal{F}(\mathbb{K}^M, \mathbb{K}^N)$ = set of vector functions from $\mathbb{K}^M$ to $\mathbb{K}^N$ : $\mathbb{K}^M \ni \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) \in \mathbb{K}^N$, with the following operations, for all $\mathbf{x} \in \mathbb{K}^M$ and $\lambda \in \mathbb{K}$:

$$(\mathbf{f} + \mathbf{g})(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})$$
$$(\lambda \mathbf{f})(\mathbf{x}) = \lambda \mathbf{f}(\mathbf{x}).$$

– $\mathcal{C}^k(I, \mathbb{K})$ = set of functions of class $\mathcal{C}^k$, defined[14] on an interval $I$ of $\mathbb{R}$ and with values in $\mathbb{K}$, with the usual operations (addition and scalar multiplication) on scalar functions of a scalar variable.

– $\mathcal{C}^0_T(\mathbb{R}, \mathbb{C}) = \{f : \mathbb{R} \to \mathbb{C}, \forall t \in \mathbb{R}, f(t+T) = f(t)\}$ = set of continuous scalar functions, $T$-periodic[15], from $\mathbb{R}$ to $\mathbb{C}$, with the operations of addition and scalar multiplication such that for all $t \in \mathbb{R}$ and $\lambda \in \mathbb{K}$, we have:

$$(f + g)(t) = f(t) + g(t)$$
$$(\lambda f)(t) = \lambda f(t).$$

The quantities $T$, $F = 1/T$, and $\omega = 2\pi F$ are the period in second the frequency in hertz, and the angular frequency in radian per second. Note that if $f$ is $2\pi$-periodic on $\mathbb{R}$, then the function $g$ defined by $g(t) = f(\frac{2\pi}{T}t)$ is $T$-periodic. Indeed, we have:

$$g(t+T) = f\Big(\frac{2\pi}{T}(t+T)\Big) = f(\frac{2\pi}{T}t + 2\pi) = f(\frac{2\pi}{T}t) = g(t). \qquad [2.5]$$

This explains that periodic functions are generally considered as $2\pi$-periodic, the shift to a $T$-periodic function being achieved by the transformation of $t$ into $\frac{2\pi}{T}t$. This will be the case in section 3.8 where we shall consider the v.s. $\mathcal{C}^0_{2\pi}(\mathbb{R}, \mathbb{C})$ for the Fourier series expansion of $2\pi$-periodic functions.

– Matrix and tensor vector spaces will be more particularly addressed in this book, matrices and tensors being introduced in Chapters 4 and 6, respectively.

---

14 A function is said to be of class $\mathcal{C}^k$ on $I$ if and only if it is $k$-times differentiable, and if its derivatives of order less than or equal to $k$ are continuous. The sets $\mathcal{C}^0$ and $\mathcal{C}^\infty$ denote, respectively, the sets of continuous functions and indefinitely differentiable functions, whose successive derivatives are all continuous.

15 A $T$-periodic function, which means of period $T$, is entirely defined by its restriction on any interval of length $T$ as $[-T/2, T/2]$, $[0, T]$ or $[a, a+T]$, $\forall a \in \mathbb{R}$, and we have $\int_{-T/2}^{T/2} f(t)dt = \int_0^T f(t)dt = \int_a^{a+T} f(t)dt$. With the change of variable $t = s - T$ and using the periodicity property $f(t+T) = f(t)$, it can be deduced that $\int_{a+T}^{b+T} f(s)ds = \int_a^b f(t)dt$.

## 2.5.10. *Vector spaces of linear maps*

### 2.5.10.1. *Linear maps*

A linear map $f$ from a $\mathbb{K}$-v.s. $E$ to a $\mathbb{K}$-v.s. $F$ is a map: $E \ni x \mapsto y = f(x) \in F$ that linearly transforms a vector of $E$ into a vector of $F$, that is, satisfying the following properties for all $(x, y) \in E^2$ and $\alpha \in \mathbb{K}$:

$$f(x+y) = f(x) + f(y)$$
$$f(\alpha\, x) = \alpha f(x),$$

which is equivalent to:

$$f(\alpha x + y) = \alpha f(x) + f(y). \qquad [2.6]$$

It is said that $f$ preserves linear combinations. $E$ and $F$ are called the domain and the codomain of $f$, respectively.

### 2.5.10.2. *Properties*

PROPOSITION 2.12.– *Linear maps satisfy the following properties:*

*– The set*[16] *of all linear maps from $E$ to $F$, denoted by $\mathcal{L}(E, F)$, endowed with the addition and scalar multiplication operations such that, for all $x \in E$ and $\alpha \in \mathbb{K}$:*

$$(f_1 + f_2)(x) = f_1(x) + f_2(x) \qquad [2.7a]$$
$$(\alpha f)(x) = \alpha f(x) \qquad [2.7b]$$

*form a vector space over $\mathbb{K}$; in other words, a $\mathbb{K}$-v.s.*

*– If $E$ and $F$ are of finite dimension, then:*

$$dim[\mathcal{L}(E, F)] = dim(E)dim(F). \qquad [2.8]$$

*– The composition $g \circ f$ of two linear maps is a linear map.*

*– A linear map $f \in \mathcal{L}(E, F)$ is entirely determined by the images by $f$ of basis vectors of $E$. This result will be used in Chapter 4 to determine the matrix associated to a linear map. For $f \in \mathcal{L}(E, F)$, with $dim(E) = J$, $dim(F) = I$, the associated matrix has dimensions $I \times J$.*

*– When $F = E$, a linear map from $E$ to itself is called an endomorphism of $E$. The set of all these endomorphisms is denoted by $\mathcal{L}(E)$, and $(\mathcal{L}(E), +, .)$ is a $\mathbb{K}$-v.s. See sections 2.5.16.3 and 2.6.3.*

16 This set will also be denoted $\mathbb{L}_{\mathbb{K}}(E, F)$ in section 4.13.

#### 2.5.10.3. *Linear forms*

A linear map from $E$ to $\mathbb{K}$, i.e. with values in $\mathbb{K}$, is called a linear form on $E$.

The set of linear forms on $E$, denoted by $E^* = \mathcal{L}(E, \mathbb{K})$, is called the dual space of $E$. This space equipped with the two operations, previously defined, is also a $\mathbb{K}$-v.s., with $\dim(E^*) = \dim(E)\dim(\mathbb{K}) = \dim(E)$. The dual of $E^*$, denoted by $E^{**}$, is called the bidual of $E$.

### 2.5.11. *Vector spaces of multilinear maps*

#### 2.5.11.1. *Multilinear maps*

Let $(E_1, \cdots, E_N; F)$ be $N+1$ vector spaces over $\mathbb{K}$. The map $f$ from $\underset{n=1}{\overset{N}{\times}} E_n$ to $F$ is $N$-linear (or multilinear) if and only if for all $n \in \langle N \rangle$, the map $E_n \to F : x \mapsto f(u_1, \cdots, u_{n-1}, x, u_{n+1}, \cdots u_N)$ is linear, which means that $f$ is linear separately in each variable $u_n$, all the other variables being held constant.

The set of $N$-linear maps, denoted by $\mathcal{ML}(E_1, \cdots, E_N; F)$, form a $\mathbb{K}$-v.s. If all of the v.s. $E_n$, with $n \in \langle N \rangle$, and $F$ are of finite dimension, we then have:

$$\dim[\mathcal{ML}(E_1, \cdots, E_N; F)] = \dim(F) \prod_{n=1}^{N} \dim(E_n). \qquad [2.9]$$

FACT 2.13.– For $(u_1, \cdots, u_N) \in \underset{n=1}{\overset{N}{\times}} E_n$ and $(\lambda_1, \cdots, \lambda_N) \in \mathbb{K}^N$, we have:

$$f(\lambda_1 u_1, \cdots, \lambda_N u_N) = (\prod_{n=1}^{N} \lambda_n) f(u_1, \cdots, u_N).$$

FACT 2.14.– In the particular case where $E_n = E$, $\forall n \in \langle N \rangle$, the map $f : E^N \to F$ is $N$-linear from $E$ to $F$, which means that for all $n \in \langle N \rangle$, the $N$ partial maps

$$E \ni x \ \mapsto \ f(u_1, \cdots, u_{n-1}, x, u_{n+1}, \cdots, u_N) \in F$$

are linear. The set of $N$-linear maps from $E^N$ to $F$ is denoted by $\mathcal{ML}_N(E, F)$.

#### 2.5.11.2. *Multilinear forms*

An $N$-linear (or multilinear) form is a special case of $N$-linear map for which $F = \mathbb{K}$.

All of the $N$-linear forms from $\underset{n=1}{\overset{N}{\times}} E_n$ to $\mathbb{K}$, denoted $\mathcal{ML}(E_1, \cdots, E_N; \mathbb{K})$, is a $\mathbb{K}$-v.s. of dimension $\prod_{n=1}^{N} \dim(E_n)$.

A multilinear form $f \in \mathcal{ML}(E_1, \cdots, E_N; \mathbb{K})$ is said to be decomposable if it can be decomposed into a product $f = f_1 ... f_N$, with $f_n \in E_n^*, n \in \langle N \rangle$, and

$$\underset{n=1}{\overset{N}{\times}} E_n \ni (u_1, \cdots, u_N) \mapsto f(u_1, \cdots, u_N) = f_1(u_1) ... f_N(u_N) \in \mathbb{K}.$$

FACT 2.15.– In the particular case where $E_n = E$, $\forall n \in \langle N \rangle$, the form: $f : E^N \to \mathbb{K}$ is said to be $N$-linear over $E$, and the set of $N$-linear forms over $E$ is denoted by $\mathcal{ML}_N(E, \mathbb{K})$.

### 2.5.11.3. *Symmetric/alternating multilinear maps/forms*

A map (respectively, form) $f \in \mathcal{ML}_N(E, F)$ (respectively, $\mathcal{ML}_N(E, \mathbb{K})$) is said to be:

– symmetric if and only if for all $(u_1, \cdots, u_N) \in \underset{n=1}{\overset{N}{\times}} E_n$, and for any permutation $\pi \in \mathcal{S}_\mathcal{N}$, where $\mathcal{S}_\mathcal{N}$ is the symmetric group of order $N$, we have:

$$f(u_{\pi(1)}, \cdots, u_{\pi(N)}) = f(u_1, \cdots, u_N)$$

that is, $f$ remains unchanged under any permutation of its variables $u_n$;

– antisymmetric (or skew-symmetric) if and only if:

$$f(u_1, \cdots, u_n, \cdots, u_p, \cdots, u_N) = -f(u_1, \cdots, u_p, \cdots, u_n, \cdots, u_N),$$

that is, $f$ changes sign if two variables $u_n$ and $u_p$ are permuted;

– alternating if and only if there exists $n \neq p$ such that

$$u_n = u_p \;\Rightarrow\; f(u_1, \cdots, u_N) = 0, \qquad [2.10]$$

that is, $f$ is cancelled if $u_n = u_p$ for at least two distinct indices $n$ and $p$.

PROPOSITION 2.16.– *Any anti-symmetric $N$-linear form $f \in \mathcal{ML}_N(E, \mathbb{K})$ satisfies the following property: for any permutation $\pi \in \mathcal{S}_\mathcal{N}$, and for all $(u_1, \cdots, u_N) \in E^N$, we have*

$$f(u_{\pi(1)}, \cdots, u_{\pi(N)}) = \sigma(\pi) f(u_1, \cdots, u_N)$$

*where $\sigma(\pi)$ designates the signature of the permutation $\pi$ (see section 2.5.5.4).*

### 2.5.11.4. *Bilinear maps and bilinear forms*

For $N = 2$, we have the set $\mathcal{BL}(E_1, E_2; F)$ of bilinear maps, that is, the set of maps $f : E_1 \times E_2 \to F$ such that, for all $u_1 \in E_1$ and $u_2 \in E_2$, the maps $y \mapsto f(u_1, y)$ and $x \mapsto f(x, u_2)$ are linear, or more specifically:

$$\forall (u_1, y_1, y_2) \in E_1 \times E_2^2 \,, \forall \alpha \in \mathbb{K} \,, f(u_1, \alpha y_1 + y_2) = \alpha f(u_1, y_1) + f(u_1, y_2)$$

$$\forall (x_1, x_2, u_2) \in E_1^2 \times E_2 \,, \forall \alpha \in \mathbb{K} \,, f(\alpha x_1 + x_2, u_2) = \alpha f(x_1, u_2) + f(x_2, u_2).$$

The set $\mathcal{BL}(E_1, E_2; F)$ is a $\mathbb{K}$-v.s.

A bilinear form is a special case of bilinear map such that $F = \mathbb{K}$.

#### 2.5.11.5. *Composition of two linear maps*

Let $E$, $F$, and $G$ be three $\mathbb{K}$-v.s., and $g \circ f$ be the composition of two linear maps $f$ and $g$ such that:

$$\mathcal{L}(E,F) \times \mathcal{L}(F,G) \to \mathcal{L}(E,G) \ : \ (f,g) \mapsto \varphi(f,g) = g \circ f.$$

It defines a bilinear map, that is, linear with respect to $f$ and $g$. Indeed, for all $f, f_1, f_2 \in \mathcal{L}(E,F)$, $g, g_1, g_2 \in \mathcal{L}(F,G)$ and $\alpha \in \mathbb{K}$, we have:

$$\varphi(\alpha f_1 + f_2, g) = g \circ (\alpha f_1 + f_2) = \alpha(g \circ f_1) + (g \circ f_2) = \alpha\varphi(f_1, g) + \varphi(f_2, g)$$
$$\varphi(f, \alpha g_1 + g_2) = (\alpha g_1 + g_2) \circ f = \alpha(g_1 \circ f) + (g_2 \circ f) = \alpha\varphi(f, g_1) + \varphi(f, g_2).$$

### 2.5.12. *Vector subspaces*

#### 2.5.12.1. *Definition and properties*

The notion of vector subspace, also called linear subspace or simply subspace, is fundamental. As we shall see in the following section, a subspace can be described by means of a generator, or more precisely by a set of vectors called a system of generators.

A non-empty subset $F$ of a $\mathbb{K}$-v.s. $E$ is a subspace of $E$ if and only if for all $x, y \in F$ and all $\lambda \in \mathbb{K}$, we have:

$$x + y \in F \ , \ \lambda x \in F$$

or equivalently:

$$\forall x, y \in F \ , \ \forall \lambda \in \mathbb{K} \ \Rightarrow \ \lambda x + y \in F.$$

It is said that a subspace is closed under addition and scalar multiplication. As a consequence, to show that a subset of $E$ is a subspace, it is not necessary to verify the eight definition axioms of a v.s. Only closure conditions are to be verified. It should be noted that the identity element $0_E$ of $E$ coincides with the identity element $0_F$ of any subspace $F$.

A subspace $F$ of the v.s. $E$ is said to be proper if it is smaller than $E$, i.e. if $F \neq E$.

Subspaces satisfy the following properties, for two subspaces $F$ and $G$ of $E$:

– $F \cap G$ is a subspace of $E$.

– $F \cup G$ in general is not a subspace of $E$.

These properties can be generalized to $N$ subspaces $F_n, n \in \langle N \rangle$, of $E$.

EXAMPLE 2.17.– Hereafter, we give two examples of subspaces relative to function spaces that will be considered in Chapter 3.

– The set $\mathcal{C}_m^0([a,b],\mathbb{K}) = \{f : [a,b] \to \mathbb{K},\ [a,b] \ni t \mapsto f(t) \in \mathbb{K}\}$ of piecewise continuous functions[17], defined on the interval $[a,b] \subseteq \mathbb{R}$ and with values in $\mathbb{K}$, is a subspace of the $\mathbb{K}$-v.s. of functions from $[a,b]$ to $\mathbb{K}$.

– Similarly, for the set $\mathcal{C}_m^k([a,b],\mathbb{K})$ of functions $f$ of class $\mathcal{C}^k$ such that $f$ and its first $k$ derivatives are piecewise continuous, or in other words, continuous except for a finite number of points of discontinuity of the first kind.

### 2.5.12.2. *Systems of generators*

Given a set of vectors $\mathcal{X} = \{x_1, \cdots, x_P\}$ of the $\mathbb{K}$-v.s. $E$, then the set:

$$F = \{y \in E : y = \sum_{p=1}^{P} \lambda_p x_p\,,\ \forall \lambda_1, \cdots, \lambda_P \in \mathbb{K}\}$$

is called the subspace of $E$ spanned by $\mathcal{X}$, or the subspace of linear combinations of $x_1, \cdots, x_P$. The set of all linear combinations of vectors from $\mathcal{X}$ is called the linear span (or just span) of $\mathcal{X}$, and it is denoted by $\text{Span}(\mathcal{X})$.

The vectors $x_1, \cdots, x_P$, called the spanning vectors, form a system of generators or a generator of $E$ if any vector $y$ of $E$ can be written as a linear combination (lc) of vectors $x_1, \cdots, x_P$, that is, if there exists scalars $\lambda_1, \cdots, \lambda_P$ such that $y = \sum_{p=1}^{P} \lambda_p x_p$. Then, we have $\text{Span}(\mathcal{X}) = E$. It is also written $E = \text{lc}(\mathcal{X})$, or $E = \text{Vect}(\mathcal{X})$.

### 2.5.12.3. *Affine and convex combinations*

Given a set $\{x_1, \cdots, x_P\}$ of $P$ vectors, and a set $\{\lambda_1, \cdots, \lambda_P\}$ of $P$ scalars satisfying $\sum_{p=1}^{P} \lambda_p = 1$, the vector $x = \sum_{p=1}^{P} \lambda_p x_p$ is called an affine combination

17 A function $f : \mathbb{R} \to \mathbb{K}$ is continuous at $x_0 \in \mathbb{R}$ if for all $\epsilon > 0$ there exists a $\delta > 0$ such that if $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < \epsilon$, namely, the value of $f$ at a point close to $x_0$ is close to $f(x_0)$. If $f$ is continuous at any point $x_0 \in [a,b] \subseteq \mathbb{R}$, then $f$ is said to be continuous on $[a,b]$.

The function $f : [a,b] \to \mathbb{K}$ is piecewise continuous on an interval $[a,b] \subseteq \mathbb{R}$ if and only if the interval can be partitioned into a finite number of points $a_n$, with $a = a_0 < a_1 < \cdots < a_{N-1} < a_N = b$, such that, for any non-negative integer $n \in \langle N \rangle$, the restriction of $f$ on $]a_{n-1}, a_n[$ is continuous, with a finite right limit $f(a_{n-1}^+)$ at the point $a_{n-1}$, and a finite left limit $f(a_n^-)$ at the point $a_n$. In other words, $f$ is piecewise continuous if it is continuous except at a finite number of points of discontinuity of the first kind, that is, $f(a_n^+)$ and $f(a_n^-)$ exist but are different.

of $x_1, \cdots, x_P$, with some of the $\lambda_p$ that may be negative. When all scalars $\lambda_p$ are non-negative, the vector $x$ is then called a convex combination of $x_1, \cdots, x_P$.

A set $C$ is said to be convex if it is closed under convex combinations, that is, if for all $x$ and $y$ in $C$ and all $t \in [0, 1]$, the point $t\,x + (1 - t)\,y$ also belongs to $C$. In geometry, a convex set is a set of points such that the line joining any two points of the set lies entirely within that set.

By definition, any subspace $C$ of a $\mathbb{R}$-v.s. is a convex set.

#### 2.5.12.4. *Linear independence/dependence*

Given a $\mathbb{K}$-v.s. $E$, it is said that a set of vectors $\{x_1, \cdots, x_P\}$ is free if:

$$\sum_{p=1}^{P} \lambda_p x_p = 0 \;\Rightarrow\; \lambda_1 = \cdots = \lambda_P = 0.$$

The vectors $x_1, \cdots, x_P$ are said to be linearly independent (or simply independent).

If there exists a set of scalars $\{\lambda_1, \cdots, \lambda_P\}$ not all zero such that $\sum_{p=1}^{P} \lambda_p x_p = 0$, the system $\{x_1, \cdots, x_P\}$ is said to be linked, and the vectors are said to be linearly dependent (or dependent). In this case, at least one of the vectors can be expressed as a linear combination of the others.

FACT 2.18.– In $\mathbb{R}^m$, a set of $n$ vectors $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ of dimension $m$ is always linearly dependent if $n > m$. This result can be interpreted in terms of an homogeneous system of linear equations $\mathbf{Ab} = \mathbf{0}$, where $\mathbf{A}$ is the $m \times n$ matrix whose columns are the vectors $\mathbf{x}_1, \cdots, \mathbf{x}_n$, and $\mathbf{b}$ is the vector of unknowns. The case $n > m$ corresponds to an underdetermined system, i.e. with more unknowns $(n)$ than equations $(m)$. Such a system has the trivial solution where all the unknowns are equal to zero, and an infinity of non-trivial solutions which form a vector space whose dimension is $n - r$, where $r$ is the rank of the matrix $\mathbf{A}$, that is, the maximal number of independent columns of $\mathbf{A}$ (See Table 4.7).

### 2.5.13. *Bases*

#### 2.5.13.1. *Definition*

A set $\mathcal{B} = \{b_1, \cdots, b_P\}$ of a $\mathbb{K}$-v.s. $E$ is a basis for $E$ if and only if $\mathcal{B}$ is both free and generator of $E$. In other words, a basis of a v.s. $E$ is a set of linearly independent vectors that spans $E$. This notion is fundamental in linear algebra.

Any vector $x$ of $E$ can then be written in a unique way as a linear combination of the vectors of $\mathcal{B}$, called basis vectors, that is:

$$\forall x \in E,\ \exists (\lambda_1, \cdots, \lambda_P) \in \mathbb{K}^P \text{ such that } x = \sum_{p=1}^{P} \lambda_p b_p.$$

The scalars $\lambda_1, \cdots, \lambda_P$ are called the coordinates (or the components) of $x$ in the basis $\mathcal{B}$. When such a basis exists, it is said that $E$ is a finite-dimensional space, of dimension $P$, and it is written that $\dim(E) = P$. It should be noted that any $\mathbb{K}$-v.s. of finite dimension admits a basis. As already mentioned, the study of linear maps between finite-dimensional vector spaces is one the main topics of linear algebra.

### 2.5.13.2. *Properties*

We have the following properties:

– A finite-dimensional v.s. $E$ has infinitely many different bases. All bases have the same number of elements, equal to $\dim(E) = P$. The dimension of a basis is an invariant[18] of a finite-dimensional v.s.

– If $E$ has dimension $P$, any free system of $P$ vectors of $E$ is a basis, and it is impossible to find more than $P$ independent vectors in $E$.

– Any subspace $F$ of $E$ is such that $\dim(F) \leq \dim(E)$, the equality taking place if and only if $F = E$.

– *Incomplete basis theorem:* Any free system of a finite-dimensional v.s. $E$ can be completed to form a basis of $E$.

### 2.5.13.3. *Examples of canonical bases*

Some bases are said to be canonical (or standard). For instance:

– For $\mathbb{R}^N$, a canonical basis consists of the following $N$ vectors[19]

$$\mathbf{e}_n = [0, \cdots 0, 1, 0, \cdots 0]^T \ , \ n \in \langle N \rangle \qquad [2.11]$$

where the $n$th component is equal to 1 and all others to 0. We have $\dim(\mathbb{R}^N) = N$. In the sequel, these basis vectors will be denoted by $\mathbf{e}_n^{(N)}$, with $n \in \langle N \rangle$.

– For $\mathbb{C}^N$, two canonical bases can be defined depending on whether $\mathbb{C}^N$ is a real v.s. (i.e. with coefficients in $\mathbb{K} = \mathbb{R}$) or a complex v.s. (namely, with coefficients in $\mathbb{K} = \mathbb{C}$).

For $\mathbb{K} = \mathbb{R}$, a canonical basis is:

$$\mathbf{e}_{2k-1} = [0, \cdots 0, 1, 0, \cdots 0]^T \ , \ \mathbf{e}_{2k} = [0, \cdots 0, \, i, \, 0, \cdots 0]^T, \, k \in \langle N \rangle,$$

18 An invariant is a property satisfied by a set of mathematical objects, that is preserved when certain transformations are applied to the objects. For a given finite-dimensional v.s. $E$, its dimension is an invariant because it is preserved by any isomorphism from $E$ to another finite-dimensional v.s. $F$. For a square matrix, its determinant, trace, rank, and characteristic polynomial are similarity invariants because they remain unchanged under changes of basis, that is, these quantities are the same for two similar matrices (see section 4.13.5).

19 The transposition operator, denoted using the superscript $^T$, transforms a row vector into a column vector and a column vector into a row vector.

where 1 and $i$ are at the $k$th position in $\mathbf{e}_{2k-1}$ and $\mathbf{e}_{2k}$, respectively, and $i^2 = -1$.

The coordinates of any vector of $\mathbb{C}^N$ in this basis are then real numbers, and we have $\dim_{\mathbb{R}}(\mathbb{C}^N)$ =2N.

For example, for $\mathbb{C}^3$, this canonical basis consists of the following six vectors:

$$\{[1,0,0]^T, [i,0,0]^T, [0,1,0]^T, [0,i,0]^T, [0,0,1]^T, [0,0,i]^T\}. \quad [2.12]$$

For $\mathbb{K} = \mathbb{C}$, a canonical basis of $\mathbb{C}^N$ is given by the canonical basis [2.11] of $\mathbb{R}^N$. The coordinates of a vector in this basis are then complex numbers, and we have $\dim_{\mathbb{C}}(\mathbb{C}^N) = N$.

– For $\mathbb{R}_N[z]$, the set of polynomials in the variable $z$, of degree $\leq N$ and with real coefficients, the set of monomials $\{1, z, z^2, \cdots, z^N\}$ constitutes a standard basis. We thus have $\dim(\mathbb{R}_N[z]) = N + 1$.

#### 2.5.13.4. *Basis of a Cartesian product of vector spaces*

Let $N$ $\mathbb{K}$-v.s. $E_n, n \in \langle N \rangle$, of finite dimension $I_n$. The Cartesian product $\underset{n=1}{\overset{N}{\times}} E_n$ is a $\mathbb{K}$-v.s. of dimension:

$$\dim(\underset{n=1}{\overset{N}{\times}} E_n) = \sum_{n=1}^{N} \dim(E_n) = \sum_{n=1}^{N} I_n$$

having the following basis:

$$\{(b_1^{(1)}, 0, \cdots, 0), \cdots, (b_{I_1}^{(1)}, 0, \cdots, 0), \cdots, (0, \cdots, 0, b_1^{(N)}), \cdots, (0, \cdots, 0, b_{I_N}^{(N)})\}$$

where $\{b_1^{(n)}, \cdots, b_{I_n}^{(n)}\}$ is a basis of $E_n$. This is illustrated by the basis of the space $\mathbb{C}^3$ previously presented, with $\{b_1, b_2\} = \{1, i\}$ a basis of $\mathbb{C}$ over $\mathbb{R}$, any $x \in \mathbb{C}$ being written as $x = a + bi$, with $a, b \in \mathbb{R}$, and $i^2 = -1$. The canonical basis for $\mathbb{C}^3$ is then given by [2.12], or equivalently:

$$\{[b_1,0,0]^T, [b_2,0,0]^T, [0,b_1,0]^T, [0,b_2,0]^T, [0,0,b_1]^T, [0,0,b_2]^T\}.$$

### 2.5.14. *Sum and direct sum of subspaces*

#### 2.5.14.1. *Definitions*

– Let $E_1$ and $E_2$ be two subspaces of $E$. We call sum of $E_1$ and $E_2$, denoted by $E_1 + E_2$, the set of all sums of a vector of $E_1$ and a vector of $E_2$, that is[20]:

$$E_1 + E_2 = \{x_1 + x_2, x_1 \in E_1 \text{ and } x_2 \in E_2\}.$$

Note that $E_1 + E_2$ is a subspace of $E$ of finite dimension equal to:

20 It should be noted that the same sign + is used to denote the sum of two vector subspaces and the union of two sets, the context allowing any ambiguity to be removed.

$$\dim(E_1 + E_2) = \dim(E_1) + \dim(E_2) - \dim(E_1 \cap E_2).$$

This equation that expresses the dimension of the sum of two subspaces of the same v.s. is called the Grassmann formula. From this equation, we deduce that $\dim(E_1 + E_2) \leq \dim(E_1) + \dim(E_2)$, with equality if and only if $E_1 \cap E_2 = \{0\}$, namely, when $E_1$ and $E_2$ are disjoint (or independent). That leads to the notion of direct sum of subspaces, as defined below.

– When the decomposition of any element of $E$ into the sum of an element of $E_1$ and an element of $E_2$ is unique, it is said that $E$ is the direct sum of $E_1$ and $E_2$, written as $E = E_1 \oplus E_2$, and $E_1$ and $E_2$ are called complement subspaces of $E$.

We have:

$$E = E_1 \oplus E_2 \Leftrightarrow \begin{cases} E = E_1 + E_2 \\ \text{and} \\ E_1 \cap E_2 = \{0\} \end{cases} \Leftrightarrow \begin{cases} \forall x \in E, \text{its decomposition} \\ x = x_1 + x_2, \text{ with } x_i \in E_i \\ \text{is unique} \end{cases}$$

$$\Leftrightarrow \dim(E) = \dim(E_1) + \dim(E_2). \qquad [2.13]$$

The direct sum $E_1 \oplus E_2$ can be interpreted as a decomposition of the v.s. $E$ into two subspaces $E_1$ and $E_2$ which span $E$ and are linearly independent.

If the subspaces $E_1$ and $E_2$ have the respective bases $\mathcal{B}_1$ and $\mathcal{B}_2$, we then have:

$\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$

$\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for E.

### 2.5.14.2. *Codimension of a subspace*

The codimension of a subspace $F$ of $E$, denoted by $\text{codim}(F)$, is the common dimension of the complements of $F$ in $E$.

If $E$ is of finite dimension, then $F$ has a complement $G$ such that $E = F \oplus G$, and thus $\dim(E) = \dim(F) + \dim(G)$, implying that: $\text{codim}(F) = \dim(E) - \dim(F)$. The codimension is the dimension of the quotient space $E/F$ (see section 2.5.15).

### 2.5.14.3. *Generalization*

It is said that $E$ is the direct sum of $N$ subspaces $E_1, \cdots, E_N$, if any $x \in E$ can be written in a unique way as $x = x_1 + \cdots, x_N$, with $x_n \in E_n, n \in \langle N \rangle$. We then write $E = E_1 \oplus E_2 \oplus \cdots \oplus E_N = \bigoplus_{n=1}^{N} E_n$, and we have:

$$\dim(\bigoplus_{n=1}^{N} E_n) = \sum_{n=1}^{N} \dim(E_n).$$

In the case of $N$ subspaces $E_n$, their direct sum corresponds to a decomposition of the v.s. $E$ into $N$ linearly independent subspaces which span $E$.

PROPOSITION 2.19.– *Given $N$ subspaces $E_n$ of the v.s. $E$, the direct sum $E = \overset{N}{\underset{n=1}{\oplus}} E_n$ is equivalent to:*

– $E = \overset{N}{\underset{n=1}{+}} E_n$

– $E_n \cap \{E_1 + \cdots + E_{n-1} + E_{n+1} + \cdots + E_N\} = \{0\}$ *for* $n \in \langle N \rangle$,

*meaning that each subspace is disjoint of the sum of the others.*

If $E_n$ has the basis $\mathcal{B}_n$, for $n \in \langle N \rangle$, then $\overset{N}{\underset{n=1}{\cup}} \mathcal{B}_n$ is a basis of $E$.

## 2.5.15. *Quotient vector spaces*

Let $(E, +, .)$ be a $\mathbb{K}$-v.s., and $G$ a subspace of $E$. Define the equivalence relation such that:

$$\forall\, u, v \in E : u \sim v \text{ if and only if } u - v \in G.$$

To any vector $u \in E$ can be associated a class of equivalence $c_u = \{v \in E : v \sim u\}$ denoted by $u + G = \{u + w, w \in G\}$. The set of equivalence classes $\{c_u : u \in E\}$, endowed with an addition and a scalar multiplication $(c_u + c_v = c_{u+v}, \lambda c_u = c_{\lambda u})$, has a structure of $\mathbb{K}$-v.s., hence the name quotient vector space.

This space, denoted by $(E/G, +, .)$, is such that:

$$E/G = \{c_u : u \in E\} = \{u + G, u \in E\}.$$

If $E$ is of finite dimension, we have: $\dim(E/G) = \dim(E) - \dim(G) = \text{codim}(G)$.

## 2.5.16. *Algebras*

### 2.5.16.1. *Definition*

An algebra over a commutative field $\mathbb{K}$, called a $\mathbb{K}$-algebra, is a set $E$ equipped with two internal laws $(+_E, \times_E)$ and an external law $(\circ)$, such that:

– $(E, +_E, \circ)$ is a $\mathbb{K}$-v.s.

– $(E, +_E, \times_E)$ is a ring:

- The law $\times_E$ is associative:

$$x \times_E (y \times_E z) = (x \times_E y) \times_E z \ , \ \forall x, y, z \in E.$$

- The law $\times_E$ is left and right distributive over $+_E$, that is, for all $x, y, z \in E$, we have:

$$x \times_E (y +_E z) = (x \times_E y) +_E (x \times_E z)$$

$$(y +_E z) \times_E x = (y \times_E x) +_E (z \times_E x)$$

– The law $\times_E$ is left and right distributive over $\circ$, that is, for all $x, y \in E$, and $\lambda \in \mathbb{K}$, we have:

$$x \times_E (\lambda \circ y) = (\lambda \circ x) \times_E y = \lambda \circ (x \times_E y).$$

A $\mathbb{K}$-algebra is therefore a $\mathbb{K}$-v.s. equipped with a second internal law noted multiplicatively.

An algebra $E$ is said to be commutative if the operation $\times_E$ is commutative:

$$x \times_E y = y \times_E x \ , \ \forall x, y \in E.$$

FACT 2.20.– Assuming that $(E, +_E, \times_E)$ is a unitary ring, that is, if there exists an identity element $1_E$ for the internal law $\times_E$ such that $1_E \times_E x = x \times_E 1_E = x$, $\forall x \in E$, the name unitary algebra is then employed.

### 2.5.16.2. *Polynomial algebra*

The set $\mathbb{K}_N[z]$ of polynomials with coefficients in the field $\mathbb{K}$, equipped with the internal addition and multiplication operations, and with the external multiplication by scalars of $\mathbb{K}$, has a commutative $\mathbb{K}$-algebra structure. Indeed, we have:

– $\mathbb{K}_N[z]$ endowed with addition and scalar multiplication is a $\mathbb{K}$-v.s.

– $\mathbb{K}_N[z]$ endowed with internal addition and multiplication operations is a commutative unitary ring.

– For any $p, q \in \mathbb{K}_N[z]$, and $\lambda \in \mathbb{K}$, we have: $p(\lambda q) = (\lambda p)q = \lambda(pq)$, namely, the internal multiplication is distributive over the (external) scalar multiplication.

– For any $p, q \in \mathbb{K}_N[z]$, we have $pq = qp$, which implies the commutativity of the $\mathbb{K}$-algebra of polynomials.

### 2.5.16.3. *Algebra of endomorphisms*

Given a $\mathbb{K}$-v.s. $E$, let $(\mathcal{L}(E), +, \circ, .)$ be the set of endomorphisms $\mathcal{L}(E)$ from $E$ to itself, equipped with internal operations (+) for the addition and $(\circ)$ for the composition of maps[21], and an external operation $(.)$ for the multiplication by a scalar of $\mathbb{K}$. The set $(\mathcal{L}(E), +, \circ, .)$ is a unitary $\mathbb{K}$-algebra.

21 Note that, contrary to the general definition of v.s. and algebras where the symbol $(\circ)$ represents the external operation, here it refers to an internal operation, as is the case in general to designate the composition of maps.

PROOF.– In section 2.5.10.2, we have seen that $(\mathcal{L}(E), +, .)$ is a $\mathbb{K}$-v.s. In addition, the map composition law $(\circ)$ is associative, and it is distributive over the operations of addition $(+)$ and multiplication $(.)$, that is, for all $\lambda \in \mathbb{K}$, and all $f, g \in \mathcal{L}(E)$, we have:

$$f \circ (g + h) = (f \circ g) + (f \circ h) \ ; \ (g + h) \circ f = (g \circ f) + (h \circ f)$$
$$f \circ (\lambda.g) = (\lambda.f) \circ g = \lambda.(f \circ g).$$

This shows that $(\mathcal{L}(E), +, \circ, .)$ is a $\mathbb{K}$-algebra. Note that it is not a commutative algebra since the composition law is not, in general, commutative. □

#### 2.5.16.4. *Algebra of matrices*

In Chapter 4, we shall see that a set of matrices, endowed with addition and scalar multiplication operations, is a v.s. In the case of square matrices, by equipping this v.s. with the usual matrix product, an algebra is obtained, but not commutative.

In section 5.12, we shall show that the set of block matrices equipped with block Kronecker and Hadamard products form an algebra and a commutative algebra, respectively.

The properties of the different algebraic structures presented in this chapter are summarized in Table 2.7.

## 2.6. Morphisms

Let $E$ and $F$ be two sets sharing the same algebraic structure. A map $f : E \to F$ is called a morphism (or homomorphism) if it preserves the algebraic structure. If $E$ and $F$ are equipped with $N$ laws $*_n$ and $\bullet_n$, with $n \in \langle N \rangle$, respectively, the map $f$ is a morphism if the following conditions hold:

– $f$ is such that

$$\forall n \in \langle N \rangle, \forall (x, y) \in E^2 \ , \ f(x *_n y) = f(x) \bullet_n f(y). \qquad [2.14]$$

– If according to the definition axioms of the structure, there exists a neutral element $(e_E)_n$ for the law $*_n$, then $f[(e_E)_n] = (e_F)_n$, where $(e_F)_n$ is the neutral element for $\bullet_n$.

### 2.6.1. *Group morphisms*

Given two groups $(E, *)$ and $(F, \bullet)$, a map $f : E \to$ F is a morphism of groups if:

$$\forall (x_1, x_2) \in E^2 \ , \ f(x_1 * x_2) = f(x_1) \bullet f(x_2).$$

| Properties of operations | Algebraic structures | | | | | |
|---|---|---|---|---|---|---|
| | Abelian group | Semi ring | Ring | Field | Vector space | Algebra |
| **Internal operation + (or $+_E$)** | | | | | | |
| Associative | yes | yes | yes | yes | yes | yes |
| Commutative | yes | yes | yes | yes | yes | yes |
| Identity element 0 | yes | | yes | yes | yes | yes |
| Opposite | yes | | yes | yes | yes | yes |
| **Internal operation . (or $\times_E$)** | | | | | | |
| Associative | | yes | yes | yes | | yes |
| Distributive over + (or $+_E$) | | yes | yes | yes | | yes |
| Distributive over $\circ$ | | | | | | yes |
| Identity element $1_{\mathbb{K}}$ (or $1_E$) | | | | yes | | yes |
| Inverse except for 0 | | | | yes | | |
| With no zero divisor | | | | yes | | |
| **External operation $\circ$** | | | | | | |
| Associative | | | | | yes | yes |
| Distributive over $+_E$ and $+$ | | | | | yes | yes |
| Identity element $1_{\mathbb{K}}$ | | | | | yes | yes |

**Table 2.7.** *Algebraic structures*

Group morphisms satisfy the following properties:

– Let $e_E$ and $e_F$ be, respectively, the neutral elements of $E$ and $F$. We have: $f(e_E) = e_F$.

– If $H$ is a subgroup of $E$, then $f(H)$ is a subgroup of $F$.

– The image of $f$, $\text{Im}(f) = f(E)$, is a subgroup of $F$.

– The kernel $\text{Ker}(f) = \{x \in E : f(x) = e_F\}$ is a subgroup of $E$.

– The composition $g \circ f$ of two morphisms of groups is a morphism of groups.

– If $f$ is a bijective morphism, it is said that $f$ is a group isomorphism, and $f^{-1}$ is also a group isomorphism. The set $f^{-1}(\{e_F\})$ is called the kernel of $f$.

– If $(E, *) = (F, \bullet)$, it is said that $f$ is an endomorphism[22] of group $E$.

– If $f$ is a bijective endomorphism, it is called an automorphism of group $E$.

### 2.6.2. *Ring morphisms*

Given two rings $(E, +_E, .E)$ and $(F, +_F, .F)$, a map $f : E \to$ F is a ring morphism if:

$$\forall (x, y) \in E^2 \ , \ f(x +_E y) = f(x) +_F f(y) \qquad [2.15a]$$

$$\forall (x, y) \in E^2 \ , \ f(x ._E y) = f(x) ._F f(y). \qquad [2.15b]$$

Property [2.15a] means that $f$ is a morphism of additive groups, while for Property [2.15b] it is said that $f$ is compatible with multiplication.

If $f$ is surjective (injective, bijective), it is said that $f$ is an epimorphism (monomorphism, isomorphism) of rings. In the case of an isomorphism of rings, the set $f^{-1}(\{0\})$ is called the kernel of $f$.

If $E$ has a neutral element, denoted by $1_E$, for multiplication $(.E)$, that is, $E$ is a unitary ring, then $F$ has a neutral element $1_F = f(1_E)$ for the operation $(.F)$. As a result, $F$ is itself a unitary ring, and $f$ is a morphism of monoids for multiplication. It is then said that $f$ is a unitary morphism.

### 2.6.3. *Morphisms of vector spaces or linear maps*

Hereafter, we provide the definitions of a few special cases of morphism of v.s., also called linear morphisms, or linear maps, namely, epimorphisms, monomorphisms, endomorphisms, isomorphisms, and automorphisms.

– Given two $\mathbb{K}$-v.s., $E$ and $F$, of finite dimension, a linear map $f \in \mathcal{L}(E, F)$ from $E$ to $F$ such as defined in [2.6] is also called a $\mathbb{K}$-linear morphism between $E$ and $F$.

– A surjective (injective) linear map is called an epimorphism (monomorphism) of vector spaces.

– A bijective linear map $f \in \mathcal{L}(E, F)$ is an isomorphism of vector spaces. It is then said that the sets $E$ and $F$ are isomorphic.

22 An endomorphism is a morphism of a structured set into itself. Thereby, this leads to the notion of endomorphism of the group or ring $E$. An endomorphism of a v.s. $E$ is a linear map $f : E \to E$.

– If $F = E$, it is said that $f$ is an endomorphism of $E$, that is to say, a linear map from $E$ to itself, and we write $f \in \mathcal{L}(E)$.

– It is said that $f$ is a nilpotent endomorphism if there exists a natural number $n \in \mathbb{N}^*$ such that $f^n = 0$. The smallest $n \in \mathbb{N}^*$ such that $f^n = 0$ is called the nilpotence index (or degree) of $f$.

The matrix associated to $f^n$ being equal to $\mathbf{A}^n$ (see Corollary 4.54), where $\mathbf{A}$ is the matrix associated to $f$, it can be concluded that the matrix associated with a nilpotent endomorphism, of nilpotence index $n$, is such that $\mathbf{A}^n = \mathbf{0}$, which corresponds to the definition of a nilpotent matrix of degree $n$ (see Table 4.3).

In section 3.5.4, we shall define the notion of orthogonal/unitary endomorphism.

– A bijective endomorphism is called an automorphism. The set of automorphisms of $E$, equipped with the composition of endomorphisms, is a group called the linear group of $E$ and denoted by $\mathrm{GL}(E)$.

#### 2.6.3.1. *Properties*

We summarize hereunder a few properties of linear maps. Let $f \in \mathcal{L}(E, F)$.

– The kernel $\mathrm{Ker}(f) = f^{-1}(\{0\}) = \{x \in E : f(x) = 0\}$ is a subspace of $E$, whereas the image $\mathrm{Im}(f) = f(E) = \{f(x) : x \in E\}$ is a subspace of $F$. This set is also called the range of $f$.

– As seen in section 2.4.2, $f$ is surjective if and only if $\mathrm{Im}(f) = F$.

– $f$ is injective if and only if $\mathrm{Ker}(f) = \{0\}$. Indeed, exploiting the linearity of $f$, we have for $u, v \in E$:

$$f(u) = f(v) \Leftrightarrow f(u - v) = 0 \Leftrightarrow u - v \in \mathrm{Ker}(f).$$

By definition, $f$ is injective if and only if $f(u) = f(v) \Rightarrow u = v$, therefore if and only if $(u - v) \in \mathrm{Ker}(f) \Rightarrow u - v = 0$, which implies $\mathrm{Ker}(f) = \{0\}$.

– Two $\mathbb{K}$-v.s., $E$ and $F$, of finite dimension are isomorphic if and only if they have the same dimension.

So, for example, the v.s. $\mathbb{K}^{I \times J}$ of matrices of dimensions $I \times J$, is isomorphic to the v.s. $\mathbb{K}^{IJ}$ of vectors of dimension $IJ$ obtained by vectorization of matrices. More generally, any $N$-dimensional $\mathbb{K}$-v.s. is isomorphic to $\mathbb{K}^N$.

In Chapter 6, we shall illustrate the notion of isomorphism of tensor spaces through the operations of matricization and vectorization.

– If $f$ is injective and if $\{u_1, \cdots, u_I\}$ is a free system of $E$, then $\{f(u_1), \cdots, f(u_I)\}$ is a free system of $F$. It is said that $f$ preserves free systems.

This result follows from the rank theorem below, with $\dim(\text{Ker}(f)) = 0$. Injective maps are also known as one-to-one maps.

– If $f$ is surjective and if $\{u_1, \cdots, u_I\}$ is a system of generators of $E$, then $\{f(u_1), \cdots, f(u_I)\}$ is a system of generators of $F$. It is said that $f$ preserves systems of generators. Surjective maps are also known as onto maps.

– If $f$ is an isomorphism, then $f^{-1}$ is also an isomorphism, and the image of any basis of $E$ by $f$ is a basis of $F$. It is said that $f$ preserves bases.

### 2.6.3.2. *Canonical factorization of linear maps*

Let $f \in \mathcal{L}(E, F)$ be a linear map from $E$ to $F$, supposed to be not injective, that is, several antecedents in $E$ can be associated to an image $f(x)$. Let us define the equivalence relation $\sim$ on $E$ such that $\forall u, v \in E$:

$$u \sim v \;\Leftrightarrow\; f(u) = f(v),$$

the equivalence class $c_u$ corresponding to the set of the elements $v$ of $E$ that have the same image as $u$. Using the linearity of $f$, we obtain:

$$f(u) = f(v) \;\Leftrightarrow\; f(v-u) = 0 \;\Leftrightarrow\; v - u \in \text{Ker}(f) \;\Leftrightarrow\; v = u + \text{Ker}(f).$$

Defining the quotient space $E/\text{Ker}(f) = \{c_u : u \in E\}$, i.e. the set of equivalence classes $c_u$ of $E$, the mapping

$$f_c : \; E/\text{Ker}(f) \to F, \;\; E/\text{Ker}(f) \ni c_u \mapsto f_c(c_u) = f(u) \in F \qquad [2.16]$$

is injective, which means that a single antecedent can be associated with an image $f(u)$ which is the equivalence class $c_u$.

The map $f$ can then be decomposed as:

$$f = g \circ f_c \circ h \qquad [2.17a]$$

$$h : \; E \to E/\text{Ker}(f), \;\; E \ni u \mapsto h(u) = c_u \in E/\text{Ker}(f) \qquad [2.17b]$$

$$g : \; \text{Im}(f) \to F, \;\; \text{Im}(f) \ni f(u) \mapsto g[f(u)] = f(u) \in F, \qquad [2.17c]$$

where $h$ is surjective, and $f_c$ is defined in [2.16].

Equivalently, we have:

$$E \overset{h}{\to} E/\text{Ker}(f) \overset{f_c}{\to} \text{Im}(f) \overset{g}{\to} F$$
$$u \mapsto c_u \mapsto f(u) \mapsto f(u).$$

This map composition [2.17a] is called canonical factorization[23] of $f$ for the law $\circ$, through quotient space. In the case where $f$ is surjective ($\text{Im}(f) = F$), we have $f = f_c \circ h$. This equation, called universal property, can be represented by the following commutative diagram:

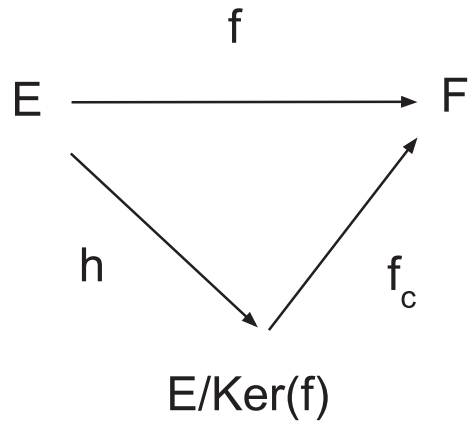



**Figure 2.1.** *Commutative diagram for universal property*

### 2.6.3.3. *Rank of a linear map and rank theorem*

The rank of a linear map $f \in \mathcal{L}(E, F)$, denoted by $\text{r}(f)$, refers to the dimension of its image $\text{Im}(f)$:

$$\text{r}(f) = \dim[\text{Im}(f)]$$

with $\text{r}(f) \leq \dim(F)$, the equality being possible if and only if $f$ is surjective. If $\{b_1, \cdots, b_I\}$ is a basis of $E$, the rank of $f$ is the rank of the set of vectors $\{f(b_1), \cdots, f(b_I)\}$, namely, the dimension of the subspace generated by $\{f(b_1), \cdots, f(b_I)\}$.

PROPOSITION 2.21.– *(Rank theorem): this theorem corresponds to the following equality:*

$$r(f) + dim(Ker(f)) = dim(E). \qquad [2.18]$$

When $f$ is an endomorphism of a finite-dimensional v.s. $E$, the rank theorem makes it possible to deduce from [2.13] that $\text{Im}(f)$ and $\text{Ker}(f)$ are complements if and only if their intersection is the null vector.

23 M. Sage, "Introduction aux espaces quotients", 2005, www.normalesup.org/sage/Enseignement/Cours/Quotient.pdf.

If $f$ is bijective, we have $\text{Ker}(f) = \{0\}$, and it can be concluded that $\text{r}(f) = \dim(E)$.

The rank of $f$ is equal to the rank of the matrix associated to $f$ (see section 4.10.1).

If $E$ and $F$ are of the same finite dimension, we have the following equivalences:

$$f \text{ bijective } \Leftrightarrow f \text{ injective } \Leftrightarrow f \text{ surjective,}$$

with:

– $\text{Ker}(f) = \{0\}$.

– $\text{Im}(f) = F$.

– $\text{r}(f) = \dim(E)$.

– Each basis of $E$ is mapped to a linearly independent set of vectors in $F$.

### 2.6.3.4. *Dual basis*

In section 2.5.10.3, we saw that the dual space $E^* = \mathcal{L}(E, \mathbb{K})$ of the $\mathbb{K}$-v.s. $E$, of dimension $I$, is such that $\dim(E^*) = \dim(E) = I$. Given a basis $\mathcal{B} = \{b_1, \cdots, b_I\}$ of $E$, it is possible to build a basis of $E^*$ as a set of $I$ linear forms $\{b_1^*, \cdots, b_I^*\}$ defined by the images of the basis vectors $\mathcal{B}$, that is: $b_j^*(b_i) = \delta_{ij}$, where $\delta_{ij}$ is the Kronecker delta equal to 1 if $i = j$ and 0 otherwise. This basis of $E^*$, denoted by $\mathcal{B}^*$, is called the dual basis of $\mathcal{B}$, and we have:

$$E \ni x = \sum_{i=1}^{I} x_i b_i \;\Rightarrow\; b_j^*(x) = \sum_{i=1}^{I} x_i b_j^*(b_i) = \sum_{i=1}^{I} x_i \delta_{ij} = x_j.$$

$b_j^*$ is thus the linear form which transforms $x \in E$ into its $j$th coordinate in the basis $\mathcal{B}$. A linear form $f \in E^*$ is written in the dual basis as:

$$f = \sum_{i=1}^{I} f_i b_i^* \text{ with } f_i = f(b_i) \in \mathbb{K}. \qquad [2.19]$$

Indeed, for any $x \in E$, we have:

$$f(x) = f(\sum_{i=1}^{I} x_i b_i) = \sum_{i=1}^{I} x_i f(b_i) = \sum_{i=1}^{I} f(b_i) b_i^*(x) = \sum_{i=1}^{I} f_i b_i^*(x),$$

from which we can deduce [2.19].

### 2.6.4. *Algebra morphisms*

Given two algebras $E(\mathbb{K})$ and $F(\mathbb{K})$ over the field $\mathbb{K}$, endowed with two internal operations $(+_E, \times_E)$ and $(+_F, \times_F)$, and an external operation $\circ_E$ and $\circ_F$, respectively, a map $f : E \to F$ is a morphism of algebras if:

$$
\begin{aligned}
f(x +_E y) &= f(x) +_F f(y) \ , \ \forall x, y \in E \\
f(\lambda \circ_E x) &= \lambda \circ_F f(x) \ , \ \forall x \in E, \forall \lambda \in \mathbb{K}. \\
f(x \times_E y) &= f(x) \times_F f(y) \ , \ \forall x, y \in E. \\
f(1_E) &= 1_F.
\end{aligned}
$$

From a practical point of view, it could be worthwhile to use an isomorphism when the operations are easier to carry out in $F$ than in $E$. We then have:

$$
x +_E y = f^{-1}(f(x) +_F f(y)) \ , \ x \times_E y = f^{-1}(f(x) \times_F f(y)).
$$

# 3

# Banach and Hilbert Spaces – Fourier Series and Orthogonal Polynomials

## 3.1. Introduction and chapter summary

This chapter is dedicated to normed vector spaces (n.v.s.) and pre-Hilbert spaces, also called inner product spaces, or more specifically v.s. equipped with a Euclidean (for real v.s.) or Hermitian (for complex v.s.) inner product. In the case of finite-dimensional v.s., we obtain Euclidian and Hermitian spaces, respectively. In the case of infinite-dimensional v.s., Banach and Hilbert spaces will be introduced.

From a signal processing perspective, this chapter can be considered as an introduction to functional analysis of both digital (time sequences) and analog (time functions) deterministic signals (Allen and Mills 2004; Reinhard 1997). This type of analysis is performed in a Hilbert space, which requires the definition of:

– the space of signals to be analyzed, in terms of time support (finite or infinite, discrete or continuous) and finite energy;

– an inner product and an associated norm making it possible to define a metric;

– a so-called Hilbertian basis $\{e_n(t), n \in I\}$ to extend signals in the form of a sum over the basis considered, $x(t) = \sum_{n \in I} c_n e_n(t)$. The approximation error, associated with a limited sum, also known as partial sum, is often evaluated as a mean squared error (MSE).

In the case of a periodic signal, the use of a trigonometric basis leads to Fourier series, and makes it possible to decompose the signal into a sum of exponential or sinusoidal signals. This is then referred to as frequency decomposition or spectral representation of the signal, in other words, a frequency-based representation.

*From Algebraic Structures to Tensors,*
First Edition. Edited by Gérard Favier.

Representation and approximation of functions, and thereby of signals and systems, play a very important role in signal and image processing, as Fourier series constitute the most famous example. It is also possible to use Hilbert bases of orthogonal polynomials, such as Legendre[1], Hermite[2], Laguerre[3], or Chebyshev[4] polynomials, for the modeling of linear systems and nonlinear systems. In addition to their application for function approximation, orthogonal polynomials are also used in combinatorics, coding and probability theory, as well as for solving interpolation problems. Other examples of Hilbert bases that are not considered in this chapter are wavelet bases used for multiresolution (or multi-scale) analysis. In image processing, this type of analysis is intended to decompose an image to extract features that enable segmentation, classification, shape recognition, or compression to be performed.

The objective of this chapter is to introduce the basic concepts underlying Banach and Hilbert spaces and to give an illustration thereof through the problem of function approximation in the form of expansions over Hilbert bases. Two types of bases will be considered: trigonometric bases leading to Fourier series and bases of orthogonal polynomials.

We will first define the notions of distance and metric space. Then, we will illustrate the use of distance for the study of convergence of sequences and of local continuity of a function. Then, the definitions of norm and inner product allow

1 Adrien-Marie Legendre (1752–1833), a French mathematician who was the author of several important contributions to the theory of numbers, abstract algebra, and statistics. He published several books including: *Eléments de géométrie* (1794), which remained a reference work for more than a century, following Euclid's *Elements, Essai sur la théorie des nombres* (1808), and *Traité des fonctions elliptiques* (in three volumes; 1826–1829). He published his polynomials, in 1794, in *Recherches sur la figure des planètes*. He is more particularly famous for the polynomials and the transformation (used in Lagrangian mechanics) that bear his name, as well as for his work on elliptic functions, not to mention the method of least squares, developed independently by Legendre and Gauss.

2 Charles Hermite (1822–1901), a French mathematician who made important contributions to the theory of orthogonal polynomials, quadratic forms, and elliptic functions. He is more particularly known for the differential equations, the polynomials, and the interpolation formula that bear his name, as well as for the structure of Hermitian space, and for Hermitian matrices.

3 Edmond Laguerre (1834–1886), a French mathematician, member of the Académie des Sciences, who is mainly known for the differential equations, the transformation and the polynomials named after him. His complete works were published in two volumes, by Hermite, Poincaré, and Rouché, in 1898 and 1905. His work in analysis and geometry were considered interesting enough to be reprinted in 1972.

4 Pafnuty Lvovich Chebyshev (1821–1894), a Russian mathematician who published his polynomials in 1854. Although Legendre's polynomials were known since the end of the 18th century, and those of Hermite since the beginning of the 19th century, Chebyshev is certainly the first mathematician to have laid the foundations for the general theory of orthogonal polynomials with applications in probability and for function interpolation and approximation.

us to introduce normed vector spaces and pre-Hilbert spaces. Various examples of norms and inner products will be provided, as well as the Hölder, Cauchy-Schwarz, and Minkowski inequalities. The notions of orthogonality, orthonormal basis, and orthogonal projection onto a subspace will be presented before describing the Gram–Schmidt orthonormalization process which is a fundamental method of linear algebra.

After defining the notion of complete metric space, we will introduce Banach[5] and Hilbert[6] spaces. Several examples of Hilbert spaces as well as Hilbert bases for these spaces will be described. We will then present the notion of Fourier series expansion of periodic functions. The convergence of these series will be illustrated through the Parseval and Dirichlet–Jordan theorems. Several examples of Fourier series will be described along with their use for establishing formulae of sums of series. Such formulae will also be highlighted by using Parseval's equality. Finally, the expansion of functions over bases of orthogonal polynomials will be illustrated using Legendre, Hermite, Laguerre, and Chebyshev polynomials.

This chapter is not intended to make an exhaustive presentation of Hilbert spaces and of the theory of function approximation. In particular, the demonstrations of convergence properties of Fourier series will not be provided here. Our goal is rather to make the link between Hilbert spaces and function approximation through the use of expansions over Hilbert bases.

## 3.2. Metric spaces

A metric space is a space equipped with a distance.

5 Stefan Banach (1892–1945), a Polish mathematician who defined, in an axiomatic way, the concept of complete normed vector space, called Banach space, an essential tool for functional analysis, or more precisely, for the study of vector spaces of infinite dimension. He made major contributions to the theory of topological vector spaces. His work focused on measure theory, set theory, and orthogonal series. He generalized those of Volterra, Fredholm, and Hilbert on integral equations.

6 David Hilbert (1862–1943), a German mathematician who contributed to the development of different branches of mathematics, including the theory of invariants, the axiomatization of Euclidean geometry, functional analysis through the study of differential and integral equations, the calculus of variations, and the algebraic theory of numbers. He is particularly known for the spaces and the finite basis theorem in ring theory, which bear his name. He also developed the use of mathematics in physics for the kinetic theory of gases, the theory of radiation, quantum mechanics, and general relativity. In 1900, he introduced his famous list of 23 unsolved problems, called the Hilbert problems, at the International Congress of Mathematicians, Paris.

### 3.2.1. *Definition of distance*

Let $E$ be a set. A map $d : E \times E \to \mathbb{R}^+$ is a distance, also called a metric, if it satisfies the following properties for all $x, y, z \in E$

$$\text{Non-negativity}: \quad d(x,y) \geq 0 \qquad [3.1]$$

$$\text{Strict positivity}: \quad d(x,y) = 0 \ \Leftrightarrow x = y \qquad [3.2]$$

$$\text{Symmetry}: \quad d(x,y) = d(y,x) \qquad [3.3]$$

$$\text{Triangle inequality}: \quad d(x,y) \leq d(x,z) + d(z,y) \qquad [3.4]$$

A geometric interpretation of the triangle inequality is that the length of one side of a triangle is smaller than the sum of the lengths of the other two sides.

The set $E$ endowed with the distance $d$ is denoted by $(E, d)$. When the property of definite positivity is not imposed, that is, if $d(x, y)$ may be zero for a couple of distinct points $x$ and $y$, it is said that $d$ defines a semi-distance.

### 3.2.2. *Definition of topology*

The notion of distance, which does not require defining a structure, makes it possible to generate a topology based on the concepts of open ball and open set:

– For all $x \in E$, the open (or closed) ball of center $x$ and radius $r$ is defined as $\mathcal{B}_r(x) = \{y \in E : d(x,y) < r\}$ (or $\overline{\mathcal{B}}_r(x) = \{y \in E : d(x,y) \leq r\}$), with $r \in \mathbb{R}^+$.

– A subset $F$ of a metric space $(E, d)$ is an open set if and only if for all $x \in F$, there is an open ball $\mathcal{B}_r(x) \subset F$.

– A subset $F$ of a metric space $(E, d)$ is closed if it is the complement of an open set.

The topology of $(E, d)$ is the set of open sets $F \subset E$. Open sets are used to define the concepts of convergence and continuity. See sections 3.2.5 and 3.2.6.

FACT 3.1.– In a metric space, open and closed sets satisfy the following properties:

– If $F_1, \cdots, F_N$ are open sets, their intersection $\bigcap_{n=1}^{N} F_n$ is an open set.

– An arbitrary union of open sets is open.

– If $F_1, \cdots, F_N$ are closed sets, their union $\bigcup_{n=1}^{N} F_n$ is a closed set.

– An arbitrary intersection of closed sets is closed.

These last two properties can be established from the first two ones using De Morgan's laws (see section 2.3.5).

When a v.s. is endowed with a topology, it is called a topological v.s. The Banach and Hilbert spaces considered in section 3.7 are two well-known examples of topological v.s.

### 3.2.3. *Examples of distances*

We give below a few examples of distances.

– For $E = \mathbb{R}$ (or $\mathbb{C}$), the map $\mathbb{R}^2 \ni (x, y) \mapsto d(x, y) = |x - y|$ defines a distance between $x$ and $y$, where $|.|$ designates the absolute value (or the modulus).

– For $E = \mathbb{R}^N$ (or $\mathbb{C}^N$), the following map is called Hölder's[7] distance:

$$d_p(x, y) = \Big( \sum_{n=1}^{N} |x_n - y_n|^p \Big)^{1/p}, \ 1 \leq p \leq \infty. \qquad [3.5]$$

In particular, for $p = 1$ and $p = 2$, we have the distance associated with the absolute value (or the modulus) and the Euclidean (or Hermitian) distance in $\mathbb{R}^N$ (or $\mathbb{C}^N$), respectively:

$$d_1(x, y) = \sum_{n=1}^{N} |x_n - y_n|$$

$$d_2(x, y) = \sqrt{\sum_{n=1}^{N} |x_n - y_n|^2}$$

$$d_\infty(x, y) = \max_n \big\{ |x_n - y_n| \big\}.$$

– In data statistical analysis, the Mahalanobis distance (Mahalanobis 1936) is very commonly used. It is a means to measure the dissimilarity between two random vectors $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^N$, having the same distribution with the covariance matrix $\mathbf{\Sigma}$, such as:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{y})}.$$

If the covariance matrix is diagonal, with diagonal elements equal to $\sigma_n^2$ , for $n \in \langle N \rangle$, we have the normalized Euclidean distance:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{n=1}^{N} \frac{(x_n - y_n)^2}{\sigma_n^2}}.$$

7 Otto Ludwig Hölder (1859–1937), a German mathematician who contributed to the development of group theory. He is more particularly known for the norm and inequalities related to spaces $l^p$ of sequences and $L^p$ of functions, which bear his name.

Unlike the Euclidean distance, which assigns the same unit weight to each component, the Mahalanobis distance weights each component inversely proportional to its dispersion, that is, its variance $\sigma_n^2$. In the case of the analysis of Gaussian signals, this amounts to reducing the influence of most noisy components.

### 3.2.4. *Inequalities and equivalent distances*

A distance $d$ satisfies the generalized triangle inequality for all $x_n \in E, n \in \langle N \rangle$, and the second triangle inequality for all $x, y, z \in E$:

$$d(x_1, x_N) \leq \sum_{n=1}^{N-1} d(x_n, x_{n+1}) \qquad [3.6a]$$

$$|d(x, z) - d(z, y)| \leq d(x, y). \qquad [3.6b]$$

Two distances $d$ and $\delta$ defined over the same metric space $E$ are said to be equivalent if there exists two positive numbers $a$ and $b$ such that, for all $x, y \in E$, we have:

$$a\, d(x, y) \leq \delta(x, y) \leq b\, d(x, y).$$

### 3.2.5. *Distance and convergence of sequences*

In a metric space $(E, d)$, a sequence $(x_n)_{n \in \mathbb{N}}$ converges to $x$ if and only if $d(x_n, x) \to 0$ when $n \to \infty$, or equivalently:

$$\forall \epsilon > 0, \exists N \in \mathbb{N}\,, \forall n \geq N\,, d(x_n, x) < \epsilon. \qquad [3.7]$$

This amounts to saying that the sequence enters and remains in the open ball $\mathcal{B}_\epsilon(x)$.

### 3.2.6. *Distance and local continuity of a function*

Given two metric spaces $(E, d)$ and $(F, \delta)$, and a map $f : E \to F$, it is said that $f$ is continuous at $x$ if and only if for any sequence $(x_n)$ of $E$ that converges to $x$, then the sequence $(f(x_n))$ of $F$ converges to $f(x)$, that is, for any $\epsilon > 0$, there exists $\eta > 0$ such that:

$$\forall y \in E,\ d(x, y) < \eta \Rightarrow \delta\big(f(x), f(y)\big) < \epsilon. \qquad [3.8]$$

### 3.2.7. *Isometries and Lipschitzian maps*

Let $f : (E, d) \to (F, \delta)$ be a map between two metric spaces.

– $f$ is an isometry if for all $x, y \in E$, we have:

$$\delta[f(x), f(y)] = d(x, y). \qquad [3.9]$$

It is then said that $f$ preserves distances.

– $f$ is Lipschitzian of constant $C$ (known as $C$-Lipschitzian) when:

$$\delta[f(x), f(y)] \leq Cd(x, y) \ , \ \forall x, y \in E. \qquad [3.10]$$

This is a regularity property stronger than continuity. $f$ is called a contracting map or a contraction if $C \in [0, 1[$.

As we have just seen, the concept of distance makes it possible to define the convergence of a sequence and the continuity of a function. The convergence of Cauchy sequences is at the base of the definition of complete spaces, and thus of Banach and Hilbert spaces that will be covered in section 3.7.

In the two sections that follow, we complete the presentation of v.s. by equipping them with a norm and an inner product, which leads to normed v.s. and pre-Hilbert spaces, respectively.

## 3.3. Normed vector spaces

A $\mathbb{K}$-v.s. $E$ is said to be normed if it is equipped with a norm.

### 3.3.1. *Definition of norm and triangle inequalities*

A norm on a v.s. $E$ is a map $E \to \mathbb{R}^+$, denoted by $\|.\|$, which to any vector $x \in E$ associates the number $\|x\|$ satisfying the following properties:

$$\text{Non-negativity :} \quad \forall x \in E : \|x\| \geq 0 \qquad [3.11]$$

$$\text{Strict positivity :} \quad \|x\| = 0 \ \Leftrightarrow x = 0 \qquad [3.12]$$

$$\text{Homogeneity :} \quad \forall (\lambda, x) \in \mathbb{K} \times E : \ \|\lambda x\| = |\lambda| \|x\| \qquad [3.13]$$

$$\text{Triangle inequality :} \quad \forall (x, y) \in E \times E : \ \|x + y\| \leq \|x\| + \|y\| \qquad [3.14]$$

where $|.|$ denotes the absolute value if $\mathbb{K} = \mathbb{R}$, or the modulus if $\mathbb{K} = \mathbb{C}$. A normed v.s. (n.v.s.) $E$ equipped with the norm $\|.\|$ is denoted by $(E, \|.\|)$.

PROPOSITION 3.2.– *The norm also satisfies a generalization of the triangle inequality as well as a second triangle inequality:*

$$\|\sum_{n=1}^{N} x_n\| \leq \sum_{n=1}^{N} \|x_n\|, \ \forall (x_1, \cdots, x_N) \in E^N. \quad [3.15a]$$

$$|\|x\| - \|y\|| \leq \|x \pm y\|, \ \forall (x, y) \in E^2. \quad [3.15b]$$

PROOF.– Use of the triangle inequality [3.14] yields:

$$\|x\| = \|y + x - y\| \leq \|y\| + \|x - y\| \ \Rightarrow \ \|x\| - \|y\| \leq \|x - y\|. \quad [3.16]$$

By permuting $x$ and $y$ in the aforementioned inequality, we get $\|y\| - \|x\| \leq \|x - y\|$, and in combination with [3.16], the following inequality can be deduced:

$$|\|x\| - \|y\|| \leq \|x - y\|, \ \forall (x, y) \in E^2. \quad [3.17]$$

By changing $y$ into $-y$, this inequality becomes:

$$|\|x\| - \|y\|| \leq \|x + y\| \ , \ \forall (x, y) \in E^2. \quad [3.18]$$

Inequalities [3.17] and [3.18] can be grouped together as shown in [3.15b]. □

NOTE 3.3.– In geometry, this inequality [3.18] means that the length of one side of a triangle is greater than or equal to the absolute value of the difference of the lengths of the other two sides.

## 3.3.2. *Examples of norms*

### 3.3.2.1. *Vector norms*

For $\mathbf{x} \in \mathbb{R}^N$, the following norms can be defined:

– Hölder's norm or $l_p$ norm

$$\|\mathbf{x}\|_p = (\sum_{n=1}^{N} |x_n|^p)^{1/p}, \ 1 \leq p \leq \infty. \quad [3.19]$$

Two important special cases of the $l_p$ norm correspond to the $l_1$ norm and the Euclidean $l_2$ norm, for $p = 1$ and $p = 2$, respectively, defined as:

$$\|\mathbf{x}\|_1 = \sum_{n=1}^{N} |x_n| \quad , \quad \|\mathbf{x}\|_2 = (\sum_{n=1}^{N} x_n^2)^{1/2}.$$

$\mathbb{R}^N$ equipped with the Euclidean norm $l_2$ is called the $N$-dimensional Euclidean space. The $l_1$ norm corresponds to the sum of the absolute values of the components of the vector $\mathbf{x}$, while the $l_p$ norm for $p \neq 1$ is obtained by taking the $p$th root of the sum of the absolute values to the power $p$ of the components of $\mathbf{x}$.

The cases $p = 1$ and $p = 2$ provide the norms of the convergence in mean and in quadratic mean, respectively. See section 3.8.5.

– Maximum norm (Max norm, or $\infty$-norm)

$$\|\mathbf{x}\|_\infty = \max_n\{|x_n|\}.$$

PROOF.– This expression results from $\lim_{p\to\infty} \|\mathbf{x}\|_p = \max_n\{|x_n|\}$. Indeed, designating by $M$ the maximum of $|x_n|$, and assuming that $m$ components of $\mathbf{x}$ have the $M$ value, $\|\mathbf{x}\|_p$ can be rewritten as:

$$\|\mathbf{x}\|_p = M\Big(m + \sum_{n:|x_n|<M} (\frac{|x_n|}{M})^p\Big)^{1/p}.$$

Consequently, by taking the logarithm and noting that $\frac{|x_n|}{M} < 1$, we obtain:

$$\lim_{p\to\infty} \log\|\mathbf{x}\|_p = \log M + \lim_{p\to\infty} \frac{1}{p}\log\Big[m + \sum_{n/|x_n|<M} (\frac{|x_n|}{M})^p\Big] = \log M,$$

from which it can be deduced that $\|\mathbf{x}\|_\infty = \lim_{p\to\infty} \|\mathbf{x}\|_p = M = \max_n\{|x_n|\}$. □

PROPOSITION 3.4.– *(Hölder's inequality):*

*Suppose $(p, q)$ is a pair of conjugated exponents, such that $\frac{1}{p} + \frac{1}{q} = 1$. For any vectors $\mathbf{x}$ and $\mathbf{y}$ of $\mathbb{R}^N$, Hölder's inequality*[8] *is given by:*

$$|\sum_{n=1}^N x_n y_n| \leq \big(\sum_{n=1}^N |x_n|^p\big)^{1/p} \big(\sum_{n=1}^N |y_n|^q\big)^{1/q} = \|\mathbf{x}\|_p \|\mathbf{y}\|_q. \qquad [3.20]$$

For $p = q = 2$, the Hölder inequality [3.20] corresponds to the Cauchy–Schwarz inequality [3.36] for the space $\mathbb{R}^N$:

$$|\mathbf{x}^T\mathbf{y}| \leq \sqrt{\sum_{n=1}^N |x_n|^2}\sqrt{\sum_{n=1}^N |y_n|^2} = \|\mathbf{x}\|_2\|\mathbf{y}\|_2.$$

8 Hölder's inequality, which plays an important role in different branches of mathematics, was proved by Hölder in 1889. Rogers (1862–1933) had given a proof of this inequality in a slightly different form in 1888. That explains why it should be referred to as the Rogers inequality, or at least as the Rogers–Hölder inequality, as suggested by L. Maligranda in "Why Hölder's inequality should be called Rogers inequality", *Mathematical Inequalities and Applications* 1(1), 1998.

PROPOSITION 3.5.– *For all* $\mathbf{x} \in \mathbb{R}^N$*, we have the following inequalities:*

$$\|\mathbf{x}\|_p \leq \|\mathbf{x}\|_q \quad , \quad 1 \leq q \leq p$$
$$\|\mathbf{x}\|_p \leq N^{1/p}\|\mathbf{x}\|_\infty \ , \ \forall p \geq 1,$$

*from which one can deduce:*

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{N}\|\mathbf{x}\|_\infty \ ; \ \|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq N\|\mathbf{x}\|_\infty \ ; \ \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{N}\|\mathbf{x}\|_2.$$

*Equalities occur when* $\mathbf{x}$ *has at most one non-zero component. Otherwise, these are strict inequalities.*

### 3.3.2.2. *Norms in spaces of infinite sequences*

PROPOSITION 3.6.– *In the space* $l^p(I, \mathbb{K})$*, with* $1 \leq p \leq \infty$ *and* $I = \mathbb{Z}$ *or* $\mathbb{N}$*, of infinite scalar sequences* $(x_n)_{n\in I}$*, defined and* $p$*-summable over* $I$*, with values in* $\mathbb{K}$*, the* $l_p$ *norm is written as:*

$$\|x\|_p = \Big(\sum_{n\in I} |x_n|^p\Big)^{1/p},$$

*where the* $p$*-summability property means that* $\sum_{n\in I} |x_n|^p < \infty$*. This norm generalizes the Hölder norm [3.19] to spaces of infinite sequences.*

When $p = 1$, the $p$-summability is called absolute summability. The corresponding $l_1$ norm allows to characterize the stability of an LTI system, the notion of stability meaning that for a bounded input, the system output is also bounded. It is then said that the system is BIBO (for bounded input–bounded output).

PROPOSITION 3.7.– *The discrete LTI system, described by the convolution equation* $y = h * x$*, is stable if its impulse response is absolutely summable, that is, if* $h$ *belongs to the space* $l^1$*, defined in [2.1].*

PROOF.– Assuming input $x$ is bounded by $M$, we have $|x_k| \leq M, \forall k$. The use of the generalized triangle inequality gives us:

$$y_n = \sum_{k=-\infty}^{\infty} h_k x_{n-k} \ \Rightarrow \ |y_n| = |\sum_{k=-\infty}^{\infty} h_k x_{n-k}|$$
$$\leq \sum_{k=-\infty}^{\infty} |h_k|\,|x_{n-k}| \leq M \sum_{k=-\infty}^{\infty} |h_k| = M\|h\|_1.$$

$|y_n|$ is thus bounded if $h$ is absolutely summable, that is, if $h \in l^1$ ($\|h\|_1 < \infty$). □

PROPOSITION 3.8.– *The $l_p$ norm can be generalized to the case of infinite vector sequences $(\mathbf{x}_n)_{n\in I}$ of the space $\mathbf{l}^p(I,\mathbb{K}^N)$, with $1 \leq p \leq \infty$ and $I = \mathbb{Z}$ or $\mathbb{N}$, defined and p-summable over $I$, with values in $\mathbb{K}^N$:*

$$\|\mathbf{x}\|_p = \Big(\sum_{n\in I} \|\mathbf{x}_n\|^p\Big)^{1/p},$$

*where $\|.\|$ can be any norm in $\mathbb{K}^N$. In general, the $l_p$ norm is considered, which gives the following mixed norm:*

$$\|\mathbf{x}\|_p = \Big(\sum_{n\in I} \|\mathbf{x}_n\|_p^p\Big)^{1/p}\ ,\ 1 \leq p \leq \infty.$$

EXAMPLE 3.9.– For $p = 2$, we have: $\|\mathbf{x}\|_2 = \Big(\sum\limits_{n\in I}\sum\limits_{i=1}^{N} |x_n(i)|^2\Big)^{1/2}$, where $x_n(i)$ is the $i$th component of $\mathbf{x}_n$, $1 \leq i \leq N$. Similarly, for $p = \infty$, we have[9]:

$$\|\mathbf{x}\|_\infty = \sup_{n\in I}\Big\{\max_{1\leq i\leq N}\{|x_n(i)|\}\Big\}.$$

### 3.3.2.3. *Norms in spaces of functions*

PROPOSITION 3.10.– *In the space $L^p(I,\mathbb{K})$ of scalar functions $f$, defined and p-integrable over the interval $I \subseteq \mathbb{R}$, with values in $\mathbb{K}$, the Hölder $L_p$ norm is defined as:*

$$\|f\|_p = \Big(\int_I |f(t)|^p dt\Big)^{1/p}\ ,\ 1 \leq p \leq \infty, \qquad [3.21]$$

*the property of p-integrability meaning that $\int_I |f(t)|^p dt < \infty$.*

We have the following special cases:

– For $p = 1$, we have the $L_1$ norm known as the norm of the convergence in mean:

$$\|f\|_1 = \int_I |f(t)|dt.$$

9 The supremum of a set $A \subset \mathbb{R}$, denoted by $\sup A$, is the least upper bound of $A$, such that $\sup A \leq M$ for all upper bounds $M$ of $A$. If $A = \{x_n : n \in I\}$, the supremum is denoted by $\sup\limits_{n\in I} x_n$.

– For $p = 2$, we have the $L_2$ norm called the norm of the convergence in quadratic mean, or in energy:

$$\|f\|_2 = \Big(\int_I |f(t)|^2 dt\Big)^{1/2}.$$

– For $p = \infty$, we have the $L_\infty$ norm called the uniform convergence norm:

$$\|f\|_\infty = \sup_{t\in I} |f(t)|.$$

We illustrate hereafter two inequalities previously introduced:

– For the norm $L_2$ and $(f, g) \in L^2\big(I, \mathbb{K}\big) \times L^2\big(I, \mathbb{K}\big)$, the inequality [3.14] is written as:

$$\sqrt{\int_I |f(t) + g(t)|^2 dt} \leq \sqrt{\int_I |f(t)|^2 dt} + \sqrt{\int_I |g(t)|^2 dt}.$$

– For $(f, g) \in L^p(I, \mathbb{K}) \times L^q(I, \mathbb{K})$, with $\frac{1}{p} + \frac{1}{q} = 1$, the Hölder inequality [3.20] becomes:

$$\Big|\int_I f(t)g(t)dt\Big| \leq \|f\|_p \|g\|_q = \Big(\int_I |f(t)|^p dt\Big)^{1/p} \Big(\int_I |g(t)|^q dt\Big)^{1/q}.$$

PROPOSITION 3.11.– *In the space $\mathbf{L}^p(I, \mathbb{K}^N)$ of vector functions $\mathbf{f}$, defined and $p$-integrable over the interval $I$, with values in $\mathbb{K}^N$, the norm $L_p$ is given by:*

$$\|\mathbf{f}\|_p = \Big(\int_I \|\mathbf{f}(t)\|^p \, dt\Big)^{1/p},$$

*where $\|.\|$ can be any norm in $\mathbb{K}^N$.*

*By considering the $l_p$ norm in $\mathbb{K}^N$, the following composite norm is defined:*

$$\|\mathbf{f}\|_p = \Big(\int_I \|\mathbf{f}(t)\|_p^p \, dt\Big)^{1/p} = \Big(\int_I \sum_{n=1}^{N} |f_n(t)|^p \, dt\Big)^{1/p},$$

*where $f_n(t)$ is the $n$th component of $\mathbf{f}(t)$.*

### 3.3.3. *Equivalent norms*

Two norms $\|.\|_{(1)}$ and $\|.\|_{(2)}$ defined over the same v.s. $E$ are said to be equivalent if there exists two positive numbers $a$ and $b$ such that, for all $x \in E$, we have:

$$a\,\|x\|_{(1)} \leq \|x\|_{(2)} \leq b\,\|x\|_{(1)}.$$

In a finite-dimensional v.s., all norms are equivalent, as illustrated by Proposition 3.5. Moreover, all norms lead to the same notion of convergence, which is not the case for infinite-dimensional v.s., for which different notions of convergence can be defined. See section 3.8.5, where different modes of convergence are described for a sequence of functions.

### 3.3.4. *Distance associated with a norm*

The norm is used as a measure of the length of a vector. It can also be used to define the distance between two elements $x$ and $y$ of $E$ as:

$$d(x,y) = \|x-y\|. \qquad [3.22]$$

Therefore, a normed $\mathbb{K}$-v.s. $E$ is a metric space for the distance (metric) induced by the norm.

The triangle inequality [3.17] can be expressed in terms of distance as:

$$\delta\big(\|x\|,\|y\|\big) \leq d(x,y),$$

which means, according to [3.10], that the map $x \mapsto \|x\|$ is 1-Lipschitzian and therefore continuous.

A metric associated with a norm satisfies the following properties:

$$\text{Translation invariance}: \quad \forall t \in E : d(x+t, y+t) = d(x,y) \qquad [3.23]$$

$$\text{Homogeneity}: \quad \forall \lambda \in \mathbb{K}: \ d(\lambda x, \lambda y) = |\lambda|\, d(x,y). \qquad [3.24]$$

The closed unit ball of a n.v.s. $E$ is defined as the set: $\overline{\mathcal{B}} = \{y \in E : \|y\| \leq 1\}$. Similarly, the closed ball of center $x$, with radius $r > 0$, is defined as the set $\overline{\mathcal{B}}_r(x) = \{y \in E : \|y - x\| \leq r\}$.

For instance, in $\mathbb{R}^N$ with the $l_p$ norm, the closed ball around the origin, with radius $r$, is defined as:

$$\overline{\mathcal{B}}_r(\mathbf{0}) = \{\mathbf{y} \in \mathbb{R}^N : \|\mathbf{y}\|_p = (\sum_{n=1}^{N} |\, y_n\,|^p)^{1/p} \leq r\}.$$

## 3.4. Pre-Hilbert spaces

Pre-Hilbert spaces, also called as inner product spaces, are real or complex (incomplete) vector spaces equipped with an inner product. When these spaces

are finite-dimensional, they are called Euclidean spaces in the real case and Hermitian (or unitary) spaces in the complex case. The notion of inner product was introduced in the 19th century by Hermann Grassmann (1809–1877) and William Kingdon Clifford (1845–1879), who were, respectively, German and English mathematicians, and the Irish mathematician, physicist, and astronomer William Rowan Hamilton (1805–1865), fathers of geometric algebra and quaternions, respectively. In section 3.7, we will see that a complete space equipped with an inner product is called a Hilbert space. Based on inner product, we can define the important notions of induced norm and distance, and also geometrical notions such as the length of a vector, the angle and orthogonality between two of them, and thus the definition of orthonormal basis.

## 3.4.1. *Real pre-Hilbert spaces*

### 3.4.1.1. *Euclidean inner product*

Given a finite-dimensional real v.s. $E$, a Euclidean inner product over $E$, denoted by $\langle ., . \rangle$, refers to any bilinear form:

$$E^2 \ni (x, y) \;\mapsto\; \langle x, y \rangle \in \mathbb{R} \qquad [3.25]$$

which is symmetric and positive definite, that is, satisfying the following properties:

$$\langle x, y \rangle = \langle y, x \rangle \;,\; \forall (x, y) \in E^2 \qquad [3.26]$$

$$\langle x, x \rangle = 0 \text{ if and only if } x = 0 \qquad [3.27]$$

$$\langle x, x \rangle \geq 0 \;,\; \forall x \in E. \qquad [3.28]$$

Due to the bilinearity property, the inner product is linear in both variables $x$ and $y$.

It is said that $E$ equipped with this inner product is a real pre-Hilbert space, and it is usually denoted by $(E, \langle ., . \rangle)$.

### 3.4.1.2. *Examples of Euclidian norms and inner products*

In Table 3.1, we present inner products and norms over three real pre-Hilbert spaces, where $\mathcal{C}^0([a, b], \mathbb{R})$ and $\mathcal{C}^0_{2\pi}(\mathbb{R}, \mathbb{R})$ denote the v.s. of continuous functions over $[a, b]$, with real values, and of $2\pi$-periodic continuous functions, from $\mathbb{R}$ to $\mathbb{R}$, respectively.

## 3.4.2. *Complex pre-Hilbert spaces*

### 3.4.2.1. *Hermitian inner product*

In the case of a finite-dimensional complex v.s. $E$, an Hermitian inner product over $E$ designates any positive definite sesquilinear form:

$$E^2 \ni (x, y) \;\mapsto\; \langle x, y \rangle \in \mathbb{C}$$

| Spaces | Inner products | Norms |
|---|---|---|
| $\mathbb{R}^N$ | $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^{N} x_n y_n = \mathbf{y}^T \mathbf{x}$ | $\|\mathbf{x}\|_2 = \left( \sum_{n=1}^{N} x_n^2 \right)^{1/2}$ |
| $\mathcal{C}^0([a,b], \mathbb{R})$ | $\langle f, g \rangle = \int_a^b f(t) g(t) dt$ | $\|f\|_2 = \left( \int_a^b f^2(t) dt \right)^{1/2}$ |
| $\mathcal{C}^0_{2\pi}(\mathbb{R}, \mathbb{R})$ | $\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t) g(t) dt$ | $\|f\|_2 = \left( \frac{1}{2\pi} \int_0^{2\pi} f^2(t) dt \right)^{1/2}$ |

**Table 3.1.** *Examples of Euclidian norms and inner products*

such that properties [3.27] and [3.28] are satisfied, and the symmetry property [3.26] becomes a Hermitian symmetry:

$$\langle x, y \rangle = \overline{\langle y, x \rangle}, \qquad [3.29]$$

where $\overline{a}$ denotes the conjugate of the complex number $a$, also denoted by $a^*$.

The aforementioned equality means that for any $y \in E$, the map $E \to \mathbb{C} : \ x \mapsto \langle x, y \rangle$ is linear, whereas for any $x \in E$, the map $E \to \mathbb{C} : \ y \mapsto \langle x, y \rangle$ is semi-linear (also called anti-linear or conjugate linear), that is:

$$\forall (x_1, x_2) \in E^2, \ \forall \lambda \in \mathbb{C}, \ \langle x_1 + \lambda x_2, y \rangle = \langle x_1, y \rangle + \lambda \langle x_2, y \rangle \qquad [3.30]$$

$$\forall (y_1, y_2) \in E^2, \ \forall \lambda \in \mathbb{C}, \ \langle x, y_1 + \lambda y_2 \rangle = \langle x, y_1 \rangle + \overline{\lambda} \langle x, y_2 \rangle. \qquad [3.31]$$

It is said that $E$ equipped with an Hermitian inner product is a complex pre-Hilbert space, also called a Hermitian (or unitary) space.

NOTE 3.12.– In some books, the map $x \mapsto \langle x, y \rangle$ is assumed to be semi-linear, while the map $y \mapsto \langle x, y \rangle$ is linear.

### 3.4.2.2. *Examples of Hermitian norms and inner products*

In Table 3.2, we present inner products and norms over three complex pre-Hilbert spaces, where $\mathcal{C}^0([a,b], \mathbb{C})$ and $\mathcal{C}^0_T(\mathbb{R}, \mathbb{C})$ denote the spaces of continuous functions on $[a,b]$, with complex values, and of $T$-periodic continuous functions from $\mathbb{R}$ to $\mathbb{C}$, respectively.

FACT 3.13.– For the Euclidean and Hermitian inner products, we have, respectively:

$$\langle \mathbf{Ax}, \mathbf{y} \rangle = \mathbf{y}^T \mathbf{Ax} = \langle \mathbf{x}, \mathbf{A}^T \mathbf{y} \rangle \text{ for } \mathbf{x}, \mathbf{y} \in \mathbb{R}^N$$

$$\langle \mathbf{Ax}, \mathbf{y} \rangle = \mathbf{y}^H \mathbf{Ax} = \langle \mathbf{x}, \mathbf{A}^H \mathbf{y} \rangle \text{ for } \mathbf{x}, \mathbf{y} \in \mathbb{C}^N.$$

| Spaces | Inner products | Norms |
|---|---|---|
| $\mathbb{C}^N$ | $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^{N} x_n \overline{y}_n = \mathbf{y}^H \mathbf{x}$ | $\|\mathbf{x}\|_2 = \left( \sum_{n=1}^{N} \|x_n\|^2 \right)^{1/2}$ |
| $\mathcal{C}^0([a,b], \mathbb{C})$ | $\langle f, g \rangle = \int_a^b f(t)\overline{g}(t)dt$ | $\|f\|_2 = \left( \int_a^b \|f(t)\|^2 dt \right)^{1/2}$ |
| $\mathcal{C}_T^0(\mathbb{R}, \mathbb{C})$ | $\langle f, g \rangle = \frac{1}{T} \int_0^T f(t)\overline{g}(t)dt$ | $\|f\|_2 = \left( \frac{1}{T} \int_0^T \|f(t)\|^2 dt \right)^{1/2}$ |

**Table 3.2.** *Examples of Hermitian norms and inner products*

### 3.4.3. *Norm induced from an inner product*

#### 3.4.3.1. *Definition and equalities*

The quantity $\|x\| = \sqrt{\langle x, x \rangle}$ represents the norm of the vector $x$, induced from the inner product $\langle ., . \rangle$. It is also known as the norm associated with the inner product. It is called Euclidean norm in the case of a real v.s. and Hermitian norm in the case of a complex v.s. As a result, a Euclidean or Hermitian v.s. is a normed v.s. If $\|x\| = 1$, it is said that $x$ is unitary.

PROPOSITION 3.14.– *The induced norm satisfies the following properties:*

*– For all* $(x_1, \cdots, x_N) \in E^N$*, taking into account the Hermitian symmetry property [3.29] gives:*

$$\begin{aligned} \| \sum_{n=1}^{N} x_n \|^2 &= \sum_{n=1}^{N} \|x_n\|^2 + \sum_{1 \leq i,j \leq N, i \neq j} \langle x_i, x_j \rangle \\ &= \sum_{n=1}^{N} \|x_n\|^2 + 2 \sum_{1 \leq i < j \leq N} Re\big(\langle x_i, x_j \rangle\big), \end{aligned} \quad [3.32]$$

*– For* $N = 2$*, the previous equality becomes:*

$$\|x + y\|^2 = \|x\|^2 + 2Re\big(\langle x, y \rangle\big) + \|y\|^2 \ , \ \forall (x, y) \in E^2. \quad [3.33]$$

*By replacing* $y$ *by* $-y$*, we obtain:*

$$\|x - y\|^2 = \|x\|^2 - 2Re\big(\langle x, y \rangle\big) + \|y\|^2 \ , \ \forall (x, y) \in E^2, \quad [3.34]$$

*and by summing the last two equalities member-wise, the equality (or identity) of the parallelogram can be deduced:*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2), \ \forall (x, y) \in E^2. \quad [3.35]$$

In geometry, this equality reflects the fact that, in the plane, the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of its sides.

It characterizes the existence of an inner product based on a norm. This means that an inner product cannot be associated with a norm that does not satisfy equality [3.35]. Thereby, for example, for the $l_p$ Hölder norm defined in [3.19] and for $E = \mathbb{R}^2$, with $\mathbf{x} = [\,0\,1\,]^T$ and $\mathbf{y} = [\,1\,0\,]^T$, the two members of [3.35] are $\|\mathbf{x}+\mathbf{y}\|_p^2 + \|\mathbf{x}-\mathbf{y}\|_p^2 = 2 \times 2^{2/p}$ and $2(\|\mathbf{x}\|_p^2 + \|\mathbf{y}\|_p^2) = 4$. As a result, the equality of the parallelogram is only satisfied for $p = 2$, which implies that an inner product can only be associated with the $l_p$ Hölder norm, for $p = 2$.

In section 3.7.3, we will see that this remark implies that a Banach space $(E, \|.\|)$ is a Hilbert space if and only if its norm $\|.\|$ satisfies the parallelogram equality in order to be able to associate it with an inner product.

In the following two sections, we introduce the Cauchy–Schwarz and Minkowski inequalities.

#### 3.4.3.2. *Cauchy–Schwarz inequality*

In the case of a pre-Hilbert space $E$, the Cauchy[10]–Schwarz[11] inequality links the inner product to the induced norm, an inequality not satisfied by any norm.

For all $(x, y) \in E^2$, the Cauchy–Schwarz inequality is written as:

$$|\langle x, y\rangle| \leq \|x\|\,\|y\|, \qquad [3.36]$$

with equality when $x$ and $y$ are collinear (i.e. when there exists $\lambda \in \mathbb{K}$ such that $y = \lambda x$), or if $x$ or $y$ is zero.

---

10 Augustin Louis Cauchy (1789–1857), a French mathematician, member of the Académie des Sciences, and professor at École Polytechnique, who is one of the most prolific mathematicians, with close to 800 articles published in his collected works ("Oeuvres complètes d'Augustin Cauchy"), in 27 volumes. He strongly influenced the development of mathematics in the 19th century. His contributions cover almost all branches of mathematics, from analysis, algebra, geometry, and probability, up to astronomy, mechanics, and optics. His major contributions include his works on the convergence of series and the theory of complex functions. He is the one who defined, for the first time, complex functions of one complex variable and introduced the concepts of curvilinear integral and of residue calculus in complex analysis. In optics, he studied the propagation of light based on that of waves. In 1836, he established an empirical relationship, called Cauchy's law, between the refractive index and wavelength of light for a particular transparent material, and he highlighted the phenomenon of diffraction. Many mathematical results (criterion, theorems, equations, integral formula, inequality, and law) bear his name. Devout Catholic and close to the Jesuits, he founded the Institut Catholique, consisting today of seven private institutions for higher education.

11 Hermann Amandus Schwarz (1843–1921), a German mathematician who was a student of Karl Weierstrass. His contributions concern real and complex analysis, differential geometry, and the study of conformal mappings, including in particular the demonstration of the conformal mapping theorem, due to Riemann.

EXAMPLE 3.15.– In $\mathbb{C}^N$ and $\mathcal{C}^0([a,b],\mathbb{C})$, the Cauchy–Schwarz inequality becomes:

$$\left|\sum_{n=1}^{N} x_n \overline{y}_n\right|^2 \leq \sum_{n=1}^{N} |x_n|^2 \sum_{n=1}^{N} |y_n|^2 \qquad [3.37a]$$

$$\left|\int_a^b f(t)\overline{g}(t)dt\right|^2 \leq \int_a^b |f(t)|^2 dt \int_a^b |g(t)|^2 dt. \qquad [3.37b]$$

NOTE 3.16.– In $E = \mathbb{R}^N$, the Cauchy–Schwarz inequality can be interpreted in terms of Euclidean geometry with:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \|\mathbf{y}\| \cos\theta \leq \|\mathbf{x}\| \|\mathbf{y}\|,$$

where $\theta$ is the angle between vectors $\mathbf{x}$ and $\mathbf{y}$. When $\theta = \frac{\pi}{2}$, we have $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, which reflects the perpendicularity of vectors $\mathbf{x}$ and $\mathbf{y}$.

### 3.4.3.3. *Minkowski inequality*

The Minkowski[12] inequality is written as, for all $(x, y) \in E^2$:

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p \ , \ 1 \leq p < \infty. \qquad [3.38]$$

Equality holds if $x$ or $y$ is the null vector. The Minkowski inequality can be seen as the triangle inequality [3.14] for the $l_p$ norm.

For instance:

– For $E = \mathbb{C}^N$, and the norm $l_p$, the Minkowski inequality is written as:

$$\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p \ , \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^N \qquad [3.39a]$$

$$\Big(\sum_{n=1}^{N} |x_n + y_n|^p\Big)^{1/p} \leq \Big(\sum_{n=1}^{N} |x_n|^p\Big)^{1/p} + \Big(\sum_{n=1}^{N} |y_n|^p\Big)^{1/p}. \qquad [3.39b]$$

12 Hermann Minkowski (1864–1909), a German mathematician and physicist who received the Grand Prize of the Académie des Sciences of Paris (shared with Henry Smith) in 1881, at the age of 18 years, before obtaining his doctorate in 1885, on the arithmetic theory of quadratic forms. He had Albert Einstein as a student at the Zurich Federal Polytechnic School and collaborated with David Hilbert. He founded the geometry of numbers, a branch of number theory, used today in cryptology, and in continuity with Hendrick Lorenz's (1853–1928) and Albert Einstein's (1879–1955) works, he introduced the notion of space-time continuum of dimension 4, called Minkowski space, providing a framework for all subsequent works on the theory of relativity. His results strongly influenced Einstein in the development of his general theory of relativity. For more detail, refer to the Sébastien Gauthier's (2009) article: "Hermann Minkowski: des formes quadratiques à la géométrie des nombres."

– Similarly, for two functions $f$ and $g$ of the space $L^p(I, \mathbb{K})$, with $I \subseteq \mathbb{R}$ and the norm $L_p$, the Minkowski inequality is written as:

$$\|f + g\|_p \leq \|\mathbf{f}\|_p + \|\mathbf{g}\|_p$$
$$\Big(\int_I |f(t) + g(t)|^p dt\Big)^{1/p} \leq \Big(\int_I |f(t)|^p dt\Big)^{1/p} + \Big(\int_I |g(t)|^p dt\Big)^{1/p}.$$

#### 3.4.3.4. *Polarization formulae*

Since $\langle x, iy\rangle = -i\langle x, y\rangle$, with $i^2 = -1$, we have:

$$\|x \pm iy\|^2 = \|x\|^2 \pm 2\mathrm{Im}\big(\langle x, y\rangle\big) + \|y\|^2, \ \forall (x, y) \in E^2. \qquad [3.40]$$

From relations [3.33], [3.34], and [3.40], the following polarization formulae can be deduced:

$$\mathrm{Re}\big(\langle x, y\rangle\big) = \frac{\|x + y\|^2 - \|x - y\|^2}{4} \ , \ \mathrm{Im}\big(\langle x, y\rangle\big) = \frac{\|x + iy\|^2 - \|x - iy\|^2}{4}.$$

These formulae allow us to define the Hermitian inner product based on the norm:

$$\langle x, y\rangle = \frac{\|x + y\|^2 - \|x - y\|^2}{4} + i\frac{\|x + iy\|^2 - \|x - iy\|^2}{4}. \qquad [3.41]$$

FACT 3.17.– In the case of a real v.s. $E$, the aforementioned formula can be simplified as:

$$\langle x, y\rangle = \frac{\|x + y\|^2 - \|x - y\|^2}{4}. \qquad [3.42]$$

From [3.33], the Euclidean inner product can also be expressed as:

$$\langle x, y\rangle = \frac{\|x + y\|^2 - \|x\|^2 - \|y\|^2}{2}. \qquad [3.43]$$

### 3.4.4. *Distance associated with an inner product*

A distance can be associated with any inner product according to the following formula:

$$d(x, y) = \sqrt{\langle x - y, x - y\rangle} = \|x - y\|, \ \forall x, y \in E,$$

where $\|x-y\|$ is the norm of $x-y$ induced from the inner product. Subsequently, any pre-Hilbert space is a metric space for the distance induced from the inner product.

### 3.4.5. *Weighted inner products*

For example, in the spaces $\mathbb{C}^N$ and $\mathcal{C}^0([a,b],\mathbb{C})$, weighted inner products can be defined as:

$$\langle \mathbf{x}, \mathbf{y} \rangle_\rho = \sum_{n=1}^{N} \rho_n x_n \overline{y}_n \qquad [3.44a]$$

$$\langle f, g \rangle_\rho = \int_a^b \rho(t) f(t) \overline{g}(t) dt, \qquad [3.44b]$$

where $\rho_n$ and $\rho(t)$ are positive weightings. This type of weighted inner product will be used in sections 3.6.6 and 3.9 to build bases of orthogonal polynomials.

In Table 3.3, we summarize the main results presented relative to Hermitian norms and inner products.

| Properties | Relations |
|---|---|
| Triangle inequality | $\|x+y\| \leq \|x\| + \|y\|$ |
| Other inequality | $\|\|x\| - \|y\|\| \leq \|x \pm y\|$ |
| Hermitian symmetry | $\langle x, y \rangle = \overline{\langle y, x \rangle}$ |
| Induced norm | $\|x\| = \sqrt{\langle x, x \rangle}$ |
| Cauchy-Schwarz inequality | $\|\langle x, y \rangle\| \leq \|x\| \, \|y\|$ |
| Norm of $x \pm y$ | $\|x \pm y\|^2 = \|x\|^2 \pm 2\mathrm{Re}\big(\langle x, y \rangle\big) + \|y\|^2$ |
| Parallelogram equality | $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$ |
| Norm of $x \pm iy$ | $\|x \pm iy\|^2 = \|x\|^2 \pm 2\mathrm{Im}\big(\langle x, y \rangle\big) + \|y\|^2$ |
| Polarization identity | $\langle x, y \rangle = \frac{\|x+y\|^2 - \|x-y\|^2}{4} + i \frac{\|x+iy\|^2 - \|x-iy\|^2}{4}$ |

**Table 3.3.** *Properties of Hermitian norms and inner products*

## 3.5. Orthogonality and orthonormal bases

The inner product allows us to define the notion of orthogonality which in turn makes it possible to define the notions of orthogonal subspaces and orthogonal complement. Next, the definition of orthonormal basis will be given before presenting the notion of orthogonal projection onto a subspace and the Gram–Schmidt orthonormalization method.

### 3.5.1. *Orthogonal/perpendicular vectors and Pythagorean theorem*

In the following, we define the concepts of orthogonal vectors and perpendicular vectors and present the Pythagorean theorem.

– Vectors $x, y \in E$, assumed to be non-null, are said to be orthogonal, and denoted by $x \perp y$, if their inner product is zero:

$$x \perp y \;\Leftrightarrow\; \langle x, y\rangle = 0 \;\Leftrightarrow\; \mathrm{Re}\big(\langle x, y\rangle\big) = \mathrm{Im}\big(\langle x, y\rangle\big) = 0. \qquad [3.45]$$

EXAMPLE 3.18.– In $\mathbb{R}^N$ and $\mathbb{C}^N$, we have:

$$\mathbf{x}, \mathbf{y} \in \mathbb{R}^N,\ \mathbf{x} \perp \mathbf{y} \Leftrightarrow \sum_{n=1}^{N} x_n y_n = \mathbf{y}^T \mathbf{x} = 0. \qquad [3.46a]$$

$$\mathbf{x}, \mathbf{y} \in \mathbb{C}^N,\ \mathbf{x} \perp \mathbf{y} \Leftrightarrow \sum_{n=1}^{N} x_n \overline{y}_n = \mathbf{y}^H \mathbf{x} = 0. \qquad [3.46b]$$

where $\mathbf{y}^H = \overline{\mathbf{y}}^T = \big[\,\overline{y}_1, \overline{y}_2, \cdots \overline{y}_N\,\big]^T$ is the transconjugated vector of $\mathbf{y}$.

– Vectors $x, y \in E$ are said to be perpendicular when only the real part of the inner product is zero, that is $\mathrm{Re}\big(\langle x, y\rangle\big) = 0$.

It is important to note that the notions of orthogonality and perpendicularity are equivalent in the case of real pre-Hilbert spaces since then we have $\mathrm{Re}\big(\langle x, y\rangle\big) = \langle x, y\rangle$, and $\mathrm{Re}\big(\langle x, y\rangle\big)$ equal to zero implies that the inner product is also equal to zero, whereas in the case of complex pre-Hilbert spaces, this equality to zero implies that $\langle x, y\rangle = i\mathrm{Im}\big(\langle x, y\rangle\big)$ which may be non-zero.

Based on [3.33], when $\mathrm{Re}\big(\langle x, y\rangle\big) = 0$, we obtain the Pythagorean theorem:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Conversely, in the case of a complex pre-Hilbert space, when the Pythagorean theorem is verified, this implies that $x$ and $y$ are perpendicular, but not that they are orthogonal.

It can be concluded that:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 \;\Leftrightarrow\; x \text{ and } y \text{ perpendicular},$$

which corresponds to the Pythagorean theorem in geometry.

## 3.5.2. *Orthogonal subspaces and orthogonal complement*

### 3.5.2.1. *Definitions*

We present here a few definitions related to the notion of orthogonality:

– Let $F$ denote a subspace of $E$. The vector $x \in E$ is orthogonal to $F$ if it is orthogonal to any vector of $F$.

– A family $\mathcal{X} = \{x_1, \cdots, x_N\}$ of non-zero vectors is orthogonal if and only if $\langle x_i, x_j \rangle = 0$ for all $i, j \in \langle N \rangle$, with $i \neq j$.

– When vectors are mutually perpendicular, such a family of vectors satisfies the generalized Pythagorean theorem, which may be inferred from [3.32]:

$$\| \sum_{n=1}^{N} x_n \|^2 = \sum_{n=1}^{N} \|x_n\|^2.$$

It can be concluded that the square of the norm of a sum of mutually perpendicular vectors is equal to the sum of the squares of the norm of each vector.

– Two subspaces $F$ and $G$ of $E$ are said to be orthogonal, denoted by $F \perp G$, if any vector of $F$ is orthogonal to any vector of $G$, namely:

$$F \perp G \;\; \Leftrightarrow \;\; \forall x \in F, \forall y \in G \;,\; x \perp y.$$

Subsequently, any basis of $F$ is orthogonal to any basis of $G$.

– Given a subspace $F$ of a finite-dimensional Euclidean v.s. $E$, its orthogonal space, denoted by $F^{\perp}$, is defined as:

$$F^{\perp} = \{x \in E : \forall y \in F, \langle x, y \rangle = 0\}.$$

that is, the set of vectors of $E$ orthogonal to all vectors of $F$.

In Table 3.4, we summarize the definitions related to the previously outlined perpendicularity and orthogonality properties.

| Membership | Properties | Definitions |
|---|---|---|
| $x, y \in E$ | Perpendicularity | $\mathrm{Re}\big(\langle x, y \rangle\big) = 0$ |
| $x, y \in E$ | Pythagorean theorem | $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ |
| $x, y \in E$ | Orthogonality $(x \perp y)$ | $\langle x, y \rangle = 0$ |
| $x \in E$ , $F \subset E$ | $x \perp F$ | $\langle x, y \rangle = 0$ , $\forall y \in F$ |
| $F, G \subset E$ | $F \perp G$ | $\langle x, y \rangle = 0$ , $\forall x \in F$, $\forall y \in G$ |
| $F \subset E$ | $F^{\perp}$ | $F^{\perp} = \{x \in E : \forall y \in F, \langle x, y \rangle = 0\}$ |

**Table 3.4.** *Perpendicularity and orthogonality properties*

### 3.5.2.2. *Orthogonal complement*

The set $F^{\perp}$ is a subspace of $E$ orthogonal to $F$. It is called the orthogonal complement of $F$, and it is said that $E = F \oplus F^{\perp}$ is the orthogonal direct sum of $F$ and $F^{\perp}$, with $\dim(F^{\perp}) = \dim(E) - \dim(F)$.

We have the following properties:

– $(F^{\perp})^{\perp} = F$.

– If $G$ is a subset of the subspace $F$, then: $G \subset F \Leftrightarrow F^{\perp} \subset G^{\perp}$.

These properties do not always hold if $E$ is infinite-dimensional.

More generally, given $N$ subspaces $F_1, \cdots, F_N$ of an inner product space $E$, it is said that $E$ is the orthogonal direct sum of $F_1, \cdots, F_N$ if $E$ is a direct sum ($E = \bigoplus_{n=1}^{N} F_n$) of the subspaces $F_n, n \in \langle N \rangle$, which are pairwise orthogonal.

### 3.5.3. *Orthonormal bases*

For an $N$-dimensional Euclidean/Hermitian space $E$, a basis $\{e_n, n \in \langle N \rangle\}$ is said to be orthonormal if $\langle e_n, e_p \rangle = \delta_{np}, \ \forall n, p \in \langle N \rangle$, where $\delta_{np}$ is the Kronecker delta. This means that all its vectors are pairwise orthogonal ($\langle e_n, e_p \rangle = 0, \forall n \neq p$), and unitary ($\|e_n\| = 1, \forall n \in \langle N \rangle$).

If the vectors of a basis $\{b_n, n \in \langle N \rangle\}$ are orthogonal but non-unitary, the basis is said to be orthogonal, and then $\{\frac{b_n}{\|b_n\|}, n \in \langle N \rangle\}$ is an orthonormal basis.

### 3.5.4. *Orthogonal/unitary endomorphisms and isometries*

Let $E$ be a finite-dimensional Euclidean v.s., and $f \in \mathcal{L}(E)$. It is said that $f$ is an orthogonal endomorphism if one of the following two equivalent conditions is satisfied:

$$\forall (x, y) \in E^2 \, , \, \langle f(x), f(y) \rangle = \langle x, y \rangle \qquad [3.47]$$

$$\forall x \in E \, , \, \|f(x)\| = \|x\|. \qquad [3.48]$$

It is then said that $f$ preserves the inner product and the norm. Such orthogonal endomorphism is also called an isometry or an orthogonal transformation.

In the case where $E$ is a Hermitian space, conditions [3.47] and [3.48] define a unitary endomorphism.

PROPOSITION 3.19.– *The set of orthogonal endomorphisms of $E$ is a group (endowed with the law of map composition) called the orthogonal group of $E$, and denoted by $\mathcal{O}(E)$.*

*Similarly, in the case where $E$ is a Hermitian space, the set of unitary endomorphisms of $E$ is called the unitary group of $E$, and denoted by $\mathcal{U}(E)$.*

PROPOSITION 3.20.– *Orthogonal endomorphisms satisfy the following properties:*

*– The matrix* $\mathbf{A}$ *associated with an orthogonal endomorphism is orthogonal, that is, it satisfies the property:* $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}$.

*– An orthogonal endomorphism transforms an orthonormal basis into an orthonormal basis.*

*– If* $f$ *and* $g$ *are two orthogonal endomorphisms, then their composite* $g \circ f$ *is itself an orthogonal endomorphism. This infers that the product of two orthogonal matrices is also an orthogonal matrix.*

PROOF.– $\mathbf{A}$ and $\mathbf{B}$ being matrices associated with $g$ and $f$, respectively, then, as will be demonstrated in Proposition 4.53, $\mathbf{AB}$ is the matrix associated with $g \circ f$, and we have:

$$(\mathbf{AB})(\mathbf{AB})^T = \mathbf{ABB}^T\mathbf{A}^T = \mathbf{A}\mathbf{A}^T = \mathbf{I},$$
$$(\mathbf{AB})^T(\mathbf{AB}) = \mathbf{B}^T\mathbf{A}^T\mathbf{AB} = \mathbf{B}^T\mathbf{B} = \mathbf{I},$$

which shows that $\mathbf{AB}$ is an orthogonal matrix . □

FACT 3.21.– In the case of a finite-dimensional Hermitian v.s., the matrix associated with a unitary endomorphism is a unitary matrix satisfying the property: $\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A} = \mathbf{I}$.

## 3.6. Gram–Schmidt orthonormalization process

### 3.6.1. *Orthogonal projection onto a subspace*

Let $F$ be a subspace of a finite-dimensional Euclidean v.s. $E$. An orthogonal projection onto $F$, denoted by $p_F$, is the projection onto $F$ parallel to $F^\perp$. For any $x \in E, p_F(x)$ is called the orthogonal projection of $x$ onto $F$ and is characterized by:

$$p_F(x) \in F \quad , \quad x - p_F(x) \in F^\perp.$$

### 3.6.2. *Orthogonal projection and Fourier expansion*

Given a basis $\mathcal{B} = \{b_1, \cdots, b_N\}$ of $F$, the orthogonal projection $p_F(x)$ is expressed as $p_F(x) = \sum_{n=1}^{N} \lambda_n b_n$. The coefficients $\lambda_n$ are determined such that $x - p_F(x) \in F^\perp$, and thus $x - p_F(x) \perp b_n$ , for $n \in \langle N \rangle$, which amounts to solving the linear system of equations:

$$\langle x - p_F(x), b_n \rangle = 0 \ , \ n \in \langle N \rangle. \qquad [3.49]$$

It is said that the projector satisfies the orthogonality principle.

When the basis $\{b_1, \cdots, b_N\}$ is orthogonal, equations [3.49] can be solved in a decoupled manner, which gives $\lambda_n = \frac{\langle x, b_n \rangle}{\|b_n\|^2}$. The orthogonal projection is then written as:

$$p_F(x) = \sum_{n=1}^{N} \frac{\langle x, b_n \rangle}{\|b_n\|^2} b_n. \qquad [3.50]$$

If the basis is orthonormal, projector [3.50] can be simplified as:

$$p_F(x) = \sum_{n=1}^{N} \langle x, b_n \rangle b_n = \sum_{n=1}^{N} c_n b_n \qquad [3.51a]$$

$$c_n = \langle x, b_n \rangle. \qquad [3.51b]$$

This projection formula is at the base of the Gram–Schmidt (GS) orthonormalization process (see section 3.6.4).

The term $c_n b_n$ represents the orthogonal projection of $x$ onto the space generated by $b_n$. The expansion $p_F(x)$, which expresses $x$ in the form of a sum of $N$ mutually orthogonal vectors, is called the Fourier expansion of $x$. The scalars $c_n$ are the coordinates of $x$ in the basis $\mathcal{B}$ and are called the Fourier coefficients.

The squared projection error, or more specifically, the square of the norm of the difference between $x$ and its orthogonal projection onto $F$, is given by:

$$\|x - p_F(x)\|^2 = \|x\|^2 - \|p_F(x)\|^2 = \|x\|^2 - \sum_{n=1}^{N} |c_n|^2. \qquad [3.52]$$

PROOF.– Note $x$ is decomposed into the sum of two orthogonal vectors $x = p_F(x) + (x - p_F(x))$, with $p_F(x) \in F$ and $x - p_F(x) \in F^{\perp}$. Application of the Pythagorean theorem, with the orthonormality of the basis $\{b_1, \cdots, b_N\}$, gives:

$$\|(x - p_F(x)) + p_F(x)\|^2 = \|x - p_F(x)\|^2 + \|p_F(x)\|^2 = \|x\|^2 \qquad [3.53a]$$

$$\|p_F(x)\|^2 = \sum_{n=1}^{N} |c_n|^2. \qquad [3.53b]$$

From these last two equations, equality [3.52] can be deduced. □

In the next section, we present the Bessel[13] inequality and Parseval's equality, also called Parseval's theorem[14].

### 3.6.3. *Bessel's inequality and Parseval's equality*

PROPOSITION 3.22.– *Bessel's inequality and Parseval's equality are given by:*

$$\sum_{n=1}^{N} |c_n|^2 \leq \|x\|^2 \quad ; \quad \sum_{n=1}^{N} |c_n|^2 = \|x\|^2 \quad \textit{when} \quad x \in F. \qquad [3.54]$$

PROOF.– The inequality directly results from equation [3.52]:

$$\sum_{n=1}^{N} |c_n|^2 = \|x\|^2 - \|x - p_F(x)\|^2 \leq \|x\|^2. \qquad [3.55]$$

The equality is also obtained from [3.52] when $x \in F$, which implies that $x = p_F(x)$, and so $x - p_F(x) = 0$. □

We can make the following observations:

– Since $p_F(x) \in F$, we have $p_F(p_F(x)) = p_F(x)$, which characterizes a projector, and the idempotence property. See the definition of an idempotent matrix in Table 4.3.

– In the context of the theory of estimation in the least squares sense, $p_F(x)$ represents the estimate $\widehat{x}_{MC}$ of $x$ using the data of the subspace $F$ and $\|x - \widehat{x}_{MC}\|^2 = \|x - p_F(x)\|^2$ is the quadratic error of estimation.

– From a geometric point of view, the orthogonal projection of $x$ onto $F$ minimizes the following criterion:

$$\min_{y \in F} d^2(x, y) = \min_{y \in F} \|x - y\|_2^2, \qquad [3.56]$$

that is, the orthogonal projection $y = p_F(x)$ is the closest point of $F$ to $x$.

---

13 Friedrich Wilhelm Bessel (1784–1846), a German astronomer and mathematician who introduced the Bessel functions for studying the movement of planets. These functions are used for solving linear differential equations of the second order named after him.

14 Marc-Antoine Parseval des Chênes (1755–1836), French mathematician who published his theorem in "Mémoire sur les séries et sur l'intégration complète d'une équation aux différences partielles linéaires du second ordre, à coefficiens constants" in 1799. This theorem is used mainly in the context of Fourier series.

– From the point of view of the theory of function approximation, equality [3.51a] can be extended to the expansion of a function $f(t)$ over a basis of orthonormal functions $\{e_n(t)), n \in \mathbb{Z}\}$ in the following form:

$$f(t) \simeq \sum_{n\in\mathbb{Z}} c_n e_n(t). \qquad [3.57]$$

Thereby, the aforementioned results presented for the orthogonal projection onto a subspace of a finite-dimensional v.s. can be generalized to the case of an infinite-dimensional Hilbert space $E$, for the orthogonal projection onto any closed subspace $F$ of $E$, generated by a finite or countable orthonormal sequence, that is, a Hilbert basis.

This is the case of Fourier series expansions that will be considered in section 3.8, and for which bases of trigonometric functions are used, within the context of a Hilbertian approach (see [3.86a]).

It is also the case of polynomial series expansions using bases of orthogonal polynomials, as considered in section 3.9.

In Table 3.5, we summarize the results on orthogonal projection.

| Properties | Relations |
|---|---|
| Orthogonal projection | $p_{\mathcal{B}}(x) = \sum_{n=1}^{N} \langle x, b_n \rangle\, b_n = \sum_{n=1}^{N} c_n b_n$ |
| Bessel's inequality | $\sum_{n=1}^{N} \lvert c_n \rvert^2 \leq \lVert x \rVert^2$ |
| Parseval's equality | $\sum_{n=1}^{N} \lvert c_n \rvert^2 = \lVert x \rVert^2 \ , \ x \in F$ |

**Table 3.5.** *Orthogonal projection of $x \in E$ onto the orthonormal basis $\mathcal{B} = \{b_1, \cdots, b_N\}$ of $F \subset E$*

### 3.6.4. *Gram–Schmidt orthonormalization process*

This method, named after the Danish and German mathematicians Jörgen Pedersen Gram (1850–1916) and Erhard Schmidt (1876–1959), was in fact already known to the French mathematician Pierre-Simon Laplace (1749–1827) in 1816.

Let $E$ denote a real pre-Hilbert v.s. of dimension $N$, admitting a free family of vectors $\{u_1, \cdots, u_N\}$. The Gram–Schmidt orthonormalization process is a method

for building, in an iterative way, an orthonormal family $\{e_n, n \in \langle N\rangle\}$ equivalent to the family of vectors $\{u_n, n \in \langle N\rangle\}$, that is, generating the same v.s. $E$. The principle of the algorithm consists of, at every step $n$, subtracting from the vector $u_n$ of the original free family its orthogonal projection onto the subspace generated by the vectors previously determined, or more precisely, Vect$(e_1, \cdots, e_{n-1})$, and then in normalizing the resulting difference $v_n$ to obtain the vector $e_n$, that is:

$$v_n = u_n - \sum_{k=1}^{n-1} \langle u_n, e_k\rangle e_k \qquad [3.58]$$

$$e_n = \frac{v_n}{\|v_n\|}. \qquad [3.59]$$

This yields the algorithm described in Table 3.6.

| **Gram–Schmidt method** | **Modified Gram–Schmidt method** |
|---|---|
| **Initialization** | |
| $v_1 = u_1\ ,\ e_1 = \frac{v_1}{\|v_1\|}$ | $v_n^{(1)} = u_n\ ,\ n \in \langle N\rangle$ |
| **Computation loops** | |
| $n = 2, \cdots, N$ | $n = 1, \cdots, N$ |
| | $e_n = \frac{v_n^{(n)}}{\|v_n^{(n)}\|}$ |
| $k = 1, \cdots, n-1$ | $k = n+1, \cdots, N$ |
| $\text{proj}_{e_k} u_n = \langle u_n, e_k\rangle e_k$ | $\text{proj}_{e_n} v_k^{(n)} = \langle v_k^{(n)}, e_n\rangle e_n$ |
| $v_n = u_n - \sum_{k=1}^{n-1} \text{proj}_{e_k} u_n$ | $v_k^{(n+1)} = v_k^{(n)} - \text{proj}_{e_n} v_k^{(n)}$ |
| $e_n = \frac{v_n}{\|v_n\|}$ | |

**Table 3.6.** *GS orthonormalization methods*

From a geometric point of view, this algorithm is illustrated, in the case of the space $\mathbb{R}^3$, as shown in Figure 3.1, for the computation of $\mathbf{e}_3 = \mathbf{v}_3/\|\mathbf{v}_3\|$, with $\mathbf{v}_3 = \mathbf{u}_3 - \sum_{k=1}^{2} \text{proj}_{\mathbf{e}_k} \mathbf{u}_3$.

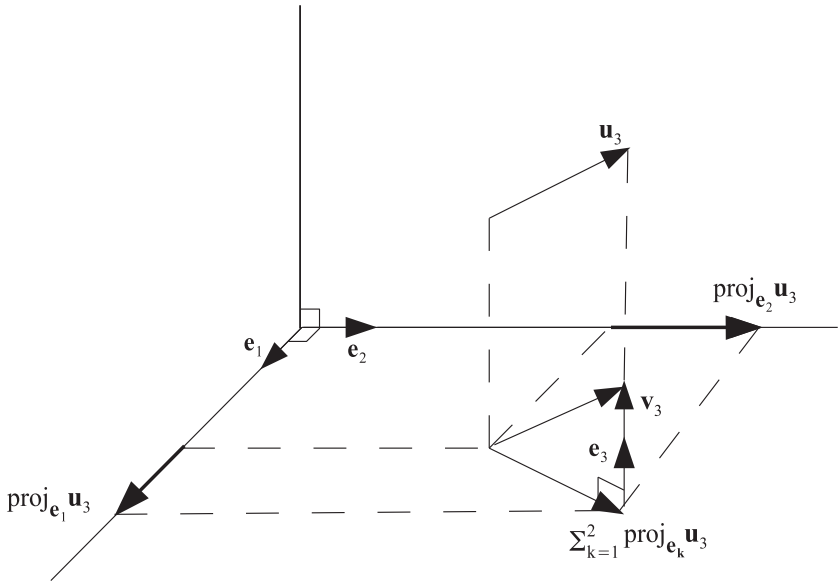



**Figure 3.1.** *Illustration of the GS method for* $\mathbb{R}^3$

### 3.6.5. *QR decomposition*

Let us assume that $\mathbf{u}_n \in \mathbb{R}^M$, for $n \in \langle N \rangle$, with $M \geq N$, and define the matrices:

$$\mathbf{U} = \left[ \mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3 \cdots \mathbf{u}_N \right] \text{ and } \mathbf{Q} = \left[ \mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3 \cdots \mathbf{e}_N \right]$$

whose columns are the vectors $\mathbf{u}_n$ and $\mathbf{e}_n$, respectively. From [3.58] and [3.59], the following equation can be deduced:

$$\mathbf{u}_n = \|\mathbf{v}_n\|\mathbf{e}_n + \sum_{k=1}^{n-1} \langle \mathbf{u}_n, \mathbf{e}_k \rangle \mathbf{e}_k, \ \mathbf{u}_1 = \|\mathbf{v}_1\|\mathbf{e}_1. \qquad [3.60]$$

The GS method can be written in the following matrix form:

$$\left[ \mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3 \cdots \mathbf{u}_N \right] =$$

$$\left[ \mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3 \cdots \mathbf{e}_N \right] \begin{bmatrix} \|\mathbf{v}_1\| & \langle \mathbf{u}_2, \mathbf{e}_1 \rangle & \langle \mathbf{u}_3, \mathbf{e}_1 \rangle & \cdots & \langle \mathbf{u}_N, \mathbf{e}_1 \rangle \\ 0 & \|\mathbf{v}_2\| & \langle \mathbf{u}_3, \mathbf{e}_2 \rangle & \cdots & \langle \mathbf{u}_N, \mathbf{e}_2 \rangle \\ 0 & 0 & \|\mathbf{v}_3\| & \cdots & \langle \mathbf{u}_N, \mathbf{e}_3 \rangle \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \|\mathbf{v}_N\| \end{bmatrix}$$

or still $\mathbf{U} = \mathbf{QR}$ with $\mathbf{Q} \in \mathbb{R}^{M\times N}$ column orthonormal ($\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_N$), that is, a matrix whose columns form an orthonormal basis for the column space of $\mathbf{U}$, and $\mathbf{R} \in \mathbb{R}^{N\times N}$ is an upper triangular matrix with positive diagonal elements. This factorization $\mathbf{U} = \mathbf{QR}$ is called the $\mathbf{QR}$ decomposition of $\mathbf{U}$. When $\mathbf{U}$ is a square matrix ($M = N$), this decomposition is unique, and $\mathbf{Q}$ is then orthogonal ($\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}_N$).

$\mathbf{QR}$ decomposition plays an important role in matrix computation and numerical analysis and will be covered in more detail in Volume 2. More specifically, it can be used to calculate the least-squares solution of a linear system of equations.

A variant of the GS method, called the modified Gram–Schmidt method (MGS), makes it possible to improve numerical stability. It is obtained by replacing in the formula for the calculation of $v_n$ the vector $u_n$ by its projection onto the subspace orthogonal to the subspace Vect$(e_1, \cdots, e_{n-1})$. This gives the algorithm described in Table 3.6. Thus, at each step $n$, the GS method requires the projections $\text{proj}_{e_k} u_n$ to be computed for $k \in \langle n-1\rangle$, while the MGS method replaces each of these projections by a sequence of projections $\text{proj}_{\perp e_{n-1}} \cdots \text{proj}_{\perp e_2} \text{proj}_{\perp e_1} u_n$, where $\text{proj}_{\perp e_k}$ is the orthogonal projection onto the subspace orthogonal to $e_k$.

### 3.6.6. *Application to the orthonormalization of a set of functions*

Consider the inner product of two functions $f$ and $g \in \mathcal{C}^0([a,b], \mathbb{R})$ as defined in Table 3.1:

$$\langle f, g\rangle = \int_a^b f(t)g(t)dt. \qquad [3.61]$$

The Gram–Schmidt algorithm is a method used to build an orthonormal set of functions $\{f_n\}$, in the interval $[a,b]$, from a set of linearly independent functions $\{g_n, n \in [0, N-1]\}$. This orthonormalization algorithm is employed, in particular, to build bases of orthogonal polynomials in the interval $[a,b]$, from the set of linearly independent polynomials $\{1, t, t^2, t^3, \cdots, t^n, \cdots\}$, using a weighted inner product:

$$\langle p_n, p_k\rangle = \int_a^b w(t)p_n(t)p_k(t)dt \qquad [3.62]$$

where $p_n$ is a polynomial of degree $n$, with $w(t) > 0$. This amounts to orthonormalizing the set of polynomials $\{\sqrt{w(t)}, t\sqrt{w(t)}, t^2\sqrt{w(t)}, \cdots, t^n\sqrt{w(t)}, \cdots\}$.

As we will see in section 3.9, different systems of orthogonal polynomials can be obtained depending on the choice of the interval $[a,b]$ and of the weighting $w(t)$. The interval $[a,b]$ may be infinite at one or both ends. In this case, we should have

$\lim_{t\to\pm\infty} w(t) = 0$ as necessary (but not sufficient) condition of existence of the integral $\int_{-\infty}^{\infty}$ (*or* $\int_0^{\infty}$) $w(t)\, p_n^2(t)\, dt = \|p_n\|^2$. See Table 3.17.

Equation [3.58] of the GS algorithm without normalization of polynomials can be written as:

$$p_n(t) = t^n - \sum_{k=0}^{n-1} \frac{\langle p_k, t^n\rangle}{\langle p_k, p_k\rangle} p_k(t) \ , \ \ p_0(t) = 1.$$

The orthogonal polynomials thus determined are unique up to a multiplicative constant that can be chosen to have unitary ($\|p_n\| = 1$ ) or monic orthogonal polynomials, such that the coefficient of the term of degree $n$ of $p_n$ is equal to 1.

EXAMPLE 3.23.– Consider the inner product $\langle f, g\rangle = \frac{1}{2}\int_{-1}^{+1} f(t)\, g(t)\, dt$ used for constructing Legendre polynomials (see Table 3.17). Application of the GS algorithm, with $p_0(t) = 1$ and $\|p_0\| = 1$, gives:

$$p_1(t) = t - \langle p_0, t\rangle\, p_0(t) = t - \frac{1}{2}\int_{-1}^{+1} t\, dt = t$$

$$\|p_1\|^2 = \frac{1}{2}\int_{-1}^{+1} t^2\, dt = \frac{1}{3}$$

$$p_2(t) = t^2 - \langle p_0, t^2\rangle\, p_0(t) - 3\langle p_1, t^2\rangle\, p_1(t) = t^2 - \frac{1}{2}\int_{-1}^{+1} t^2\, dt - \frac{3t}{2}\int_{-1}^{+1} t^3\, dt$$

$$= t^2 - \frac{1}{3}$$

$$\|p_2\|^2 = \frac{1}{2}\int_{-1}^{+1} (t^2 - \frac{1}{3})^2\, dt = \frac{1}{2}\int_{-1}^{+1} (t^4 - \frac{2}{3}t^2 + \frac{1}{9})dt = \frac{4}{45}$$

$$p_3(t) = t^3 - \langle p_0, t^3\rangle\, p_0(t) - 3\langle p_1, t^3\rangle\, p_1(t) - \frac{45}{4}\langle p_2, t^3\rangle\, p_2(t)$$

$$= t^3 - \frac{1}{2}\int_{-1}^{+1} t^3\, dt - \frac{3t}{2}\int_{-1}^{+1} t^4\, dt - \frac{45}{8}(t^2 - \frac{1}{3})\int_{-1}^{+1} (t^2 - \frac{1}{3})t^3\, dt$$

$$= t^3 - \frac{3t}{5}.$$

Bases of orthogonal polynomials $(p_n)_{n\in\mathbb{N}}$ consist of a complete set of polynomials, the notion of completeness meaning that the polynomial basis must contain enough polynomials to represent a function $f$ in the form of a polynomial series:

$$S^f(t) = \sum_{n\in\mathbb{N}} c_n p_n(t) \qquad [3.63a]$$

$$c_n = \frac{\langle f, p_n\rangle}{\|p_n\|^2} = \frac{\int_a^b w(t) f(t) p_n(t)\, dt}{\int_a^b w(t) p_n^2(t)\, dt}. \qquad [3.63b]$$

This gives the Bessel inequality:

$$\|S^f(t)\|^2 = \sum_{n\in\mathbb{N}} |c_n|^2 \|p_n\|^2 \leq \|f\|^2 = \int_a^b w(t) f^2(t)\, dt \qquad [3.64]$$

where equality and inequality signs are valid in the case of a complete system of orthogonal polynomials and an incomplete system, respectively.

## 3.7. Banach and Hilbert spaces

Banach and Hilbert spaces constitute the study framework of numerical sequences $(x_n)$, with $n \in \mathbb{Z}$ or $n \in \mathbb{N}$, and functions $f(t)$, with $t \in \mathbb{R}$ or $t \in \mathbb{R}^+$, as found in signal processing for the analysis of sampled and analog signals, that is of discrete- and continuous-time signals, respectively. Due to the fact that these signals can have an infinite support, a generalization of norms and inner products defined in sections 3.3 and 3.4, to the case of infinite variation intervals for $n$ and $t$, requires a convergence property to be satisfied to ensure the limited character of the quantities involved.

Two signal spaces play an important role in signal processing. These are $l^2$ space of square-summable series and $L^2$ space of square-integrable functions, which correspond to spaces of finite energy signals. For these spaces, one can generalize the concepts of norm, inner product, orthonormal basis, and orthogonal projection in the sense of an approximation.

In the following, first, we define the notion of complete metric space, before defining Banach and Hilbert spaces and the notion of Hilbert basis. Various Hilbert bases will be presented as trigonometric bases and bases of orthogonal polynomials. The Hilbertian approach generalizes the Euclidean approach considered in previous sections to infinite-dimensional v.s. This approach plays an important role in the theory of function approximation using expansions over Hilbert bases. We will end this chapter by presenting examples of Fourier series expansions and expansions over bases of orthogonal polynomials.

### 3.7.1. *Complete metric spaces*

After defining the notion of Cauchy sequence, we provide the definition of a complete metric space and summarize some of its properties. The notion of complete space is fundamental for the construction of Hilbert bases in a Hilbert space and for obtaining methods of approximation through series expansions over these bases.

### 3.7.1.1. *Definition of a Cauchy sequence*

A unilateral sequence is a function $f : \mathbb{N}^* \to \mathbb{C},\ \mathbb{N}^* \ni n \mapsto x_n \in \mathbb{C}$, written as $(x_n)_{n\in\mathbb{N}^*}$, or $(x_n)_{n\geq 1}$, or simply $(x_n)$. In the case of a bilateral sequence, the domain of $f$ is $\mathbb{Z}$ instead of $\mathbb{N}^*$.

In the following, we define the important concepts of convergent sequence and of Cauchy sequence.

– It is said that a sequence $(x_n)_{n\geq 1}$ in a n.v.s. $(E, \|.\|)$ converges (in norm) to the limit $x$ if the sequence $\|x_n - x\|$ tends to 0 when $n$ tends to infinity:

$$\lim_{n\to\infty} \|x_n - x\| = 0, \qquad [3.65]$$

and it is written that $\lim\limits_{n\to\infty} x_n = x$. Equivalently, $(x_n)$ converges to $x$ if for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $\lim\|x_n - x\| < \epsilon$ for all $n \geq N$. If not, $(x_n)$ diverges.

– In a metric space $(E, d)$, it is said that $(x_n)_{n\geq 1}$ is a Cauchy sequence if:

$$\forall \epsilon > 0,\ \exists N \in \mathbb{N} \ :\ d(x_n, x_p) < \epsilon\ ,\ \forall n, p \geq N.$$

The notion of Cauchy sequence is more general than the notion of convergent sequence, in the sense that it expresses the arbitrarily large proximity of two arbitrary terms of the sequence, from a given rank $N$. Nonetheless, this is not synonymous with a convergent sequence, whereas every convergent sequence is a Cauchy sequence.

Indeed, if $(x_n)_{n\geq 1}$ converges to $x$, considering the distance $d$ induced by the norm, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \epsilon/2, \forall n \geq N$, and the use of the triangle inequality gives:

$$d(x_n, x_p) \leq d(x_n, x) + d(x, x_p) < \epsilon\ ,\ \forall n, p \geq N,$$

from which it can be inferred that $(x_n)$ is a Cauchy sequence.

### 3.7.1.2. *Series and sequence of partial sums*

A series is the sum of a sequence. In the following, we consider sequences in a n.v.s. $E$. So, given a sequence $(x_n)_{n\geq 1}$ in $E$, the associated series is defined as $\sum_{n=1}^{\infty} x_n$. A sequence is in turn associated with this series, called the sequence of partial sums, such that the $N$th partial sum, denoted by $S_N^x$, is defined as $S_N^x = \sum_{n=1}^{N} x_n$.

If the sequence $(S_N^x)$ of partial sums converges to $S$ when $N \to \infty$, then the series $\sum x_n$ converges to $S$. Otherwise, the series diverges. This notion of partial sum will be used in section 3.8.4 for the study of Fourier series.

#### 3.7.1.3. *Definition of complete metric space and examples*

A metric space $E$ is said to be complete if any Cauchy sequence in $E$ converges to a limit in $E$.

Some examples of complete metric spaces are outlined below.

– Any finite-dimensional n.v.s. equipped with the metric [3.22] associated with the norm is a complete metric space.

– Any finite-dimensional pre-Hilbert space is complete. On the other hand, any infinite-dimensional pre-Hilbert space is not complete.

– The Cartesian product of a finite number of complete metric spaces is a complete metric space.

NOTE 3.24.– In the case of an incomplete metric space $(E, d)$, its extension to a complete metric space is called the completion of $E$ with respect to $d$. The completion of a metric space is unique up to isometry.

Any incomplete metric space can be completed by preserving its structure. Thus, a n.v.s. can be completed to a Banach space, and an inner product space to a Hilbert space.

### 3.7.2. *Adherence, density and separability*

We summarize here a few fundamental results related to sequences and more generally to metric spaces (Coste 2016). In the following, $(x_n)$ is a sequence of a metric set $(E, d)$, and $F$ is a subset of $E$.

– A subsequence (or extracted sequence) of a sequence $(x_n)_{n\in\mathbb{N}}$ is a sequence $(x_{n_i})_{i\in\mathbb{N}}$, where $(n_i)_{i\in\mathbb{N}}$ is a strictly increasing sequence of elements of $\mathbb{N}$, that is, a sequence obtained taking only some elements of the initial sequence.

– If $(x_n)$ converges to $x$, then every extracted sequence converges to the same limit.

– The limits of convergent subsequences of $(x_n)$ are called adherence values of $(x_n)$, and the set of sequence limits in $F \subseteq E$ is called the adherence (or closure) of $F$, and denoted by $\overline{F}$:

$$x \in \overline{F} \Leftrightarrow \exists (x_n) \in F \text{ such that } \lim_{n\to\infty} x_n = x.$$

– If a Cauchy sequence has an adherence value, then it converges.

– A subspace $F$ of a n.v.s. $E$ is said to be closed if and only if any convergent sequence of $F$ converges to a limit in $F$, namely, if $(x_n) \in F$ and $\lim_{n\to\infty} x_n = x$, then $x \in F$.

For a finite-dimensional n.v.s. $E$, any subspace $F$ is closed. This is not the case in general, if $E$ is infinite-dimensional.

– A subspace $F$ of a complete metric space $E$ is itself complete if and only if it is closed.

– A subset $F$ of a metric space $E$ is said to be dense in $E$ if and only if the adherence $\overline{F}$ is $E$. From a topological point of view, the concept of density makes it possible to formalize the fact that for any point $x \in E$, there exists a point of $F$ as close to $x$ as desired.

– A metric space $(E, d)$ is said to be separable if it contains a countable dense subset $F \subset E$, that is, for which $\overline{F} = E$.

## 3.7.3. *Banach and Hilbert spaces*

### 3.7.3.1. *Definitions*

A Banach space is a complete n.v.s., or in other words, a complete metric space for the distance induced by the norm.

A Hilbert space is a (real or complex) pre-Hilbert space (or inner product space) in which the norm associated with the inner product makes it a complete space and therefore a Banach space. Complete means that any sequence of functions in a Hilbert space converges to a limit belonging to the space.

### 3.7.3.2. *Properties*

– Any finite-dimensional n.v.s. is a Banach space.

– Any finite-dimensional pre-Hilbert space is a Hilbert space.

– It is important to point out that Hilbert spaces generalize Banach spaces in the sense that they are equipped with an inner product, which is not the case of Banach spaces. This is because it is not always possible to associate an inner product with a norm, while a norm can always be induced from an inner product. Subsequently, any Hilbert space is a Banach space, but the converse is not necessarily true.

– A Banach space is a Hilbert space if and only if its norm satisfies the parallelogram equality [3.35]. There exists a single inner product built using this norm and given by [3.41].

In Figure 3.2, we summarize the links between metric space, n.v.s., and pre-Hilbert space, on the one hand, and Banach and Hilbert spaces, on the other hand.

### 3.7.3.3. *Examples of Hilbert spaces*

As we saw in section 3.4.3, $l^p$ and $L^p$ spaces do not admit any inner product associated with $l_p$ and $L_p$ Hölder norms, for $p \neq 2$. This means that only $l^2$ and $L^2$

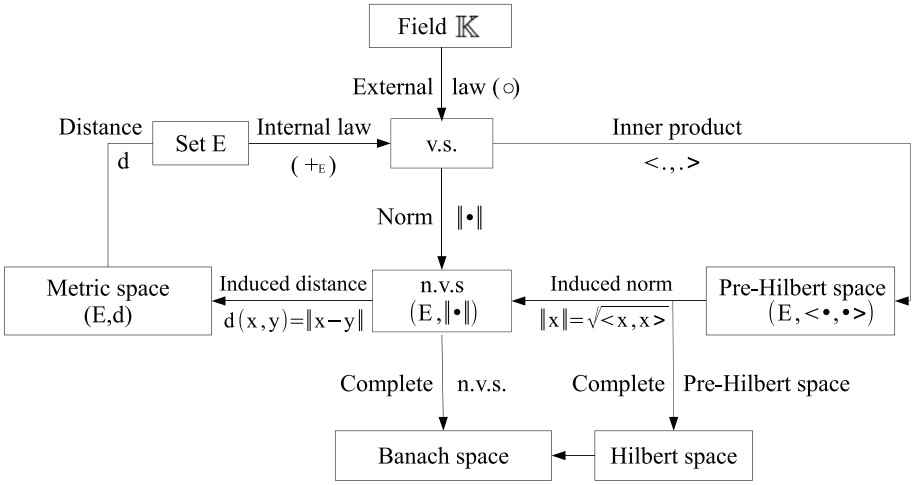



**Figure 3.2.** *Banach and Hilbert spaces*

spaces corresponding to $p = 2$ are Hilbert spaces. We describe hereafter three infinite-dimensional Hilbert spaces, corresponding to the sets $l^2(\mathbb{Z}, \mathbb{K})$ of square-summable sequences, $L^2(\mathbb{R}, \mathbb{K})$ of square-integrable functions, and $L^2([0, T], \mathbb{C})$ of $T$-periodic square-integrable functions.

– $l^2(\mathbb{Z}, \mathbb{K})$: space of real (or complex) sequences, for which the square of the absolute value (or of the modulus) is summable:

$$l^2(\mathbb{Z}, \mathbb{K}) = \left\{ x : \mathbb{Z} \ni n \mapsto x_n \in \mathbb{K}\,,\ \sum_{n\in\mathbb{Z}} |x_n|^2 < \infty \right\} \qquad [3.66a]$$

$$\langle x, y \rangle = \sum_{n\in\mathbb{Z}} x_n y_n \quad (\text{or } \sum_{n\in\mathbb{Z}} x_n \overline{y}_n). \qquad [3.66b]$$

Similarly, one defines subspaces $l^2((-\infty, 0), \mathbb{K})$ and $l^2((0, \infty), \mathbb{K})$.

– $L^2(\mathbb{R}, \mathbb{K})$: space of real- (complex-) valued functions, for which the square of the absolute value (of the modulus) is integrable:

$$L^2(\mathbb{R}, \mathbb{K}) = \left\{ f : \mathbb{R} \ni t \mapsto f(t) \in \mathbb{K}\,,\ \int_{\mathbb{R}} |f(t)|^2 dt < \infty \right\} \qquad [3.67a]$$

$$\langle f, g \rangle = \int_{\mathbb{R}} f(t) g(t) dt \quad (\text{or } \int_{\mathbb{R}} f(t) \overline{g}(t) dt). \qquad [3.67b]$$

In signal processing, the quantities $\|x\|_2^2 = \sum_{n\in\mathbb{Z}} |x_n|^2$ and $\|f\|_2^2 = \int_{\mathbb{R}} |f(t)|^2 dt$ represent the energy of the digital signal $x_n$ and of the analog signal $f(t)$, respectively, and the $l^2$ and $L^2$ spaces are those of finite energy signals. In the case of causal signals, that is, equal to zero for $n < 0$ and $t < 0$, the sets $\mathbb{Z}$ and $\mathbb{R}$ in definitions [3.66a] and [3.67a] are to be replaced by $\mathbb{N}$ and $\mathbb{R}^+$, respectively.

– $L^2([0,T],\mathbb{C})$: space of $T$-periodic functions, defined and square integrable on $[0,T]$, that is:

$$L^2([0,T],\mathbb{C}) = \left\{ f : [0,T] \ni t \mapsto f(t) \in \mathbb{C}, \int_0^T |f(t)|^2 dt < \infty \right\}, \quad [3.68a]$$

$$\langle f, g \rangle = \frac{1}{T}\int_0^T f(t)\overline{g}(t)dt = \frac{1}{T}\int_{-T/2}^{T/2} f(t)\overline{g}(t)dt. \quad [3.68b]$$

This space is the one considered for $T$-periodic functions on $\mathbb{R}$, restricted to an interval[15] of length $T$ often chosen as $[-T/2, T/2]$ or $[0,T]$.

In the context of signal processing, the quantity $\|f\|_2^2 = \frac{1}{T}\int_0^T |f(t)|^2\, dt$ equal to the square of the norm associated with the inner product [3.68b] represents the power of the analog signal $f(t)$ dissipated over one period.

The inner product can also be defined as:

$$\langle f, g \rangle = \int_0^T f(t)\overline{g}(t)dt = \int_{-T/2}^{T/2} f(t)\overline{g}(t)dt. \quad [3.69]$$

The quantity $\|f\|_2^2 = \int_0^T |f(t)|^2\, dt$ then represents the energy of the signal $f(t)$ dissipated over one period. It is equal to the square of the norm associated with the inner product [3.69].

## 3.7.4. *Hilbert bases*

### 3.7.4.1. *Definitions*

– A family $\{b_n\}_{n\in\mathbb{N}^*}$ of vectors of a Hilbert space $E$ is said to be complete (or total) if the space generated by the vectors $b_n$, $n \in \mathbb{N}^*$, that is, $\text{Vect}(b_n, n \in \mathbb{N}^*)$, is equal to the space $E$. This means that $\text{Vect}(b_n, n \in \mathbb{N}^*)$ is dense in $E$. Any vector $x \in E$ can then be written as the sum of a series $\sum_{n\in\mathbb{N}^*} x_n b_n$, namely:

$$\lim_{N\to\infty} \|x - \sum_{n=1}^{N} x_n b_n\| = 0, \quad [3.70]$$

where $(x_n)_{n\in\mathbb{N}^*}$ represent the coordinates of $x$ in the basis $\{b_n\}_{n\in\mathbb{N}^*}$.

15 As it has been demonstrated in section 2.5.9.2, the integrals of a $T$-periodic function can be calculated on any interval of length $T$.

– As in the case of finite-dimensional v.s., a family of vectors $\{e_n\}_{n\in I}$ of a Hilbert space is called orthonormal, if $\langle e_n, e_p\rangle = \delta_{np}$ for all $n, p \in I$.

– A complete orthonormal family is called a Hilbert basis. The orthonormality property can be demonstrated in a similar way to the case of a finite-dimensional v.s. On the other hand, to demonstrate that a family of vectors is complete often requires to perform a more difficult analysis. See Young (1988) for such an analysis which will not be considered here.

In the following, we consider the case of infinite countable Hilbert bases, in other words, indexed by $I = \mathbb{N}$ or $\mathbb{Z}$.

– Any separable Hilbert space has a countable Hilbert basis.

The choice of some bases such as trigonometric bases allow certain characteristics of signals to be extracted, such as their frequency content.

### 3.7.4.2. *Examples of Hilbert bases*

– In $l^2(\mathbb{N}^*, \mathbb{C})$, with the inner product $\langle x, y\rangle = \sum_{n\in\mathbb{N}^*} x_n\overline{y}_n$, a canonical Hilbert basis is given by:

$$e_n = \{\underbrace{0, \cdots, 0}_{\text{(n-1) terms}}, 1, 0, \cdots\}\ ,\ n \in \mathbb{N}^*.$$

– In $L^2([-\pi, \pi], \mathbb{C})$, the space of $2\pi$-periodic functions, with:

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\overline{g}(t)dt, \qquad [3.71]$$

complex exponential functions $e_n, n \in \mathbb{Z}$, defined as:

$$e_n(t) = e^{int}\ ,\ t \in [-\pi, \pi]\ ,\ i^2 = -1, \qquad [3.72]$$

form a complete orthonormal family, called Fourier Hilbert basis.

The orthonormality of the basis functions $e_n$ and $e_p$, for $n, p \in \mathbb{Z}$, can be verified based on the definition of the inner product:

$$\langle e_n, e_p\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(n-p)t}dt = \begin{cases} 1 & \text{if } n = p \\ \left[\frac{e^{i(n-p)t}}{2\pi i(n-p)}\right]_{t=-\pi}^{t=\pi} = 0 & \text{if } n \neq p \end{cases} \qquad [3.73]$$

The equality to zero is a result of the fact that $n - p \in \mathbb{Z}$. We thus have $\langle e_n, e_p\rangle = \delta_{np}$. Observing that $e_n(t + 2\pi) = e^{in(t+2\pi)} = e^{int} = e_n(t)$, it can be concluded that this basis of functions can be used to represent $2\pi$-periodic functions.

In the case of $T$-periodic functions, with the inner product [3.68b], the basis functions [3.72] become:

$$e_n(t) = e^{\frac{in2\pi t}{T}} = e^{in\omega t}, \qquad [3.74]$$

where $\omega = \frac{2\pi}{T}$ represents the angular frequency associated with $T$. We then have $e_n(t+T) = e_n(t)$.

– Similarly, the family of trigonometric functions $(\phi_n), n \in \mathbb{N}$ defined as:

$$\phi_0 = 1 \qquad [3.75a]$$

$$\phi_{2k}(t) = \sqrt{2}\cos kt \ , \ k \in \mathbb{N}^* \qquad [3.75b]$$

$$\phi_{2k-1}(t) = \sqrt{2}\sin kt \ , \ k \in \mathbb{N}^* \qquad [3.75c]$$

is a Hilbert basis, also called trigonometric basis, in $L^2([-\pi, \pi], \mathbb{C})$.

The orthonormality of the aforementioned functions can be demonstrated using trigonometric formulae. For instance:

$$\|\phi_{2k}(t)\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} 2\cos^2 kt\, dt = \frac{1}{2\pi}\int_{-\pi}^{\pi}(1+\cos 2kt)\, dt = 1$$

$$\|\phi_{2k-1}(t)\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} 2\sin^2 kt\, dt = \frac{1}{\pi}\int_{-\pi}^{\pi}(1-\cos^2 kt)\, dt = 1.$$

Similarly, we have:

$$\begin{aligned}\langle \phi_{2k}(t), \phi_{2p-1}(t)\rangle &= \frac{1}{2\pi}\int_{-\pi}^{\pi} 2\cos kt \sin pt\, dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}[\sin(p+k)t + \sin(p-k)t]\, dt \\ &= -\frac{1}{2\pi}\left[\frac{\cos(p+k)t}{p+k}\right]_{-\pi}^{\pi} - \frac{1}{2\pi}\left[\frac{\cos(p-k)t}{p-k}\right]_{-\pi}^{\pi} = 0.\end{aligned}$$

This equality to zero can also be directly deduced due to the fact that $\cos kt \sin pt$ is an odd function, and for any odd function $f$, we have $\int_{-\pi}^{\pi} f(t)\, dt = 0$.

We can make the following observations:

– If the inner product is defined as:

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(t)\overline{g}(t)dt, \qquad [3.78]$$

then the basis function [3.72] becomes:

$$e_n(t) = \frac{1}{\sqrt{2\pi}}e^{int}. \qquad [3.79]$$

Similarly, functions $(\phi_n)$ must be divided by $\frac{1}{\sqrt{2\pi}}$, and definitions [3.75b] and [3.75c] become:

$$\phi_{2k}(t) = \frac{1}{\sqrt{\pi}} \cos kt \ , \ \phi_{2k-1}(t) = \frac{1}{\sqrt{\pi}} \sin kt \ , \ k \in \mathbb{N}^*. \qquad [3.80]$$

Indeed, we then have, for example:

$$\|\phi_{2k}(t)\|^2 = \int_{-\pi}^{\pi} \frac{1}{\pi} \cos^2 kt \, dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 + \cos 2kt) \, dt = 1.$$

– In $L^2([a,b],\mathbb{C})$, with the inner product $\langle f,g\rangle = \int_a^b f(t)\overline{g}(t)dt$, the Hilbert basis [3.79] becomes:

$$e_n(t) = \frac{1}{\sqrt{b-a}} e^{in2\pi\frac{t-a}{b-a}} \ , \ t \in [a,b] \ , \ n \in \mathbb{N}. \qquad [3.81]$$

– For the space $L^2(\mathbb{R},\mathbb{K})$, it is possible to divide the real axis in intervals of length $2\pi$ such that $I_m = [(2m-1)\pi,(2m+1)\pi]$. The family of vectors defined by Allen and Mills (2004):

$$e_{m,n}(t) = \chi_m e^{int} \ , \ m,n \in \mathbb{N} \qquad [3.82]$$

is a Hilbert basis in $L^2(\mathbb{R},\mathbb{K})$, with $\chi_m$ being the characteristic function of the subset $I_m$ (see the definition in section 2.3.6).

In Table 3.7, we summarize the complex exponential and trigonometric Hilbert bases previously presented for the space $L^2([-\pi,\pi],\mathbb{C})$, equivalent to $L^2([0,2\pi],\mathbb{C})$, as well as those inferred from [3.81] with $(a=0, b=T)$, for the space $L^2([0,T],\mathbb{C})$ of $T$-periodic functions, specifying the inner product associated with each one of these bases.

| **Spaces** | $\langle .,. \rangle$ | $e_n(t)$ | cos() | sin() |
|---|---|---|---|---|
| $L^2([0,2\pi],\mathbb{C})$ | $\langle f,g\rangle = \int_0^{2\pi} f(t)\overline{g}(t)dt$ | $\frac{1}{\sqrt{2\pi}} e^{int}$ | $\frac{1}{\sqrt{\pi}} \cos\, nt$ | $\frac{1}{\sqrt{\pi}} \sin\, nt$ |
| $L^2([0,2\pi],\mathbb{C})$ | $\langle f,g\rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t)\overline{g}(t)dt$ | $e^{int}$ | $\sqrt{2} \cos\, nt$ | $\sqrt{2} \sin\, nt$ |
| $L^2([0,T],\mathbb{C})$ | $\langle f,g\rangle = \int_0^{T} f(t)\overline{g}(t)dt$ | $\frac{1}{\sqrt{T}} e^{in\omega t}$ | $\sqrt{\frac{2}{T}} \cos\, n\omega t$ | $\sqrt{\frac{2}{T}} \sin\, n\omega t$ |
| $L^2([0,T],\mathbb{C})$ | $\langle f,g\rangle = \frac{1}{T} \int_0^{T} f(t)\overline{g}(t)dt$ | $e^{in\omega t}$ | $\sqrt{2} \cos\, n\omega t$ | $\sqrt{2} \sin\, n\omega t$ |

**Table 3.7.** *Examples of Hilbert bases* $(\omega = \frac{2\pi}{T})$

## 3.8. Fourier series expansions

### 3.8.1. *Fourier series, Parseval's equality and Bessel's inequality*

Let $f$ denote a function of a Hilbert space $(E, \langle ., . \rangle)$ and $F$ a complete subspace of $E$ admitting as Hilbert basis $\{e_n, n \in \mathbb{Z}\}$. The expansion of $f \in E$ over this basis is written as:

$$p_F(f) = \sum_{n \in \mathbb{Z}} c_n(f) e_n \ , \ c_n(f) = \langle f, e_n \rangle.$$

This expansion generalizes equation [3.51a] to the case of infinite-dimensional Hilbert spaces. It corresponds to the orthogonal projection of $f$ onto $F$, such that $f - p_F(f) \in F^{\perp}$. Parseval's equality [3.54] then becomes:

$$\sum_{n \in \mathbb{Z}} |c_n(f)|^2 = \|f\|^2, \qquad [3.83]$$

and for any finite subset $I \subset \mathbb{Z}$, the Bessel inequality can be written as:

$$\sum_{n \in I} |c_n(f)|^2 \leq \|f\|^2. \qquad [3.84]$$

### 3.8.2. *Case of $2\pi$-periodic functions from $\mathbb{R}$ to $\mathbb{C}$*

As we have seen in section 2.5.9, the shifting from a $2\pi$-periodic function $f$ to a $T$-periodic function $g$ can be carried out through the transformation [2.5], which explains why Fourier[16] series expansions are often considered for $2\pi$-periodic functions.

Let $f$ be a function of the v.s. $\mathcal{C}^0_{2\pi}(\mathbb{R}, \mathbb{C})$, namely, continuous and of period $2\pi$, from $\mathbb{R}$ to $\mathbb{C}$. Its Fourier series expansion over the orthonormal basis $\mathcal{B} = \{e_n : t \mapsto e^{int}, n \in \mathbb{Z}\}$, with the following inner product:

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t) \overline{g}(t) dt, \qquad [3.85]$$

16 Joseph Fourier (1768–1830), a French mathematician and physicist who was a student of Lagrange, Laplace, and Monge, at the École Normale Supérieure. He participated in the French Revolution. Appointed prefect of Isère by Napoleon in 1802, he created the Faculty of Grenoble in 1810. He studied the propagation of heat, and he is famous for the series and the transform that bear his name. His series that he introduced in 1822 were controversial at the time of their publication. The study of their convergence properties gave rise to many works (closely related to the theory of integration), which include those of Dirichlet, who was the first to provide conditions for the convergence of trigonometric series in 1828; these conditions were then generalized by Jordan in 1881.

is written as:

$$S^f(t) = \sum_{n\in\mathbb{Z}} c_n(f)e^{int} \tag{3.86a}$$

$$c_n(f) = \langle f, e_n\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{-int}dt, \tag{3.86b}$$

where $c_n(f)$ is called the exponential Fourier coefficient of rank $n$ of function $f$. To alleviate writing, $c_n$ will be used in the following. Taking Euler's formula $e^{int} = \cos nt + i\sin nt$ into account, this Fourier series expansion can be rewritten as:

$$\begin{aligned} S^f(t) &= c_0 + \sum_{n\in\mathbb{N}^*} (c_n e^{int} + c_{-n}e^{-int}) \\ &= c_0 + \sum_{n\in\mathbb{N}^*} (c_n + c_{-n})\cos nt + i(c_n - c_{-n})\sin nt \\ &= a_0 + \sum_{n\in\mathbb{N}^*} (a_n \cos nt + b_n \sin nt), \end{aligned} \tag{3.87}$$

with:

$$a_0 = c_0 = \frac{1}{2\pi}\int_0^{2\pi} f(t)\,dt \tag{3.88a}$$

$$a_n = c_n + c_{-n} = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos nt\,dt = \sqrt{2}\langle f, \varphi_{2n}\rangle, \quad n \in \mathbb{N}^* \tag{3.88b}$$

$$b_n = i(c_n - c_{-n}) = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin nt\,dt = \sqrt{2}\langle f, \varphi_{2n-1}\rangle, \ n \in \mathbb{N}^*, \tag{3.88c}$$

the basis functions $\varphi_{2n}$ and $\varphi_{2n-1}$ being defined in [3.75b] and [3.75c].

The coefficients $a_n$ and $b_n$ are called trigonometric Fourier coefficients of $f$. We also have:

$$c_n = \frac{1}{2}(a_n - ib_n)\ ,\ c_{-n} = \frac{1}{2}(a_n + ib_n). \tag{3.89}$$

If $f$ is of class $\mathcal{C}^k$ over $\mathbb{R}$, that is, $k$ times continuously differentiable (see definition in section 2.5.9.2), then the complex Fourier coefficients of its $k$th derivative $f^{(k)}$ are given by $c_n(f^{(k)}) = (in)^k c_n(f)$, $\forall n \in \mathbb{Z}$. This can be obtained by deriving $k$ times the Fourier series of $f$, defined in [3.86a]. From this relation, it can be deduced that:

$$|c_n(f)| = \frac{1}{n^k}|c_n(f^{(k)})| = \frac{1}{2\pi n^k}\left|\int_0^{2\pi} f^{(k)}(t)e^{-int}\,dt\right| \le \frac{1}{2\pi n^k}\int_0^{2\pi} \left|f^{(k)}(t)\right| dt,$$

from which it can be concluded that the coefficients $c_n(f)$ decrease when $n$ increases, and tend to zero when $n$ tends to infinity.

On the other hand, the complex Fourier coefficients of the translated function defined as $f_\tau(t) = f(t-\tau)$ are given by $c_n(f_\tau) = e^{-in\tau}c_n(f)$. This result follows directly from definition [3.86b] applied to the function $f_\tau(t)$, with the change of variable $s = t - \tau$, and taking into account the assumption that $f$ is $2\pi$-periodic.

In Table 3.8, we summarize the existing relations between the exponential and trigonometric coefficients of $f$, and between the exponential coefficients of $f$ and those of $f^{(k)}$ on the one hand and of the translated function $f_\tau$ on the other.

It should be noted that the multiplication of $f$ by a constant factor $\alpha$ results in the same multiplication by $\alpha$ of coefficients $c_n, a_n$, and $b_n$.

| $a_0 = c_0\ ;\ a_n = c_n + c_{-n}\ ,\ b_n = i(c_n - c_{-n}),\ \forall n \in \mathbb{N}^*$ |
|---|
| $c_0 = a_0\ ;\ c_n = \frac{1}{2}(a_n - ib_n)\ ,\ c_{-n} = \frac{1}{2}(a_n + ib_n),\ \forall n \in \mathbb{N}^*$ |
| kth order derivative |
| $c_n(f^{(k)}) = (in)^k c_n(f),\ \forall n \in \mathbb{Z}$ |
| Translated function: $f_\tau(t) = f(t-\tau)$ |
| $c_n(f_\tau) = e^{-in\tau}c_n(f),\ \forall n \in \mathbb{Z}$ |

**Table 3.8.** *Relations between Fourier coefficients*

Equation [3.87], called as a trigonometric series of period $2\pi$, corresponds to the expansion of $f$ over the Hilbert basis [3.75a]–[3.75c]. The term $a_0$, which represents the average value of $f$ over a period, is called the harmonic of zero rank. The sum $h_n(t) = a_n \cos nt + b_n \sin nt$ represents the harmonic of rank $n$, of period $\frac{2\pi}{n}$. For $n = 1$, we have the fundamental.

The Fourier series expansions [3.86a] and [3.87] are expressed in complex exponential and sine-cosine forms, respectively. They are also called complex form and real form of the Fourier series.

The expansion [3.87] presents the following advantages:

– The sine-cosine form is a unilateral series ($n \in \mathbb{N}^*$), while the complex exponential form is a bilateral series ($n \in \mathbb{Z}$).

– For a real-valued function $f$, we can deduce from [3.88a] to [3.88c] that the coefficients $a_n$ and $b_n$ are real, whereas from [3.86b], we deduce that the coefficients $c_n$ are complex. As a result, the Fourier series [3.87] is equal to a sum of real $2\pi$-periodic functions. This is not the case of the series [3.86a]. That is why, in general, the real form [3.87] is preferred to the complex form [3.86a] in the case of real-valued functions.

For real $a_n$ and $b_n$, we deduce the following formulae from relations [3.89]:

$$c_{-n} = \overline{c_n} \;\Rightarrow\; a_n = 2\,\mathrm{Re}(c_n) \;\text{ and }\; b_n = -2\,\mathrm{Im}(c_n) \;\; \forall n \in \mathbb{N}^*. \qquad [3.90]$$

We can also define the cosine Fourier series. Using the trigonometric relation $a\cos\varphi + b\sin\varphi = \sqrt{a^2+b^2}\cos\left(\varphi + \arctan(-\frac{b}{a})\right)$, we obtain:

$$S^f(t) = a_0 + \sum_{n=1}^{\infty} A_n \cos(nt + \alpha_n) \text{ with } A_n = \sqrt{a_n^2 + b_n^2} \;\;,\;\; \alpha_n = \arctan(-\frac{b_n}{a_n}).$$

In Table 3.9, we summarize the results relative to the Fourier series expansion of a function $f \in \mathcal{C}^0_{2\pi}(\mathbb{R},\mathbb{C})$.

| Properties | Relations |
|---|---|
| Fourier series in exponential form | $S^f(t) = \sum_{n\in\mathbb{Z}} c_n e^{int}$ <br> $c_n = \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{-int}dt$ |
| Fourier series in sine-cosine form | $S^f(t) = a_0 + \sum_{n\in\mathbb{N}^*}(a_n\cos nt + b_n \sin nt)$ <br> $a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(t)\,dt$ <br> $a_n = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos nt\,dt, \quad n \in \mathbb{N}$ <br> $b_n = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin nt\,dt, \;\; n \in \mathbb{N}^*$ |
| Fourier series in cosine form | $S^f(t) = a_0 + \sum_{n=1}^{\infty}\sqrt{a_n^2+b_n^2}\cos\left(nt + \arctan(-\frac{b_n}{a_n})\right)$ |

**Table 3.9.** *Fourier series expansion of* $f \in \mathcal{C}^0_{2\pi}(\mathbb{R},\mathbb{C})$

The computation of the Fourier coefficients is simplified in the case of an even or odd function[17]. Indeed, for an even $f$, by carrying out a change of variable from $t$ into $-t$ and taking into account the parity property of $f$, we obtain:

$$c_{-n} = \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{int}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{int}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}dt = c_n,$$

while for an odd $f$, we have:

$$c_{-n} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{int}dt = -\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}dt = -c_n.$$

Using definitions [3.88b] and [3.88c] of the coefficients $a_n$ and $b_n$ according to the coefficients $c_n$, it is easy to deduce the simplified expressions for the Fourier coefficients and associated Fourier series, which are summarized in Table 3.10.

| Coefficients | $f$ even | $f$ odd |
|---|---|---|
| $a_0$ | $\frac{1}{\pi}\int_0^{\pi} f(t)\,dt$ | 0 |
| $a_n\ ,\ n \in \mathbb{N}^*$ | $\frac{2}{\pi}\int_0^{\pi} f(t)\cos nt\,dt$ | 0 |
| $b_n\ ,\ n \in \mathbb{N}^*$ | 0 | $\frac{2}{\pi}\int_0^{\pi} f(t)\sin nt\,dt$ |
| | Fourier series | |
| $S^f(t)$ | $a_0 + \sum_{n\in\mathbb{N}^*} a_n \cos nt$ | $\sum_{n\in\mathbb{N}^*} b_n \sin nt$ |

**Table 3.10.** *Fourier series of even and odd $f$ on $[-\pi, \pi]$*

From Table 3.10, we can conclude that the Fourier series of an even function contains only cosine terms, whereas for an odd function, it contains only sine terms.

17 Definitions:

- $f$ is said to be even if $f(-t) = f(t)$ for all $t \in \mathbb{R}$.
- $f$ is said to be odd if $f(-t) = -f(t)$ for all $t \in \mathbb{R}$.

Properties:

- $f, g$ even $\Rightarrow fg$ is even.
- $f, g$ odd $\Rightarrow fg$ is even.
- $f$ even and $g$ odd $\Rightarrow fg$ is odd.
- If $f$ is even: $\int_{-\pi}^{\pi} f(t)dt = 2\int_0^{\pi} f(t)dt$.
- If $f$ is odd: $\int_{-\pi}^{\pi} f(t)dt = 0$.

EXAMPLE 3.25.– Consider the $2\pi$-periodic function defined on $[-\pi, \pi]$ as $f(t) = t$. Since this function is odd, we have $a_n = 0, \forall n \in \mathbb{N}$, and using an integration by parts, we get:

$$b_n = \frac{2}{\pi}\int_0^{\pi} t \sin nt\, dt = \frac{2}{\pi}\Big(\Big[-\frac{t\cos nt}{n}\Big]_0^{\pi} + \frac{1}{n}\int_0^{\pi}\cos nt\, dt\Big) = \frac{2}{n}(-1)^{n+1}$$

which gives:

$$S^f(t) = 2\Big(\sin t - \frac{1}{2}\sin 2t + \cdots + \frac{(-1)^{n+1}}{n}\sin nt + \cdots\Big).$$

Note that this expression of $S^f(t)$ is valid for $-\pi < t < \pi$, but not at the discontinuity points $\pm\pi$ of $f$. The Dirichlet–Jordan theorem allows the convergence of the Fourier series to be studied at these points of discontinuity (see Table 3.12).

FACT 3.26.– In the case of an odd function with a vertical offset, that is, for a function $A + f(t)$ where $A$ is a constant and $f$ is odd, we have $a_0 = A$ and $a_n = 0\,,\ \forall n \in \mathbb{N}^*$.

### 3.8.3. $T$-periodic functions from $\mathbb{R}$ to $\mathbb{C}$

As already pointed out, the shift from a period $2\pi$ to a period $T$ can be achieved by changing the variable of integration $t$ in $\frac{2\pi}{T}t = \omega t$ and therefore $dt$ in $\frac{2\pi}{T}dt = \omega dt$ in the calculation formulae of $c_n$, $a_n$ and $b_n$. In Table 3.11, we give the Fourier coefficients for a function of the v.s. $\mathcal{C}^0_T(\mathbb{R}, \mathbb{C})$ for the two inner products defined in Table 3.7 using the angular frequency $\omega = 2\pi/T$.

| $\langle f, g\rangle = \frac{1}{T}\int_0^T f(t)\overline{g}(t)dt$ |
|---|
| $c_n \ ;\ a_0\ ;\ a_n\ ;\ b_n$ |
| $\frac{1}{T}\int_0^T f(t)e^{-in\omega t}\,dt \ ;\ \frac{1}{T}\int_0^T f(t)\,dt\ ;\ \frac{2}{T}\int_0^T f(t)\cos n\omega t\,dt\ ;\ \frac{2}{T}\int_0^T f(t)\sin n\omega t\,dt$ |
| $\langle f, g\rangle = \int_0^T f(t)\overline{g}(t)dt$ |
| $c_n \ ;\ a_0\ ;\ a_n\ ;\ b_n$ |
| $\frac{1}{\sqrt{T}}\int_0^T f(t)e^{-in\omega t}\,dt \ ;\ \frac{1}{\sqrt{T}}\int_0^T f(t)\,dt\ ;\ \sqrt{\frac{2}{T}}\int_0^T f(t)\cos n\omega t\,dt\ ;\ \sqrt{\frac{2}{T}}\int_0^T f(t)\sin n\omega t\,dt$ |

**Table 3.11.** *Fourier coefficients for a function $f \in \mathcal{C}^0_T(\mathbb{R}, \mathbb{C})\,,\ \omega = 2\pi/T$*

### 3.8.4. Partial Fourier sums and Bessel's inequality

An $N$th-order expansion can be obtained from an orthogonal projection onto the subspace generated by the set of vectors $\mathcal{B}_N = \{e^{int}, n \in [-N, N]\} =$

Vect$\{e_{-N}, \cdots, e_0, \cdots, e_N\}$, or equivalently Vect$\big(\sin nt, \cos nt, n \in \langle N \rangle\big)$, which gives the $N$th partial sum, denoted by $S_N^f$, of the Fourier series, at point $t$:

$$S_N^f(t) = \sum_{n=-N}^{N} c_n e^{int} = a_0 + \sum_{n=1}^{N} (a_n \cos nt + b_n \sin nt) \qquad [3.91a]$$

$$c_n = \langle f, e_n \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int} dt. \qquad [3.91b]$$

$S_N^f$, which is a linear combination of the vectors $e_n, n \in [-N, N]$, is called a trigonometric series. It corresponds to the sum of the first $N+1$ harmonics of $f$. This is the best approximation of $f$ in the sense of the minimization of $\|f - S_N^f\|_2^2$.

With the simplified expressions of the Fourier coefficients of Table 3.10, we have:

– For an even $f$:

$$S_N^f(t) = a_0 + \sum_{n=1}^{N} a_n \cos nt. \qquad [3.92]$$

– For an odd $f$:

$$S_N^f(t) = \sum_{n=1}^{N} b_n \sin nt. \qquad [3.93]$$

Applying the Pythagorean theorem gives the Bessel inequality:

$$\|S_N^f\|_2^2 \leq \|f\|_2^2 \;\Leftrightarrow\; \sum_{n=-N}^{N} |c_n|^2 \leq \frac{1}{2\pi} \int_0^{2\pi} |f(t)|^2 \, dt, \qquad [3.94]$$

with

$$\sum_{n=-N}^{N} |c_n|^2 = |a_0|^2 + \frac{1}{2} \sum_{n=1}^{N} (|a_n|^2 + |b_n|^2). \qquad [3.95]$$

From the Bessel inequality, it can be concluded that the sequences $(c_n)$, $(a_n)$, and $(b_n)$ tend to 0 when $|n|$ tends to infinity.

### 3.8.5. *Convergence of Fourier series*

The Fourier series of $f$ is said to be convergent if the sequence of the partial sums $(S_N^f)$ converges. The study of convergence of the Fourier series is a difficult

problem which Fourier himself disregarded. The resolution of this problem has been the subject of several theorems depending on the membership space and therefore on the properties of the function $f$ to be represented. These theorems are associated with several types of convergence. The following modes of convergence can thus be distinguished:

– Pointwise (or simple) convergence: It is said that the sequence of functions $(f_n)$ defined on the interval $[a, b]$ of $\mathbb{R}$ and with values in $\mathbb{K}$ converges pointwise (or simply) to $f$ on $[a, b]$ if the sequence $(f_n(t))$ converges to $f(t)$ for all $t \in [a, b]$, or equivalently, if for every $t \in [a, b]$ and for some $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $|f_n(t) - f(t)| < \epsilon$ for $n \geq N$.

– Uniform convergence: It is said that the sequence $(f_n)$ converges uniformly to $f$ on $[a, b]$ if the sequence of difference $|f_n(t) - f(t)|$, for all $t \in [a, b]$, tends to zero when $n \to +\infty$.

Uniform convergence implies pointwise convergence. Indeed, uniform convergence requires that all sequences $f_n(t)$ converge simultaneously to $f(t)$. Thus, from a specific rank $N$, the difference $|f_n(t) - f(t)|$ becomes very small for all $t$, whereas for pointwise convergence, the difference $|f_n(t) - f(t)|$ strongly depends on the value of $t$, and therefore this convergence does not imply uniform convergence. Pointwise convergence is said to be weaker than uniform convergence.

– Almost everywhere convergence: It is said that the sequence $(f_n)$ converges almost everywhere to $f$ on $[a, b]$ if the sequence $(f_n(t))$ converges to $f(t)$ on $[a, b]$ except for a set of values of $t$ of zero measure. Pointwise convergence implies almost everywhere convergence, but the converse does not hold.

– Convergence in mean or in $L_1$ norm: It is said that the sequence $(f_n)$ converges in mean to $f$ if:

- All functions $f_n$ and $f$ belong to the space $L^1([a, b], \mathbb{C})$.

- $\|f_n - f\|_1 \to 0$ when $n \to +\infty$, that is, $\int_a^b |f_n(t) - f(t)|\, dt = 0$ when $n \to +\infty$.

– Convergence in quadratic mean or in $L_2$ norm: It is said that the sequence $(f_n)$ converges in quadratic mean to $f$ if:

- All functions $f_n$ and $f$ belong to the space $L^2([a, b], \mathbb{C})$.

- $\|f_n - f\|_2 \to 0$ when $n \to +\infty$, that is, $\int_a^b |f_n(t) - f(t)|^2\, dt = 0$ when $n \to +\infty$.

In signal processing, the convergence in quadratic mean is convergence in energy for the norm associated with the inner product [3.69].

We have the following implications:

– Uniform convergence implies $L_1$ convergence and $L_2$ convergence.

– Convergence in $L_2$ norm implies convergence in $L_1$ norm. This implication results from the Cauchy–Schwarz inequality [3.37b]:

$$\left| \int_a^b |f_n(t) - f(t)|\, dt \right|^2 \leq (b-a) \int_a^b |f_n(t) - f(t)|^2\, dt.$$

– It should be noted that there is no link between convergence in $L_2$ norm and pointwise and almost everywhere convergences.

The convergence of Fourier series is essentially linked to the following questions: under what conditions and with which mode of convergence does the sequence $(S_N^f)$ converge, and is the Parseval equality satisfied ($f = \lim\limits_{N\to+\infty} S_N^f$)?

The Parseval and Dirichlet–Jordan theorems presented below provide answers to these questions for $2\pi$-periodic functions, from $\mathbb{R}$ to $\mathbb{C}$, piecewise continuous ($f \in \mathcal{C}^0_{m,2\pi}(\mathbb{R},\mathbb{C})$), in the first case, and continuous and piecewise smooth functions[18] ($f \in \mathcal{C}^0_{m,2\pi}(\mathbb{R},\mathbb{C})$ and $f \in \mathcal{C}^1_{m,2\pi}(\mathbb{R},\mathbb{C})$, in the second case.

**Parseval's theorem:** For $f \in \mathcal{C}^0_{m,2\pi}(\mathbb{R},\mathbb{C})$, the sequence of partial sums $(S_N^f)_{N\in\mathbb{N}}$ converges to $f$ in $L_2$ norm, that is:

$$\lim_{N\to+\infty} \|f - S_N^f\|_2^2 = \lim_{N\to+\infty} \frac{1}{2\pi} \int_0^{2\pi} |f(t) - S_N^f(t)|^2 dt = 0,$$

from which can be deduced the Parseval equality by application of the Pythagorean theorem (see section 3.8.1):

$$\lim_{N\to+\infty} \|S_N^f\|_2^2 = \sum_{n\in\mathbb{Z}} |c_n|^2 = |a_0|^2 + \frac{1}{2} \sum_{n\in\mathbb{N}^*} (|a_n|^2 + |b_n|^2) \qquad [3.96a]$$

$$= \|f\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(t)|^2\, dt. \qquad [3.96b]$$

This convergence does not imply that the sequence $S_N$ converges pointwise to $f$. This pointwise convergence at any point $t \in \mathbb{R}$ is established below in the Dirichlet–Jordan theorem.

18 A function $f$ is piecewise smooth if its first derivative $f'$ is piecewise continuous, that is, $f$ is of class $\mathcal{C}^1$, which means that $f$ has left and right derivatives at any point of discontinuity $a$, more specifically $\lim\limits_{t\to a^-} \frac{f(t)-f(a^-)}{t-a^-}$ and $\lim\limits_{t\to a^+} \frac{f(t)-f(a^+)}{t-a^+}$ exist and can be different. See Example 2.17 for the definition of the set $\mathcal{C}^k_m$.

With the inner product [3.85], the Parseval theorem makes it possible to connect the power $P = \|f\|_2^2 = \frac{1}{2\pi}\int_0^{2\pi} |f(t)|^2 dt$ of signal $f$ dissipated over a period, to the sum of the squares of the moduli of its Fourier coefficients, using equalities [3.96a] and [3.96b]. From this expression of power, the curve $\|h_n\|^2 = \frac{1}{2}(|a_n|^2 + |b_n|^2)$ can be plotted according to the frequency $\frac{n}{2\pi}$. This curve, which represents the distribution of the power of $f$ for each harmonic $h_n$, is called the power spectrum of $f$.

In signal processing, the Fourier series expansion of a deterministic signal provides a spectral representation thereof. This is referred to as frequency or harmonic analysis, particularly suited to signals composed of several harmonics. In the case of non-periodic signals, their analysis can be performed using the Fourier transform.

FACT 3.27.– In the literature, one often defines $a_0 = 2c_0 = \frac{1}{\pi}\int_0^{2\pi} f(t)\,dt$ instead of $a_0 = c_0$ as in [3.88a]. The Fourier series [3.87] and the Parseval equality [3.96b] then become:

$$S^f(t) = \frac{1}{2}a_0 + \sum_{n\in\mathbb{N}^*} (a_n \cos nt + b_n \sin nt)$$

$$2\|f\|_2^2 = \frac{1}{\pi}\int_0^{2\pi} |f(t)|^2 dt = \frac{1}{2}|a_0|^2 + \sum_{n\in\mathbb{N}^*} (|a_n|^2 + |b_n|^2).$$

**Dirichlet**[19]**–Jordan**[20] **theorem:** If $f$ is a $2\pi$-periodic function, from $\mathbb{R}$ to $\mathbb{C}$, continuous and piecewise smooth ($f \in \mathcal{C}^0_{m,2\pi}(\mathbb{R},\mathbb{C})$ and $f \in \mathcal{C}^1_{m,2\pi}(\mathbb{R},\mathbb{C})$), then $f$ has a Fourier series:

$$S_N^f(t) = \sum_{n=-N}^{N} c_n e^{int} = a_0 + \sum_{n=1}^{N}(a_n \cos nt + b_n \sin nt), \qquad [3.98]$$

19 Lejeune Dirichlet (1805–1859), a German mathematician who created the analytic theory of numbers and who, because of his work on the convergence of Fourier series and their use to represent arbitrary functions, is considered to be the founder of the theory of Fourier series. In 1837, he proposed the modern definition of a function. Several theorems and mathematical concepts are named after him, as for example, the Dirichlet kernel, integral, function and test.

20 Marie Ennemond Camille Jordan (1838–1922), a French mathematician, polytechnician, and member of the Académie des Sciences, who made many fundamental contributions to the theory of groups and to complex analysis for the calculus of integrals. Several lemmas and theorems bear his name, such as the Jordan–Holder and the Jordan–Schur theorems. In matrix calculus, he is known for the Jordan canonical form, also called Jordan matrix, which is a block diagonal matrix formed of Jordan blocks (see section 5.5.4) and which is linked to the block-diagonal matrix representation of an endomorphism.

such that, at all points $t$ of discontinuity of $f$, we have:

$$\lim_{N\to+\infty} S_N^f(t) = \frac{f(t^+) + f(t^-)}{2}, \qquad [3.99]$$

where $f(t^+)$ and $f(t^-)$ are the right and left limits of $f(s)$, as $s \to t$, and $S_N^f(t)$ converges pointwise to $f(t)$, at all points $t$ where $f$ is continuous:

$$\lim_{N\to+\infty} S_N^f(t) = f(t). \qquad [3.100]$$

FACT 3.28.– The piecewise class $\mathcal{C}^1$ assumption means that $f$ is such that at any point $t$ of discontinuity of the first kind, we have $f(t) = \frac{f(t^-)+f(t^+)}{2}$, namely, $f(t)$ is equal to the average value of its left and right limits at $t$.

Note that the convergence of the partial sums $S_N^f$ in the neighborhood of a discontinuity point is characterized by the Gibbs–Wilbraham[21] phenomenon, consisting in some overshoot with oscillations.

In Table 3.12, we summarize the results related to the Dirichlet–Jordan theorem, for $T$-periodic functions.

| $f \in \mathcal{C}^1_{m,T}(\mathbb{R},\mathbb{C})\ ;\ \omega = \frac{2\pi}{T}$ |
|---|
| $\langle f,g\rangle = \frac{1}{T}\int_0^T f(t)\overline{g}(t)dt\ ;\ \|f\|_2^2 = \frac{1}{T}\int_0^T \|f(t)\|^2 dt$ |
| $S_N^f(t) = \sum_{n=-N}^{N} c_n e^{in\omega t}\ ;\ c_n = \frac{1}{T}\int_0^T f(t)e^{-in\omega t}\,dt$ |
| **Convergence at any continuity point** |
| $\lim_{N\to+\infty} S_N^f(t) = f(t)$ |
| **Convergence at any discontinuity point** |
| $\lim_{N\to+\infty} S_N^f(t) = \frac{f(t^+)+f(t^-)}{2}$ |
| **Parseval's equality** |
| $\frac{1}{T}\int_0^T \|f(t)\|^2\,dt = \sum_{n\in\mathbb{Z}} \|c_n\|^2 = \|a_0\|^2 + \frac{1}{2}\sum_{n\in\mathbb{N}^*}\left(\|a_n\|^2 + \|b_n\|^2\right)$ |

**Table 3.12.** *Dirichlet-Jordan theorem for $T$-periodic functions*

21 This phenomenon has been highlighted for the first time in 1848 by Henry Wilbraham, then mathematically explained, in 1899, by Josiah Willard Gibbs (1839–1903), an American mathematician and physicist whose work focused on thermodynamics, electromagnetism, and statistical mechanics.

### 3.8.6. *Examples of Fourier series*

To illustrate the previous results, we consider six examples of periodic functions whose expressions over one period are given in Table 3.13. It should be noted that, for a function $f : [a, a+T] \to \mathbb{R}$, defined on the interval $[a, a+T]$ of length $T > 0$, its $T$-periodic extension to the whole real line ($\mathbb{R}$), denoted by $f_p$, is such that for $t \in \mathbb{R}$:

$$f_p(t) = f(t - n_t T) \ , \ n_t = \text{Ent}\Big(\frac{t-a}{T}\Big), \qquad [3.101]$$

where Ent (.) represents the integer part.

| **Functions** | $a_0$ | $a_n$ | $b_n$ |
|---|---|---|---|
| $f_1(t) = \begin{cases} -t+1 & \text{if } 0 \le t \le 1 \\ t+1 & \text{if } -1 \le t \le 0 \end{cases}$ | $\frac{1}{2}$ | $\frac{2}{n^2\pi^2}(1-\cos n\pi) = \begin{cases} 0 & \text{if } n = 2k \\ \frac{4}{(2k+1)^2\pi^2} & \text{if } n = 2k+1 \end{cases}$ | $0$ |
| $f_2(t) = \lvert t \rvert \ , \quad \text{if } t \in [-\pi, \pi]$ | $\frac{\pi}{2}$ | $\frac{2}{n\pi^2}((-1)^n - 1) = \begin{cases} 0 & \text{if } n = 2k \\ -\frac{4}{(2k+1)^2\pi} & \text{si } n = 2k+1 \end{cases}$ | $0$ |
| $f_3(t) = t(2\pi - t) \ , \ t \in [0, 2\pi]$ | $\frac{2\pi^2}{3}$ | $-\frac{4}{n^2}$ | $0$ |
| $f_4(t) = t \quad , \quad -\pi \le t \le \pi$ | $0$ | $0$ | $2\frac{(-1)^{n+1}}{n}$ |
| $f_5(t) = \begin{cases} 1 & \text{if } 0 < t < \pi \\ -1 & \text{if } -\pi < t < 0 \end{cases}$ | $0$ | $0$ | $\frac{2}{n\pi}(1-\cos n\pi) = \begin{cases} 0 & \text{if } n = 2k \\ \frac{4}{(2k+1)\pi} & \text{if } n = 2k+1 \end{cases}$ |
| $f_6(t) = \begin{cases} t(\pi - t) \ , \ t \in [0, \pi] \\ t(\pi + t) \ , \ t \in [-\pi, 0] \end{cases}$ | $0$ | $0$ | $\frac{4}{n^3\pi}(1-\cos n\pi) = \begin{cases} 0 & \text{if } n = 2k \\ \frac{8}{(2k+1)^3\pi} & \text{if } n = 2k+1 \end{cases}$ |

**Table 3.13.** *Examples of Fourier series*

After determining the Fourier coefficients of each of these functions, their Fourier series expansions and associated Parseval formulae are used to compute sums of numerical sequences.

The functions $f_1$ and $f_2$ are defined on $[-1, 1]$ and on $[-\pi, \pi]$, respectively. They are extended to $\mathbb{R}$ as 2-periodic and $2\pi$-periodic functions, respectively. These extended functions, which are shaped as triangular sine waves, are piecewise continuous on $\mathbb{R}$ and even. Thereby, their Fourier coefficients $b_n$ are zero.

Figures 3.3 and 3.4 show functions $f_1$ and $f_4$ and their extensions.

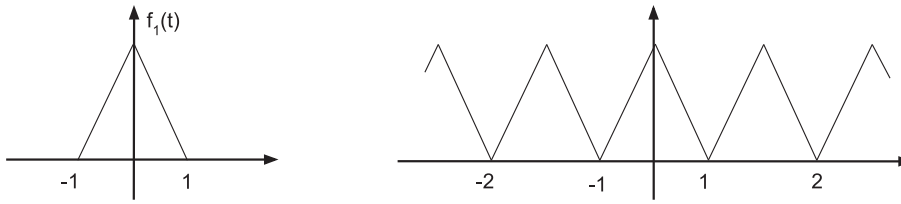



**Figure 3.3.** *Function $f_1$ and its extension*

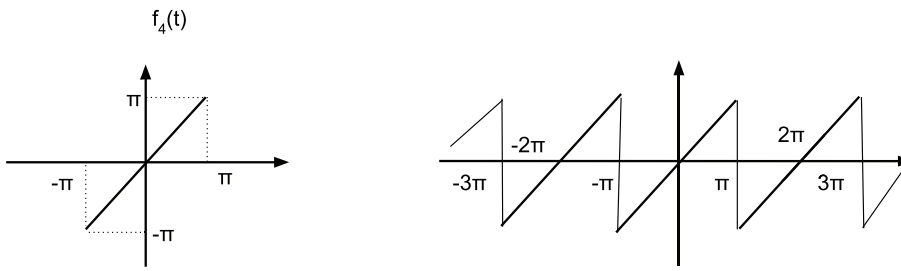



**Figure 3.4.** *Function $f_4$ and its extension*

The function $f_3$, of parabolic shape, defined on $[0, 2\pi]$ and extended as a $2\pi$-periodic function, is such that $f_3(-t) = -t(2\pi + t) = f_3(t + 2\pi) = f_3(t)$. Therefore, this is also an even function, with coefficients $b_n$ equal to zero. This extension is then called an even extension of $f_3$.

On the other hand, the function $f_4$ extended to $\mathbb{R}$ is a sawtooth wave, whereas $f_5$ has the shape of a square wave. The corresponding extended functions are of period $2\pi$ and odd, implying that their coefficients $a_n$ are zero.

Finally, the extended function of $f_6$ is of parabolic shape, $2\pi$-periodic and odd, implying that its coefficients $a_n$ are also zero.

It should be noted that there are discontinuities at points $t \in \{-1, 0, 1\}$ for $f_1$, $t \in \{-\pi, 0, \pi\}$ for $f_2$, and $t \in \{0, 2\pi\}$ for $f_3$, whereas $f_4$ and $f_5$ have jump discontinuities at points $t \in \{-\pi, \pi\}$ and $t \in \{-\pi, 0, \pi\}$, respectively.

The Fourier coefficients of these functions are given in Table 3.13, and the corresponding Fourier series are represented by the following equations:

$$S^{f_1}(t) = \frac{1}{2} + \frac{4}{\pi^2} \sum_{k \in \mathbb{N}} \frac{\cos(2k+1)\pi t}{(2k+1)^2} \quad [3.102a]$$

$$S^{f_2}(t) = \frac{\pi}{2} - \frac{4}{\pi} \sum_{k \in \mathbb{N}} \frac{\cos(2k+1)t}{(2k+1)^2} \quad [3.102b]$$

$$S^{f_3}(t) = \frac{2\pi^2}{3} - 4 \sum_{n \in \mathbb{N}^*} \frac{\cos nt}{n^2} \quad [3.102c]$$

$$S^{f_4}(t) = 2 \sum_{n \in \mathbb{N}^*} (-1)^{n+1} \frac{\sin nt}{n} \quad [3.102d]$$

$$S^{f_5}(t) = \frac{4}{\pi} \sum_{k \in \mathbb{N}} \frac{\sin(2k+1)t}{2k+1} \quad [3.102e]$$

$$S^{f_6}(t) = \frac{8}{\pi} \sum_{k \in \mathbb{N}} \frac{\sin(2k+1)t}{(2k+1)^3}. \quad [3.102f]$$

NOTE 3.29.– We can make the following remarks:

– The functions $f_1$, $f_2$, and $f_3$ being non-negative, they have a positive average value, and therefore their Fourier series contain a constant term.

– The functions $f_1$, $f_2$, and $f_3$ being even, their Fourier series contain only cosine terms which are themselves even functions, whereas the functions $f_4$, $f_5$, and $f_6$ being odd, their Fourier series contain only sine terms which are odd functions.

– The Fourier series $S^{f_2}$, $S^{f_5}$, and $S^{f_6}$ contain only odd harmonics, with frequencies equal to 1, 3, 5, ... times the fundamental frequency $\frac{1}{2\pi}$. Similarly, $S^{f_1}$ contains only odd harmonics, with the fundamental frequency $\frac{1}{2}$. On the other hand, $S^{f_3}$ and $S^{f_4}$ contain all the harmonics of frequency $\frac{n}{2\pi}$, with $n \in \mathbb{N}^*$.

In Table 3.14, we give some examples of sums of series whose expressions are deduced from the series expansions [3.102a]–[3.102f] at specific points $t$.

Similarly, the Parseval equality [3.96b] can be employed to compute sums of series as shown in Table 3.15, with $2\pi$ replaced by $T = 2$ in the case of function $f_1$.

| Values of $f$ | Sums of series |
|---|---|
| $f_1(0) = \frac{1}{2} + \frac{4}{\pi^2} \sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^2} = 1$ | $\sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8}$ |
| $f_2(0) = \frac{\pi}{2} - \frac{4}{\pi} \sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^2} = 0$ | $\sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8}$ |
| $f_3(0) = \frac{2\pi^2}{3} - 4 \sum_{n\in\mathbb{N}^*} \frac{1}{n^2} = 0$ | $\sum_{n\in\mathbb{N}^*} \frac{1}{n^2} = \frac{\pi^2}{6}$ |
| $f_4(\frac{\pi}{2}) = 2 \sum_{n\in\mathbb{N}^*} \frac{(-1)^{n+1}}{n} \sin(\frac{n\pi}{2}) = \frac{\pi}{2}$ | $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots = \frac{\pi}{4}$ |
| $f_5(\frac{\pi}{2}) = \frac{4}{\pi} \sum_{n\in\mathbb{N}} \frac{(-1)^n}{2n+1} = 1$ | $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots = \frac{\pi}{4}$ |
| $f_6(\frac{\pi}{2}) = \frac{8}{\pi} \sum_{n\in\mathbb{N}} \frac{(-1)^n}{(2n+1)^3} = \frac{\pi^2}{4}$ | $\sum_{n\in\mathbb{N}} \frac{(-1)^n}{(2n+1)^3} = \frac{\pi^3}{32}$ |

**Table 3.14.** *Examples of sums of series*

| Parseval equalities | Sums of series |
|---|---|
| $\|f_1\|_2^2 = \frac{1}{4} + \frac{8}{\pi^4} \sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^4} = \frac{1}{3}$ | $\sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^4} = \frac{\pi^4}{96}$ |
| $\|f_2\|_2^2 = \frac{\pi^2}{4} + \frac{8}{\pi^2} \sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^4} = \frac{\pi^2}{3}$ | $\sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^4} = \frac{\pi^4}{96}$ |
| $\|f_3\|_2^2 = \frac{4\pi^4}{9} + 8 \sum_{n\in\mathbb{N}^*} \frac{1}{n^4} = \frac{8\pi^4}{15}$ | $\sum_{n\in\mathbb{N}^*} \frac{1}{n^4} = \frac{\pi^4}{90}$ |
| $\|f_4\|_2^2 = 2 \sum_{n\in\mathbb{N}^*} \frac{1}{n^2} = \frac{\pi^2}{3}$ | $\sum_{n\in\mathbb{N}^*} \frac{1}{n^2} = \frac{\pi^2}{6}$ |
| $\|f_5\|_2^2 = \frac{8}{\pi^2} \sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^2} = 1$ | $\sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8}$ |
| $\|f_6\|_2^2 = \frac{32}{\pi^2} \sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^6} = \frac{\pi^4}{30}$ | $\sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^6} = \frac{\pi^6}{960}$ |

**Table 3.15.** *Sums of series inferred from the Parseval equality*

PROOF.– The extended function of $f_1$ being of period $T = 2$, by making use of its parity property, and by integrating twice by parts, we obtain the following coefficients $a_n$:

$$a_0 = \frac{1}{2}\int_{-1}^{1} f_1(t)dt = \int_{-1}^{0} (t+1)dt = \left[\frac{t^2}{2} + t\right]_{-1}^{0} = \frac{1}{2}$$

$$a_n = \int_{-1}^{1} f_1(t)\cos n\pi t\, dt = 2\int_{-1}^{0} (t+1)\cos n\pi t\, dt\ ,\ n \geq 1$$

$$= 2\Big[\frac{(t+1)\sin n\pi t}{n\pi}\Big]_{-1}^{0} - 2\int_{-1}^{0}\frac{\sin n\pi t}{n\pi}\,dt = \frac{2}{n^2\pi^2}\Big[\cos n\pi t\Big]_{-1}^{0}$$

$$= \frac{2}{n^2\pi^2}(1-\cos n\pi) = \begin{cases} 0 & \text{if } n = 2k, \quad \text{for } k \geq 1 \\ \frac{4}{\pi^2(2k+1)^2} & \text{if } n = 2k+1, \text{ for } k \geq 0 \end{cases}$$

and the Fourier series of $f_1$ is given by:

$$S^{f_1}(t) = a_0 + \sum_{n\in\mathbb{N}^*} a_n \cos n\pi t = \frac{1}{2} + \frac{4}{\pi^2}\sum_{k\in\mathbb{N}}\frac{\cos(2k+1)\pi t}{(2k+1)^2}.$$

According to the Dirichlet–Jordan theorem, $f_1$ converges pointwise, which implies that $f_1(0) = S^{f_1}(0)$, with:

$$f_1(0) = 1 \text{ and } S^{f_1}(0) = \frac{1}{2} + \frac{4}{\pi^2}\sum_{k\in\mathbb{N}}\frac{1}{(2k+1)^2}$$

from which the following can be deduced:

$$\sum_{k\in\mathbb{N}}\frac{1}{(2k+1)^2} = \frac{\pi^2}{8}.$$

In addition, the Parseval equality is written as:

$$\|f_1\|_2^2 = a_0^2 + \frac{1}{2}\sum_{n\geq 1} a_n^2 = \frac{1}{4} + \frac{8}{\pi^4}\sum_{k\in\mathbb{N}}\frac{1}{(2k+1)^4}$$

$$= \frac{1}{2}\int_{-1}^{1} f_1^2(t)dt = \int_{-1}^{0}(t+1)^2 dt = \Big[\frac{(t+1)^3}{3}\Big]_{-1}^{0} = \frac{1}{3},$$

from which the following can be deduced:

$$\sum_{k\in\mathbb{N}}\frac{1}{(2k+1)^4} = \frac{\pi^4}{96}.$$

Noting that $f_2$ can be obtained from $f_1$ after applying a multiplication factor $\pi$ and a time offset of half a period (1), after a time-scale transformation from $t$ into $t/\pi$, or more specifically $f_2(t) = \pi f_1(\frac{t}{\pi}+1)$, the Fourier series of $f_2$ can be deduced from that of $f_1$ by applying this transformation to the expression [3.102a] of $S^{f_1}(t)$, which gives:

$$S^{f_2}(t) = \pi S^{f_1}(\frac{t}{\pi}+1) = \pi\Big[\frac{1}{2} + \frac{4}{\pi^2}\sum_{k\in\mathbb{N}}\frac{\cos(2k+1)\pi(\frac{t}{\pi}+1)}{(2k+1)^2}\Big]$$

$$= \frac{\pi}{2} - \frac{4}{\pi}\sum_{k\in\mathbb{N}}\frac{\cos(2k+1)t}{(2k+1)^2},$$

from which are deduced the coefficients $a_n$ given in Table 3.13. In addition, we have $\|f_2\|_2^2 = \pi^2\|f_1\|_2^2$, which explains the same sum of series for $f_1$ and $f_2$ in Table 3.15. In Table 3.14, it is verified that the sums of series obtained from $f_1(0) = 1$ and $f_2(0) = 0$, at the points separated by half a period, are the same.

For $f_3$, the coefficients $a_n$ are given by formulae [3.88a] and [3.88b]. By integrating by parts, we obtain:

$$\begin{aligned} a_0 &= \frac{1}{2\pi}\int_0^{2\pi} f_3(t)dt = \frac{1}{2\pi}\int_0^{2\pi} t(2\pi - t)dt \\ &= \frac{1}{2\pi}\Big[-\frac{t}{2}(2\pi - t)^2\Big]_0^{2\pi} + \frac{1}{4\pi}\int_0^{2\pi}(2\pi - t)^2 dt \\ &= \frac{1}{12\pi}\Big[-(2\pi - t)^3\Big]_0^{2\pi} = \frac{2\pi^2}{3} \\ a_n &= \frac{1}{\pi}\int_0^{2\pi} t(2\pi - t)\cos nt\, dt\ ,\ n \geq 1. \end{aligned}$$

By using a double integration by parts, we get:

$$\begin{aligned} a_n &= \Big[\frac{t(2\pi - t)\sin nt}{n\pi}\Big]_0^{2\pi} - \frac{1}{n\pi}\int_0^{2\pi}(2\pi - 2t)\sin nt\, dt \\ &= \frac{2}{n^2\pi}\Big[(\pi - t)\cos nt\Big]_0^{2\pi} + \frac{2}{n^2\pi}\int_0^{2\pi}\cos nt\, dt \\ &= -\frac{4}{n^2} + \frac{2}{n^3\pi}\Big[\sin nt\Big]_0^{2\pi} = -\frac{4}{n^2}. \end{aligned}$$

The Fourier series of $f_3$ is thus given by:

$$S^{f_3}(t) = a_0 + \sum_{n\in\mathbb{N}^*} a_n \cos nt = \frac{2\pi^2}{3} - 4\sum_{n\in\mathbb{N}^*}\frac{\cos nt}{n^2}.$$

For $t = 0$, we have $f_3(0) = 0$ and $S^{f_3}(0) = \frac{2\pi^2}{3} - 4\sum\limits_{n\in\mathbb{N}^*}\frac{1}{n^2}$.

Given that the function $f_3$ belongs to the space $\mathcal{C}^1_{m,2\pi}(\mathbb{R},\mathbb{R})$, it is pointwise convergent according to the Dirichlet–Jordan theorem, which implies $S^{f_3}(0) = f_3(0)$ and therefore the following equality:

$$\frac{2\pi^2}{3} - 4\sum_{n\in\mathbb{N}^*}\frac{1}{n^2} = 0 \ \Rightarrow\ \sum_{n\in\mathbb{N}^*}\frac{1}{n^2} = \frac{\pi^2}{6}.$$

On the other hand, the Parseval equality [3.96b] is written as:

$$\|f_3\|_2^2 = a_0^2 + \frac{1}{2}\sum_{n\in\mathbb{N}^*} a_n^2 = \frac{4\pi^4}{9} + 8\sum_{n\in\mathbb{N}^*}\frac{1}{n^4}$$

and using a double integration by parts, we get:

$$\begin{aligned}\|f_3\|_2^2 &= \frac{1}{2\pi}\int_0^{2\pi} t^2(2\pi - t)^2 dt \\ &= \frac{1}{3\pi}\int_0^{2\pi} t(2\pi - t)^3 dt = \frac{1}{12\pi}\int_0^{2\pi} (2\pi - t)^4 dt \\ &= \Big[-\frac{(2\pi - t)^5}{60\pi}\Big]_0^{2\pi} = \frac{8\pi^4}{15}.\end{aligned}$$

The sum of the following series is thus deduced:

$$\sum_{n\in\mathbb{N}^*} \frac{1}{n^4} = \frac{\pi^4}{90}.$$

The extended function of $f_4$ being odd and of period $2\pi$, the coefficients $b_n$ are given by the formula in Table 3.10:

$$\begin{aligned}b_n &= \frac{2}{\pi}\int_0^{\pi} f_4(t)\sin nt\, dt = \frac{2}{\pi}\int_0^{\pi} t\, \sin nt\, dt \\ &= \frac{2}{\pi}\Big[-\frac{t\cos nt}{n}\Big]_0^{\pi} + \frac{2}{\pi}\int_0^{\pi}\frac{\cos nt}{n} dt \\ &= -\frac{2\cos n\pi}{n} + \frac{2}{n^2\pi}\Big[\sin nt\Big]_0^{\pi} = \frac{2(-1)^{n+1}}{n}.\end{aligned}$$

The Fourier series of $f_4$ can be written as:

$$S^{f_4}(t) = \sum_{n\in\mathbb{N}^*} b_n \sin nt = 2\sum_{n\in\mathbb{N}^*} \frac{(-1)^{n+1}}{n}\sin nt.$$

In addition, we have:

$$f_4(\frac{\pi}{2}) = \frac{\pi}{2} \quad \text{and} \quad S^{f_4}(\frac{\pi}{2}) = 2\sum_{n\in\mathbb{N}^*} \frac{(-1)^{n+1}}{n}\sin(\frac{n\pi}{2})$$

from which the following can be deduced:

$$\sum_{n\in\mathbb{N}^*} \frac{(-1)^{n+1}}{n}\sin(\frac{n\pi}{2}) = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + ... = \frac{\pi}{4}.$$

On the other hand, we have:

$$\begin{aligned}\|f_4\|_2^2 &= \frac{1}{2}\sum_{n\geq 1} b_n^2 = 2\sum_{n\in\mathbb{N}^*}\frac{1}{n^2} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} t^2 dt = \frac{\pi^2}{3} \;\Rightarrow\; \sum_{n\in\mathbb{N}^*}\frac{1}{n^2} = \frac{\pi^2}{6}.\end{aligned}$$

The function $f_4$ comprises two singularities on the interval $[-\pi, \pi]$, corresponding to jumps from $\pi$ to $-\pi$ at points $t = -\pi$ and $t = \pi$, with $f_4(-\pi^-) = f_4(\pi^-) = \pi$ and $f_4(-\pi^+) = f_4(\pi^+) = -\pi$. According to the Dirichlet–Jordan theorem, we verify that at these singularity points, we have:

$$S^{f_4}(-\pi) = \frac{f_4(-\pi^+) + f_4(-\pi^-)}{2} = 0$$
$$S^{f_4}(\pi) = \frac{f_4(\pi^+) + f_4(\pi^-)}{2} = 0.$$

Since the extended function of $f_5$ is odd, the coefficients $b_n$ are given by:

$$b_n = \frac{2}{\pi}\int_0^\pi f_5(t)\sin nt\,dt = \frac{2}{\pi}\int_0^\pi \sin nt\,dt$$
$$= \frac{2}{n\pi}\Big[-\cos nt\Big]_0^\pi = \frac{2}{n\pi}(1-\cos n\pi) = \begin{cases} 0 & \text{si } n = 2k \\ \frac{4}{(2k+1)\pi} & \text{if } n = 2k+1 \end{cases}.$$

As a result, the Fourier series of $f_5$ can be written as:

$$S^{f_5}(t) = \frac{4}{\pi}\sum_{k\in\mathbb{N}} \frac{\sin(2k+1)t}{2k+1}.$$

For $t = \frac{\pi}{2}$, we have $f_5(\frac{\pi}{2}) = 1$ and $S^{f_5}(\frac{\pi}{2}) = \frac{4}{\pi}\sum\limits_{n\in\mathbb{N}} \frac{(-1)^n}{2n+1}$. The function $f_5$ being continuous at point $t = \frac{\pi}{2}$, the Dirichlet–Jordan theorem gives:

$$S^{f_5}(\frac{\pi}{2}) = f_5(\frac{\pi}{2}) \;\Rightarrow\; \sum_{n\in\mathbb{N}} \frac{(-1)^n}{2n+1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots = \frac{\pi}{4}.$$

On the other hand, the Parseval equality gives us:

$$\|f_5\|_2^2 = 8\sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^2\pi^2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} dt = 1 \;\Rightarrow\; \sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8}.$$

The extended function of $f_6$ being odd, with double integration by parts, we get:

$$b_n = \frac{2}{\pi}\int_0^\pi f_6(t)\sin nt\,dt = \frac{2}{\pi}\int_0^\pi t(\pi - t)\sin nt\,dt$$
$$= \frac{4}{n^3\pi}(1-\cos n\pi) = \begin{cases} 0 & \text{if } n = 2k \\ \frac{8}{(2k+1)^3\pi} & \text{si } n = 2k+1 \end{cases}.$$

As a result, the Fourier series of $f_6$ is given by:

$$S^{f_6}(t) = \frac{8}{\pi}\sum_{n\in\mathbb{N}} \frac{\sin(2n+1)t}{(2n+1)^3}.$$

The extended function of $f_6$ being $2\pi$-periodic, continuous, and of class $\mathcal{C}^1_{m,2\pi}$, the Dirichlet–Jordan theorem implies that $S^{f_6}$ simply converges to $f$. By considering the point $t = \frac{\pi}{2}$, we have $f_6(\frac{\pi}{2}) = \frac{\pi^2}{4}$, and:

$$S^{f_6}(\frac{\pi}{2}) = \frac{8}{\pi}\sum_{n\in\mathbb{N}} \frac{\sin(2n+1)\pi/2}{(2n+1)^3} = \frac{8}{\pi}\sum_{n\in\mathbb{N}} \frac{(-1)^n}{(2n+1)^3},$$

from which the following is deduced:

$$\sum_{n\in\mathbb{N}} \frac{(-1)^n}{(2n+1)^3} = \frac{\pi^3}{32}.$$

On the other hand, the Parseval equality provides:

$$\begin{aligned}\|f_6\|_2^2 &= \frac{1}{2}\sum_{n\in\mathbb{N}} b_n^2 = \frac{32}{\pi^2}\sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^6} \\ &= \frac{1}{\pi}\int_0^\pi t^2(\pi-t)^2 dt = \frac{\pi^4}{30} \;\Rightarrow\; \sum_{n\in\mathbb{N}} \frac{1}{(2n+1)^6} = \frac{\pi^6}{960}.\end{aligned}$$

This completes the demonstration of the results provided in Tables 3.13–3.15. □

In Table 3.16, we provide the normalized mean squared error (NMSE) for the reconstruction of every function $f$, with $N \in [0, 5]$, as:

$$\text{NMSE}_N^f = \frac{\|f\|_2^2 - \|S_N^f\|_2^2}{\|f\|_2^2} = \frac{\|f\|_2^2 - a_0^2 - \frac{1}{2}\sum_{n=1}^N (a_n^2 + b_n^2)}{\|f\|_2^2},$$

that is:

$$\text{NMSE}_N^{f_1} = \frac{1}{4}(1 - \frac{96}{\pi^4}\sum_{n=0}^N \frac{1}{(2n+1)^4})$$

$$\text{NMSE}_N^{f_3} = \frac{1}{6}(1 - \frac{90}{\pi^4}\sum_{n=1}^N \frac{1}{n^4})$$

$$\text{NMSE}_N^{f_4} = 1 - \frac{6}{\pi^2}\sum_{n=1}^N \frac{1}{n^2}$$

$$\text{NMSE}_N^{f_5} = 1 - \frac{8}{\pi^2}\sum_{n=0}^N \frac{1}{(2n+1)^2}$$

$$\text{NMSE}_N^{f_6} = 1 - \frac{960}{\pi^6}\sum_{n=0}^N \frac{1}{(2n+1)^6}.$$

| $f$ | $N=0$ | $N=1$ | $N=2$ | $N=3$ | $N=4$ | $N=5$ |
|---|---|---|---|---|---|---|
| $f_1$ | 0.25 | 0.0036 | | 0.0006 | | 0.0002 |
| $f_3$ | 0.1667 | 0.0127 | 0.0031 | 0.0012 | 0.0006 | 0.0003 |
| $f_4$ | | 0.3921 | 0.2401 | 0.1726 | 0.1346 | 0.1102 |
| $f_5$ | | 0.1894 | | 0.0994 | | 0.067 |
| $f_6$ | | 0.0015 | | $7.6\,10^{-5}$ | | |

**Table 3.16.** *Normalized mean squared error*

Based on the results shown in Table 3.16, it can be concluded that for $f_1$, $f_3$, and $f_6$, the convergence of the Fourier series (to the function) is very fast, since $N \leq 3$ terms are sufficient to achieve a NMSE smaller than or of the order of $10^{-3}$. On the other hand, for $f_4$ and $f_5$, convergence is much slower, given that values of $N$ greater than 10 are needed to reconstruct the function in an accurate way. This is explained by the form of the functions being considered, those with a shape of triangular sinusoidal waves ($f_1, f_2$) and of parabolas ($f_3, f_6$) being much closer to a pure sinusoidal form than the sawtooth $f_4$ and rectangular pulse $f_5$ functions that are characterized by two abrupt discontinuities in the form of jumps.

These results highlight the limitations of Fourier series expansions. These expansions are suitable for the approximation of periodic functions, with periodicity created by extension of the interval of definition, corresponding in practice to an interval of observation of finite length.

Nonetheless, as we have illustrated through the six examples of functions, the convergence of partial Fourier sum $S_N^f$ is highly dependent on the shape of the signal to be analyzed. In addition, a local perturbation in the signal causes all the Fourier coefficients to be modified. This is not the case of wavelet bases that allow for a multiresolution expansion of the signal to be analyzed, by way of decomposition of the Hilbert space $L^2(\mathbb{R}, \mathbb{R})$ into a sum of orthogonal subspaces, each subspace being associated with a level of resolution.

## 3.9. Expansions over bases of orthogonal polynomials

In the same way as trigonometric Hilbert bases allow periodic functions to be represented and analyzed using Fourier series, it is possible to use polynomial bases for the representation and analysis of functions of the Hilbert space $L^2([a, b], \mathbb{R})$. In the following, we consider Legendre, Hermite, Laguerre, and Chebyshev (of the first kind) orthogonal polynomials that lead to the series named after these authors.

Our goal here is to present, in a unified way, four families of orthogonal polynomials leading to polynomial series, which are to be compared to Fourier series for function approximation. For a more detailed presentation of orthogonal polynomials, consult the book by Beckmann (1973). Numerical examples of function approximation using Legendre, Hermite, and Chebyshev polynomials are given in Guillopé (2010).

In the v.s. of polynomials, we saw that the set of monomials $\{1, t, \cdots, t^n, \cdots\}$ is independent. Considering the weighted inner product:

$$\langle f, g\rangle = \int_a^b w(t) f(t) g(t) dt,$$

various polynomial Hilbert bases can be built by applying the Gram–Schmidt method to the set $\{1, t, t^2, \cdots\}$, for various intervals $[a, b]$ and various weighting functions $w(t)$, associated with different spaces of functions.

In Table 3.17, we summarize the spaces of functions and inner products considered for the construction of orthogonal Legendre, Hermite, Laguerre, and Chebyshev polynomials, which will be denoted by $L_n(t), H_n(t), \Lambda_n(t)$, and $T_n(t)$, respectively. We can observe that Laguerre and Hermite polynomials involve infinite intervals of integration ($[0, \infty]$ and $[-\infty, \infty]$, respectively), whereas for Legendre and Chebyshev polynomials, we have the same finite interval $[-1, 1]$.

| Polynomials | Spaces | Inner products |
|---|---|---|
| Legendre | $L^2([-1,1],\mathbb{R}) = \{f : [-1,1] \to \mathbb{R}, \int_{-1}^{1} f^2(t)dt < \infty\}$ | $\langle f, g\rangle = \frac{1}{2}\int_{-1}^{1} f(t)g(t)dt$ |
| Hermite | $L^2(\mathbb{R},\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R}, \int_{-\infty}^{+\infty} f^2(t)e^{-t^2}dt < \infty\}$ | $\langle f, g\rangle = \int_{-\infty}^{+\infty} f(t)g(t)e^{-t^2}dt$ |
| Laguerre | $L^2(\mathbb{R}^+,\mathbb{R}) = \{f : \mathbb{R}^+ \to \mathbb{R}, \int_{0}^{\infty} f^2(t)e^{-t}dt < \infty\}$ | $\langle f, g\rangle = \int_{0}^{\infty} f(t)g(t)e^{-t}dt$ |
| Chebyshev | $L^2([-1,1],\mathbb{R}) = \{f : [-1,1] \to \mathbb{R}, \int_{-1}^{1} f^2(t)\frac{dt}{\sqrt{1-t^2}} < \infty\}$ | $\langle f, g\rangle = \frac{2}{\pi}\int_{-1}^{1} f(t)g(t)\frac{dt}{\sqrt{1-t^2}}$ |

**Table 3.17.** *Spaces of polynomials and associated inner products*

There are different manners of defining these polynomials.

In Table 3.18, we present Rodrigues formulae (with $n \in \mathbb{N}^*$) and three-term recurrence relations satisfied by the four families of polynomials. Rodrigues formulae allow the orthogonal polynomials to be generated from successive derivations of different functions. The original formula was obtained by Olinde Rodrigues for Legendre polynomials in 1816.

| Polynomials | Rodrigues formulae | Recurrence relations |
|---|---|---|
| Legendre | $L_n(t) = \frac{1}{2^n n!}\frac{d^n}{dt^n}(t^2-1)^n$ | $(n+1)L_{n+1}(t) = (2n+1)\,tL_n(t) - nL_{n-1}(t)$<br>$L_0(t) = 1, L_1(t) = t\ ,\ \forall n \geq 1$ |
| Hermite | $H_n(t) = (-1)^n e^{t^2}\frac{d^n}{dt^n}(e^{-t^2})$ | $H_{n+1}(t) = 2tH_n(t) - 2nH_{n-1}(t)$<br>$H_0(t) = 1, H_1(t) = 2t\ ,\ \forall n \geq 1$ |
| Laguerre | $\Lambda_n(t) = \frac{e^t}{n!}\frac{d^n}{dt^n}(t^n e^{-t})$ | $(n+1)\Lambda_{n+1}(t) = (2n+1-t)\Lambda_n(t) - n\Lambda_{n-1}(t)$<br>$\Lambda_0(t) = 1, \Lambda_1(t) = 1-t\ ,\ \forall n \geq 1$ |
| Chebyshev | $T_n(t) = \frac{(-1)^n 2^n n!}{(2n)!}\sqrt{1-t^2}\frac{d^n}{dt^n}$<br>$\times\left((1-t^2)^{n-\frac{1}{2}}\right)$ | $T_{n+1}(t) = 2t\,T_n(t) - T_{n-1}(t)$<br>$T_0(t) = 1, T_1(t) = t\ ,\ \forall n \geq 1$ |

**Table 3.18.** *Rodrigues formulae and recurrence relations*

For Chebyshev polynomials, defining the polynomial $T_n(t)$ by using the following relation:

$$T_n(t) = \cos n\theta \text{ where } t = \cos\theta \Leftrightarrow T_n(t) = \cos\left(n\arccos(t)\right)\ t \in [-1,1],$$

the recurrence relation satisfied by $T_n$ derives from the trigonometric identity:

$$\cos((n+1)\theta) + \cos((n-1)\theta) = 2\cos\theta\cos n\theta \Rightarrow T_{n+1}(t) = 2t\,T_n(t) - T_{n-1}(t).$$

NOTE 3.30.– Note that the recurrence relation for Legendre polynomials ensures a normalization such that $L_n(1) = 1$, for all $n$, which explains the difference between the polynomials given in Table 3.20 and the ones obtained using the GS algorithm in Example 3.23.

In Table 3.19, we present the differential equations satisfied by every family of polynomials.

| Polynomials | Differential equations, $\forall n \in \mathbb{N}$ |
|---|---|
| Legendre | $(1-t^2)L_n''(t) - 2tL_n'(t) + n(n+1)L_n(t) = 0$ |
| Hermite | $H_n''(t) - 2tH_n'(t) + 2nH_n(t) = 0$ |
| Laguerre | $t\Lambda_n''(t) + (1-t)\Lambda_n'(t) + n\Lambda_n(t) = 0$ |
| Chebyshev | $(1-t^2)T_n''(t) - tT_n'(t) + n^2T_n(t) = 0$ |

**Table 3.19.** *Differential equations*

In Table 3.20, we present the first four polynomials of each of these polynomial bases $(L_n, H_n, \Lambda_n, T_n, n \in [0,3])$.

| Polynomials | Legendre | Hermite | Laguerre | Chebyshev |
|---|---|---|---|---|
| $n=0$ | $1$ | $1$ | $1$ | $1$ |
| $n=1$ | $t$ | $2t$ | $-t+1$ | $t$ |
| $n=2$ | $\frac{3t^2-1}{2}$ | $4t^2-2$ | $\frac{1}{2}(t^2-4t+2)$ | $2t^2-1$ |
| $n=3$ | $\frac{5t^3-3t}{2}$ | $8t^3-12t$ | $\frac{1}{6}(-t^3+9t^2-18t+6)$ | $4t^3-3t$ |

**Table 3.20.** *First four Legendre, Hermite, Laguerre and Chebyshev polynomials*

In Table 3.21, we present the orthogonality relations for each family of orthogonal polynomials. As we can see from this table, the polynomials built in this way are indeed orthogonal, but Legendre and Hermite polynomials are not unitary. To make them unitary, we have to simply divide them by their norm, which gives $\sqrt{2n+1}L_n(t)$ and $\frac{1}{(2^n n!\sqrt{\pi})^{1/2}}H_n(t)$, respectively.

| Polynomials | Orthogonality relations |
|---|---|
| Legendre | $\langle L_n, L_p\rangle = \frac{1}{2}\int_{-1}^{1} L_n(t)L_p(t)dt = \frac{1}{2n+1}\delta_{np}$ |
| Hermite | $\langle H_n, H_p\rangle = \int_{-\infty}^{+\infty} H_n(t)H_p(t)e^{-t^2}dt = 2^n n!\sqrt{\pi}\delta_{np}$ |
| Laguerre | $\langle \Lambda_n, \Lambda_p\rangle = \int_0^{\infty} \Lambda_n(t)\Lambda_p(t)e^{-t}dt = \delta_{np}$ |
| Chebyshev | $\langle T_n, T_p\rangle = \frac{2}{\pi}\int_{-1}^{1} T_n(t)T_p(t)\frac{dt}{\sqrt{1-t^2}} = \begin{cases} 0, n \neq p \\ 2, n = p = 0 \\ 1, n = p \neq 0 \end{cases}$ |

**Table 3.21.** *Orthogonality relations*

In Table 3.22, we describe the series of orthogonal polynomials associated with the four families of polynomials.

It should be noted that, by analogy with formula [3.50], Legendre and Hermite coefficients are given by $\|L_n\|^{-2}\langle f, L_n\rangle$ and $\|H_n\|^{-2}\langle f, H_n\rangle$, respectively.

For instance, expanding a function $f(t)$, defined in $[-1,1]$, over the Hilbert basis of Legendre polynomials, gives the Legendre series of $f(t)$ defined as:

$$f(t) = \sum_{n=0}^{\infty} c_n L_n(t). \qquad [3.115]$$

| Polynomials | Series | Coefficients |
|---|---|---|
| Legendre | $S^f(t) = \sum_{n\in\mathbb{N}} c_n L_n(t)$ | $c_n = \frac{2n+1}{2}\int_{-1}^{1} f(t)L_n(t)dt$ |
| Hermite | $S^f(t) = \sum_{n\in\mathbb{N}} c_n H_n(t)$ | $c_n = \frac{1}{2^n n!\sqrt{\pi}}\int_{-\infty}^{+\infty} f(t)H_n(t)e^{-t^2}dt$ |
| Laguerre | $S^f(t) = \sum_{n\in\mathbb{N}} c_n \Lambda_n(t)$ | $c_n = \int_0^{\infty} f(t)\Lambda_n(t)e^{-t}dt$ |
| Chebyshev | $S^f(t) = \frac{1}{2}c_0 + \sum_{n\in\mathbb{N}^*} c_n T_n(t)$ | $c_n = \frac{2}{\pi}\int_{-1}^{1} f(t)T_n(t)\frac{dt}{\sqrt{1-t^2}}$ |

**Table 3.22.** *Polynomial series*

Multiplying both sides of this equation by $L_p(t)$, integrating over the interval $[-1,+1]$, and using the orthogonality relationship given in Table 3.21, we obtain:

$$\frac{1}{2}\int_{-1}^{1} f(t)L_p(t)dt = \frac{1}{2}\sum_{n=0}^{\infty} c_n \int_{-1}^{1} L_n(t)L_p(t)dt = \frac{1}{2p+1}c_p \qquad [3.116]$$

which leads to:

$$c_p = \frac{2p+1}{2}\int_{-1}^{1} f(t)L_p(t)dt = \|L_p\|^{-2}\langle f, L_p\rangle. \qquad [3.117]$$

Polynomial series behave like Fourier series with changes of variable. They thus have the same convergence properties.

For example, in the case of Legendre series, the Dirichlet–Jordan theorem becomes: If $f$ is a piecewise continuous function and of class $\mathcal{C}^1$ on $[-1,1]$, with left- and right-hand side limits at any point, then:

$$\lim_{N\to\infty}\sum_{n=0}^{N} c_n L_n(t) = \frac{f(t^+)+f(t^-)}{2},\ t\in[-1,1].$$

On the other hand, the Parseval formula for Legendre series is given by:

$$\|f\|_2^2 = \sum_{n\in\mathbb{N}}|c_n|^2 = \sum_{n\in\mathbb{N}}\frac{|\langle f, L_n\rangle|^2}{\|L_n\|^2} = \sum_{n\in\mathbb{N}}\frac{2n+1}{2}\int_{-1}^{1} f(t)L_n(t)dt.$$

Similar formulae exist for Hermite, Laguerre, and Chebyshev series.

# 4

# Matrix Algebra

## 4.1. Chapter summary

The objective of this chapter[1] is to present an algebraic framework for the study of matrices and to introduce bilinear/sesquilinear maps and bilinear/sesquilinear forms with their associated matrices. Having defined the main notations and a few special matrices in sections 4.2 and 4.3, we describe the transposition and vectorization operations in sections 4.4 and 4.5. In section 4.6, the notions of inner product, norm, and orthogonality are illustrated in the case of Euclidean vectors. Next, in section 4.7, matrix multiplication is considered, including the definition of periodic, nilpotent, and idempotent matrices. Matrix trace and Frobenius norm are defined in section 4.8.

In section 4.9, we present the fundamental subspaces associated with a matrix, based on which the matrix rank is defined in section 4.10. Then, the notions of determinant, inverse, auto-inverse, and generalized inverse are described in section 4.11, along with the definition and a summary of the properties of the Moore–Penrose pseudo-inverse. Different structures of multiplicative groups of matrices are defined in section 4.12.

An important part of the chapter is dedicated to matrix representations of a linear map (section 4.13) and of a bilinear/sesquilinear form (section 4.14), with the study of the effect of changes of bases resulting in the definition of equivalent, similar, and congruent matrices. Note that bilinear maps and bilinear forms are particular cases of multilinear maps and multilinear forms which play a fundamental role in multilinear algebra, as will be shown in Chapter 6 with the introduction of the notions of hypermatrices, tensor product spaces, and tensors.

1 This chapter was co-written with J. Henrique DE MORAIS GOULART.

*From Algebraic Structures to Tensors,*
First Edition. Edited by Gérard Favier.

Quadratic forms and Hermitian forms are considered in section 4.15, with the introduction of symmetric and Hermitian matrices. A brief presentation of the Gauss reduction method and of the Sylvester's inertia law is made.

Finally, in the last two sections, we define the notions of eigenvalue and eigenvector, and then that of generalized eigenvalue. Their main properties are described with a particular attention to the cases of symmetric/Hermitian matrices as well as orthogonal/unitary matrices. The interpretation of eigenvalues as extrema of the Rayleigh quotient is demonstrated.

## 4.2. Matrix vector spaces

### 4.2.1. *Notations and definitions*

Scalars, column vectors, matrices, and hypermatrices/tensors of order higher than two will be denoted by lowercase letters $(a, b, \ldots)$, bold lowercase letters $(\mathbf{a}, \mathbf{b}, \ldots)$, bold uppercase letters $(\mathbf{A}, \mathbf{B}, \ldots)$, and calligraphic letters $(\mathcal{A}, \mathcal{B}, \ldots)$, respectively.

$\mathbf{1}_I$ is the column vector of dimension $I$, in which all the elements are equal to 1. Matrices $\mathbf{0}_{I\times J}$ and $\mathbf{1}_{I\times J}$ of dimensions $I \times J$ have all their elements equal to 0 and 1, respectively. $\mathbf{I}_N = [\delta_{ij}]$, with $i, j \in \langle N\rangle$, designates the $N$th-order identity matrix, $\delta_{ij}$ being the Kronecker delta.

A matrix $\mathbf{A}$ of dimensions $I \times J$, with $I$ and $J \in \mathbb{N}^*$, denoted by $\mathbf{A}(I, J)$, is an array of $IJ$ elements stored in $I$ rows and $J$ columns; the elements belong to a field $\mathbb{K}$. Its $i$th row and $j$th column, denoted by $\mathbf{A}_{i.}$ and $\mathbf{A}_{.j}$, respectively, are called $i$th row vector and $j$th column vector. The element located at the intersection of $\mathbf{A}_{i.}$ and $\mathbf{A}_{.j}$ is designated by $a_{ij}$ or $a_{i,j}$ or $(\mathbf{A})_{ij}$ or still $(\mathbf{A})_{i,j}$. We will use the notation $\mathbf{A} = [a_{ij}]$, with $a_{i,j} \in \mathbb{K}$, $i \in \langle I\rangle$, and $j \in \langle J\rangle$.

A matrix is said to be real (complex, quaternionic, or octonionic) if its elements are real numbers (complex numbers, quaternions, or octonions). In this book, we will mainly consider real and complex matrices and hypermatrices. Quaternionic matrices and octonionic matrices will be considered in Volume 2.

The sets of real or complex matrices of dimensions $I \times J$ are denoted by $\mathbb{K}^{I\times J}$, with $\mathbb{K} = \mathbb{R}$ for real matrices and $\mathbb{K} = \mathbb{C}$ for complex matrices[2].

2 Sets of quaternionic and octonionic matrices of dimensions $I \times J$ will be denoted by $\mathbb{Q}^{I\times J}$ and $\mathbb{O}^{I\times J}$, respectively. In the case of polynomial matrices, the elements are polynomials that may belong to the ring $\mathbb{K}[z]$, $\mathbb{K}[z^{-1}]$, or $\mathbb{K}[z, z^{-1}]$, that is, the rings of polynomials in indeterminates $z$, $z^{-1}$, or $z$ with positive and negative powers, respectively, and coefficients in $\mathbb{K}$. The ring $\mathbb{K}[z, z^{-1}]$ is that of Laurent polynomials. The set of polynomial matrices of dimensions $I \times J$, whose elements belong to these rings, are denoted by $\mathbb{K}^{I\times J}[z]$, $\mathbb{K}^{I\times J}[z^{-1}]$, and $\mathbb{K}^{I\times J}[z, z^{-1}]$, respectively. Polynomial matrices will be covered in a chapter of Volume 2.

A matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ is written in the form:

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,J} \\ a_{2,1} & a_{2,2} & \dots & a_{2,J} \\ \vdots & \vdots & & \vdots \\ a_{I,1} & a_{I,2} & \dots & a_{I,J} \end{bmatrix}.$$

If $I = J$, it is said that $\mathbf{A}$ is square of order $I$, whereas if $I \neq J$, $\mathbf{A}$ is said to be rectangular. Coefficients $a_{ii}, i \in \langle I \rangle$, of a square matrix of order $I$, are called the diagonal elements; they form the main diagonal. The elements $a_{i,I-i+1}, i \in \langle I \rangle$, constitute the secondary diagonal also known as the antidiagonal. In the case of a rectangular matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ with $I > J$, the elements $a_{jj}$, $j \in \langle J \rangle$, called pseudo-diagonal elements, form the main pseudo-diagonal, while $a_{j,J-j+1}, j \in \langle J \rangle$, constitute the secondary pseudo-diagonal. If $J > I$, it is necessary to swap $J$ with $I$ in the definitions of the pseudo-diagonals.

The special cases $I = 1$ and $J = 1$ correspond respectively to row vectors of dimension $J$ and to column vectors of dimension $I$:

$$\mathbf{v} = [v_1 \cdots v_J] = \begin{bmatrix} v_1 \\ \vdots \\ v_J \end{bmatrix}^T \in \mathbb{K}^{1 \times J} \quad , \quad \mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_I \end{bmatrix} \in \mathbb{K}^{I \times 1}.$$

In the following, for column vectors, $\mathbb{K}^I$ will be used instead of $\mathbb{K}^{I \times 1}$.

FACT 4.1.– The matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ can be defined as a map $f$ from the Cartesian product $\langle I \rangle \times \langle J \rangle$ to $\mathbb{K}$ such that $(i, j) \mapsto f(i, j) = a_{ij} \in \mathbb{K}$, $f(i, j)$ representing the value of $f$ at position $(i, j)$ of $\mathbf{A}$.

FACT 4.2.– In sections 4.13 and 4.14, we will see that a matrix can be associated with a linear map and a bilinear form.

### 4.2.2. *Partitioned matrices*

Another notation consists in partitioning a matrix into blocks which can be matrices or vectors themselves. Such a matrix is written as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \dots & \mathbf{A}_{1,S} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \dots & \mathbf{A}_{2,S} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_{R,1} & \mathbf{A}_{R,2} & \dots & \mathbf{A}_{R,S} \end{bmatrix} \in \mathbb{K}^{(I_1+\cdots+I_R)\times(J_1+\cdots+J_S)}, \qquad [4.1]$$

where $\mathbf{A}_{r,s} \in \mathbb{K}^{I_r \times J_s}$, with $r \in \langle R \rangle$ and $s \in \langle S \rangle$.

In [4.1], we have $(\mathbf{A})_{i,j} = (\mathbf{A}_{r,s})_{i-I_1-\cdots-I_{r-1},j-J_1-\cdots-J_{s-1}}$, where $r$ is the smallest integer such that $i - I_1 - \cdots - I_{r-1} > 0$ and $s$ is the smallest integer such that $j - J_1 - \cdots - J_{s-1} > 0$. If $I_r = 1$ for a value of $r$, then the $r$th row contains blocks that are row vectors (for columns such that $J_s > 1$) or scalars (for $J_s = 1$). A similar reasoning applies to columns if $J_s = 1$. Therefore, for example, if $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$ are $I \times J$ matrices, then

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

is a matrix $2I \times 2J$ that satisfies:

$$(\mathbf{A})_{i,j} = a_{i,j} = \begin{cases} (\mathbf{A}_{11})_{i,j}\,, & \text{if } 1 \leq i \leq I \text{ and } 1 \leq j \leq J, \\ (\mathbf{A}_{12})_{i,j-J}\,, & \text{if } 1 \leq i \leq I \text{ and } J+1 \leq j \leq 2J, \\ (\mathbf{A}_{21})_{i-I,j}\,, & \text{if } I+1 \leq i \leq 2I \text{ and } 1 \leq j \leq J, \\ (\mathbf{A}_{22})_{i-I,j-J}\,, & \text{if } I+1 \leq i \leq 2I \text{ and } J+1 \leq j \leq 2J. \end{cases}$$

A second example involving blocks of different sizes, is:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \mathbf{a}_{12}^T \\ \mathbf{a}_{21} & \mathbf{A}_{22} \end{bmatrix} \in \mathbb{K}^{I \times J},$$

where $a_{11} \in \mathbb{K}$, $\mathbf{a}_{12} \in \mathbb{K}^{J-1}$, $\mathbf{a}_{21} \in \mathbb{K}^{I-1}$, and $\mathbf{A}_{22} \in \mathbb{K}^{I-1 \times J-1}$.

Partitioned matrices will be studied in detail in Chapter 5.

### 4.2.3. *Matrix vector spaces*

In the following, we present addition and scalar multiplication operations that allow to equip the set of matrices $\mathbb{K}^{I \times J}$ with a v.s. structure:

– Addition[3] of two matrices $\mathbf{A}$ and $\mathbf{B}$ of $\mathbb{K}^{I \times J}$ such that:

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}] \in \mathbb{K}^{I \times J}.$$

– Multiplication of a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ by a scalar $\alpha \in \mathbb{K}$ such that:

$$\alpha\mathbf{A} = [\alpha\, a_{ij}] \in \mathbb{K}^{I \times J}.$$

Equipped with these two operations, the set $\mathbb{R}^{I \times J}$ is a $\mathbb{R}$-v.s., while $\mathbb{C}^{I \times J}$ is a $\mathbb{C}$-v.s.

3 Addition of two matrices needs that they have the same dimensions. In this case, they are said to be conformable for addition.

Indeed, it is easy to verify that for the set $\mathbb{K}^{I\times J}$, the axioms of definition of a v.s. (see Table 2.6) are satisfied for all matrices $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{K}^{I\times J}$ and for all scalars $\alpha, \beta \in \mathbb{K}$.

I. (i) $\mathbf{A} + \mathbf{B} \in \mathbb{K}^{I\times J}$

(ii) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

(iii) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$

(iv) $\mathbf{A} + \mathbf{0}_{I\times J} = \mathbf{A}$

(v) $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}_{I\times J}$.

II. (i) $\alpha\mathbf{A} \in \mathbb{K}^{I\times J}$

(ii) $\alpha(\beta\mathbf{A}) = (\alpha\beta)\mathbf{A}$

(iii) $(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$

(iv) $\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$

(iv) $1\mathbf{A} = \mathbf{A}$.

The vector space $\mathbb{K}^{I\times J}$ is of dimension $IJ$. In the case where $I = J$, by equipping the set $\mathbb{K}^{I\times I}$ with matrix multiplication (see section 4.7), this set has an algebra structure. It should be noted that this algebra is not commutative because, in general, $\mathbf{AB} \neq \mathbf{BA}$.

NOTE 4.3.– In the literature, the set $\mathbb{K}^{I\times J}$ of rectangular matrices, of dimensions $I \times J$, is often denoted $\mathcal{M}_{I,J}(\mathbb{K})$. Similarly, the set $\mathbb{K}^{I\times I}$ of square matrices of order $I$ is denoted $\mathcal{M}_I(\mathbb{K})$.

Every real or complex matrix $\mathbf{A} = [a_{ij}] \in \mathbb{K}^{I\times J}$ can be written in the canonical basis as:

$$\mathbf{A} = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}\mathbf{E}_{ij}^{(I\times J)}.$$

The elements $a_{ij}$ represent the coordinates of $\mathbf{A}$ in the canonical basis $\{\mathbf{E}_{ij}^{(I\times J)}\}$, $\mathbf{E}_{ij}^{(I\times J)} \in \mathbb{R}^{I\times J}$ containing 1 at position $(i, j)$ and 0s elsewhere.

## 4.3. Some special matrices

In Table 4.1, we define a few special matrices. For a doubly stochastic matrix, the sum of the elements of each row and of each column is equal to 1. This type of matrices appears in certain probabilistic models such as Markov chains.

| Properties | | Conditions |
|---|---|---|
| $\mathbf{A}$ positive (denoted $\mathbf{A} > 0$) | If | $a_{ij} > 0 \ \ \forall i \in \langle I \rangle \ , \ \ \forall j \in \langle J \rangle$ |
| $\mathbf{A}$ non-negative (denoted $\mathbf{A} \geq 0$) | If | $a_{ij} \geq 0 \ \ \forall i \in \langle I \rangle \ , \ \ \forall j \in \langle J \rangle$ |
| $\mathbf{A}$ (square) diagonal | If | $a_{ij} = 0 \ \ \forall i \neq j$ |
| $\mathbf{A}$ diagonally dominant | If | $\mid a_{ii} \mid \geq \sum_{j \neq i} \mid a_{ij} \mid \ \ \forall i \in \langle I \rangle$ |
| $\mathbf{A}$ strongly diagonally dominant | If | the above inequality is strict for a value of $i$ at least |
| $\mathbf{A}$ strictly diagonally dominant | If | the inequality is strict for all $i$ |
| $\mathbf{A}$ (square) stochastic | If | $\mathbf{A} \geq 0$ and $\sum_{j=1}^{J} a_{ij} = 1, \ \ \forall i \in \langle I \rangle$ |
| $\mathbf{A}$ (square) doubly stochastic | If | $\mathbf{A} \geq 0$ and $\sum_{j=1}^{I} a_{ij} = \sum_{j=1}^{I} a_{ji} = 1, \ \ \forall i \in \langle I \rangle$ |

**Table 4.1.** *Special matrices*

In Table 4.2, we present four types of upper triangular matrices. Note that for an upper triangular matrix, all the elements below the main diagonal are zero, whereas for an upper Hessenberg matrix, the elements below the first subdiagonal (i.e. the diagonal below the main diagonal) are the ones that are zero. Symmetrically, with respect to the main diagonal, we can define four types of lower triangular matrices.

| Structures | | Conditions |
|---|---|---|
| Upper triangular | If | $a_{ij} = 0$ for $i > j$ |
| Unit upper triangular | If | $a_{ij} = 0$ for $i > j$ and $a_{ii} = 1$ |
| Strictly upper triangular | If | $a_{ij} = 0$ for $i \geq j$ |
| Upper Hessenberg | If | $a_{ij} = 0$ for $i > j + 1$ |

**Table 4.2.** *Upper triangular matrices*

FACT 4.4.– The set of upper (or lower) triangular matrices, denoted by $\mathbb{K}_{ut}^{I \times I}$ (or $\mathbb{K}_{lt}^{I \times I}$), is a subspace of $\mathbb{K}^{I \times I}$, of dimension $I(I+1)/2$, with the canonical basis $\{\mathbf{E}_{ij}^{(I \times J)}, \ i \leq j\}$ (respectively, $\{\mathbf{E}_{ij}^{(I \times J)}, \ j \leq i\}$).

## 4.4. Transposition and conjugate transposition

The transpose and the conjugate transpose (also called transconjugate) of a column vector

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_I \end{bmatrix} \in \mathbb{C}^I,$$

denoted by $\mathbf{u}^T$ and $\mathbf{u}^H$, respectively, are the row vectors defined as:

$$\mathbf{u}^T = [u_1, \cdots, u_I] \quad \text{and} \quad \mathbf{u}^H = [u_1^*, \cdots, u_I^*],$$

where $u_i^*$ is the conjugate of $u_i$ also denoted by $\bar{u}_i$.

The transpose of $\mathbf{A} \in \mathbb{K}^{I\times J}$ is the matrix denoted by $\mathbf{A}^T$, of dimensions $J \times I$, such that $\mathbf{A}^T = [a_{ji}]$, with $i \in \langle I\rangle$ and $j \in \langle J\rangle$.

In the case of a complex matrix, the conjugate transpose, also known as Hermitian transpose and denoted by $\mathbf{A}^H$, is defined as:

$$\mathbf{A}^H = (\mathbf{A}^*)^T = (\mathbf{A}^T)^* = [a^*_{ji}],$$

where $\mathbf{A}^* = [a^*_{ij}]$ is the conjugate of $\mathbf{A}$. By decomposing $\mathbf{A}$ using its real and imaginary parts, we have[4]:

$$\mathbf{A} = \mathrm{Re}(\mathbf{A}) + j\,\mathrm{Im}(\mathbf{A}) \Rightarrow \left\{ \begin{array}{l} \mathbf{A}^T = [\mathrm{Re}(\mathbf{A})]^T + j\,[\mathrm{Im}(\mathbf{A})]^T \\ \mathbf{A}^H = [\mathrm{Re}(\mathbf{A})]^T - j\,[\mathrm{Im}(\mathbf{A})]^T \end{array} \right\}.$$

PROPOSITION 4.5.– *The operations of transposition and conjugate transposition satisfy:*

$$(\mathbf{A}^T)^T = \mathbf{A}, \quad (\mathbf{A}^H)^H = \mathbf{A}, \tag{4.2a}$$

$$(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T, \quad (\mathbf{A}+\mathbf{B})^H = \mathbf{A}^H + \mathbf{B}^H, \tag{4.2b}$$

$$(\alpha\mathbf{A})^T = \alpha\mathbf{A}^T, \quad (\alpha\mathbf{A})^H = \alpha^*\mathbf{A}^H, \tag{4.2c}$$

*for any matrix* $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{I\times J}$ *and any scalar* $\alpha \in \mathbb{C}$.

FACT 4.6.– For a real matrix $\mathbf{A} \in \mathbb{R}^{I\times J}$, we have $\mathbf{A}^* = \mathbf{A}$ and therefore, $\mathbf{A}^H = \mathbf{A}^T$.

According to the canonical basis introduced in section 4.2.3, $\mathbf{A}^T$ and $\mathbf{A}^H$ can be developed as follows:

$$\mathbf{A}^T = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}\mathbf{E}^{(J\times I)}_{ji}, \quad \mathbf{A}^H = \sum_{i=1}^{I}\sum_{j=1}^{J} a^*_{ij}\mathbf{E}^{(J\times I)}_{ji}.$$

The transpose (or conjugate transpose) of a matrix $\mathbf{A}$ partitioned into $RS$ blocks $\mathbf{A}_{rs}$, with $r \in \langle R\rangle, s \in \langle S\rangle$, is obtained by transposing the blocks, and then by blockwise transposition (or conjugate transposition). For example, for $R = S = 2$, we have:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \Rightarrow \mathbf{A}^T = \begin{bmatrix} \mathbf{A}^T_{11} & \mathbf{A}^T_{21} \\ \mathbf{A}^T_{12} & \mathbf{A}^T_{22} \end{bmatrix}, \quad \mathbf{A}^H = \begin{bmatrix} \mathbf{A}^H_{11} & \mathbf{A}^H_{21} \\ \mathbf{A}^H_{12} & \mathbf{A}^H_{22} \end{bmatrix}.$$

4 It should be noted that the symbol $j$ is used in the text both as an index (usually associated with the columns of a matrix) and for denoting $\sqrt{-1}$. The context of each instance makes it possible to eliminate any ambiguity. We also use $i$ for $\sqrt{-1}$.

## 4.5. Vectorization

A very widely used operation in matrix computation is vectorization which consists of stacking the columns of a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ on top of each other to form a column vector of dimension $JI$:

$$\mathbf{A} = \left[\mathbf{A}_{.1} \cdots \mathbf{A}_{.J}\right] \in \mathbb{K}^{I \times J} \quad \Rightarrow \quad \text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{A}_{.1} \\ \vdots \\ \mathbf{A}_{.J} \end{bmatrix} \in \mathbb{K}^{JI}.$$

This operation defines an isomorphism between the vector space $\mathbb{K}^{JI}$ of vectors of dimension $JI$ and the vector space $\mathbb{K}^{I \times J}$ of matrices $I \times J$. Indeed, the canonical basis of $\mathbb{K}^{JI}$, denoted by $\left\{\mathbf{e}^{(JI)}_{(j-1)I+i}\right\}$, allows us to write $\text{vec}(\mathbf{A})$ as:

$$\mathbf{A} = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}\, \mathbf{e}^{(I)}_i \circ \mathbf{e}^{(J)}_j \quad \Rightarrow \quad \text{vec}(\mathbf{A}) = \sum_{i=1}^{I}\sum_{J=1}^{J} a_{ij}\, \mathbf{e}^{(JI)}_{(j-1)I+i},$$

with $\mathbf{e}^{(JI)}_{(j-1)I+i} = \text{vec}\left(\mathbf{e}^{(I)}_i \circ \mathbf{e}^{(J)}_j\right) = \text{vec}\left[\mathbf{e}^{(I)}_i (\mathbf{e}^{(J)}_j)^T\right]$, where the symbol $\circ$ denotes the outer product (see section 6.3).

FACT 4.7.– Since the operator vec satisfies $\text{vec}(\alpha\,\mathbf{A} + \beta\,\mathbf{B}) = \alpha\ \text{vec}(\mathbf{A}) + \beta\ \text{vec}(\mathbf{B})$ for all $\alpha, \beta \in \mathbb{K}$, it is linear.

## 4.6. Vector inner product, norm and orthogonality

### 4.6.1. *Inner product*

In this section, we recall the definition of the inner product (also called dot or scalar product) of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{K}^I$, denoted by $\langle \mathbf{a}, \mathbf{b} \rangle$ or $\mathbf{b}^T\mathbf{a}$ if $\mathbb{K} = \mathbb{R}$, and $\mathbf{b}^H\mathbf{a}$ if $\mathbb{K} = \mathbb{C}$. This binary operation satisfies the properties of an Euclidean inner product, as introduced in section 3.4.1.1. In $\mathbb{R}^I$, it is defined as:

$$\langle \cdot, \cdot \rangle : \mathbb{R}^I \times \mathbb{R}^I \to \mathbb{R}\,, \quad (\mathbf{a}, \mathbf{b}) \mapsto \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{b}^T\mathbf{a} = \sum_{i=1}^{I} a_i\, b_i.$$

It is easy to verify that this operation is a bilinear form.

In $\mathbb{C}^I$, the definition of the inner product is given by:

$$\langle \cdot, \cdot \rangle : \mathbb{C}^I \times \mathbb{C}^I \to \mathbb{C}, \quad (\mathbf{a}, \mathbf{b}) \mapsto \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{b}^H\mathbf{a} = \sum_{i=1}^{I} a_i\, b_i^*.$$

In this case, the inner product is a sesquilinear form, which means (see section 3.4.2.1):

– linearity with respect to the first argument:

$$\langle \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2, \mathbf{b} \rangle = \alpha_1 \langle \mathbf{a}_1, \mathbf{b} \rangle + \alpha_2 \langle \mathbf{a}_2, \mathbf{b} \rangle \quad \text{for all } \alpha_1, \alpha_2 \in \mathbb{C};$$

– Hermitian symmetry:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle^*.$$

The inner product in $\mathbb{C}^I$ is said to be semilinear with respect to the second argument, namely:

$$\langle \mathbf{a}, \alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2 \rangle = \alpha_1^* \langle \mathbf{a}, \mathbf{b}_1 \rangle + \alpha_2^* \langle \mathbf{a}, \mathbf{b}_2 \rangle \quad \text{for all } \alpha_1, \alpha_2 \in \mathbb{C}.$$

By equipping $\mathbb{K}^I$ with an inner product, we obtain a Euclidean $\mathbb{R}$-v.s. if $\mathbb{K} = \mathbb{R}$ or Hermitian $\mathbb{C}$-v.s. if $\mathbb{K} = \mathbb{C}$.

### 4.6.2. *Euclidean/Hermitian norm*

The Euclidean (Hermitian) norm of a vector $\mathbf{a}$, denoted $\|\mathbf{a}\|$, associates to $\mathbf{a} \in \mathbb{R}^I$ ($\mathbf{a} \in \mathbb{C}^I$) a non-negative real number according to the following definition:

$$\begin{aligned} \|\cdot\|_2 : \mathbb{K}^I &\to \mathbb{R}^+ \\ \mathbf{a} &\mapsto \|\mathbf{a}\|_2 = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}. \end{aligned}$$

The properties of the inner product guarantee that $\|\mathbf{a}\|_2 = 0$ if and only if $\mathbf{a} = \mathbf{0}$.

This quantity has a natural geometric interpretation, representing the length of a vector in $\mathbb{K}^I$, that is, the distance between the origin of $\mathbb{K}^I$ and the point whose coordinates are $(a_1, \ldots, a_I)$. For example, in $\mathbb{R}^2$, we have $\|\mathbf{a}\|_2^2 = \langle \mathbf{a}, \mathbf{a} \rangle = a_1^2 + a_2^2$, which is consistent with the Pythagorean theorem. This fact is illustrated in Figure 4.1, where the vector $\mathbf{a} = [3 \;\; 4]^T \in \mathbb{R}^2$ is represented by an arrow from the origin of the coordinate system and directed to the point $(3, 4)$; the first coordinate is associated with the $x$-axis while the second is associated with the $y$-axis. Segments $P$, $Q$, and $R$ form a right triangle whose hypotenuse is $R$. Denoting by $|S|$ the length of segment $S$, the Pythagorean theorem gives $|R|^2 = |P|^2 + |Q|^2 = a_1^2 + a_2^2 = 3^2 + 4^2 = 25 = \|\mathbf{a}\|_2^2$.

### 4.6.3. *Orthogonality*

As discussed in Chapter 3, two vectors $\mathbf{a}$ and $\mathbf{b}$ of $\mathbb{K}^I$ are said to be orthogonal if and only if $\langle \mathbf{a}, \mathbf{b} \rangle = 0$. This terminology comes from the fact that the inner product in the Euclidean space $\mathbb{R}^I$ satisfies:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \|\mathbf{a}\|_2 \|\mathbf{b}\|_2 \cos\theta,$$

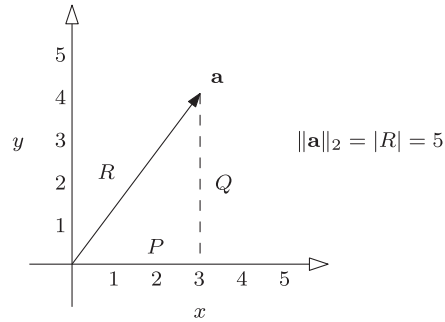



**Figure 4.1.** *Geometric interpretation of the norm of* $\mathbf{a} = [3 \;\; 4]^T \in \mathbb{R}^2$

where $\theta$ is the angle between vectors $\mathbf{a}$ and $\mathbf{b}$. Figure 4.2 provides an example in the case of $\mathbb{R}^2$, where the vectors are $\mathbf{a} = [-3 \;\; 3]^T$ and $\mathbf{b} = [2 \;\; 2]^T$. It is easy to see that $\langle \mathbf{a}, \mathbf{b} \rangle = a_1 b_1 + a_2 b_2 = -6 + 6 = 0$, and vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal. Geometrically, this amounts to saying that the angle between these two vectors is $\theta = \frac{\pi}{2}$ radians, which is visible in the figure. From the definition, it can be concluded that a vector $\mathbf{a} \in \mathbb{K}^I$ is orthogonal to any other vector $\mathbf{b} \in \mathbb{K}^I$ if and only if $\mathbf{a} = \mathbf{0}$.

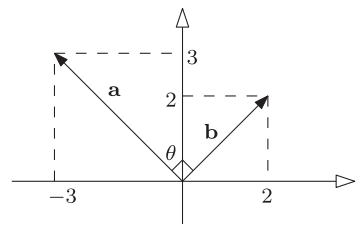



**Figure 4.2.** *Geometric representation of orthogonality between vectors* $\mathbf{a} = [-3 \;\; 3]^T$ *and* $\mathbf{b} = [2 \;\; 2]^T$ *in* $\mathbb{R}^2$

## 4.7. Matrix multiplication

### 4.7.1. *Definition and properties*

Given matrices $\mathbf{B} \in \mathbb{K}^{I \times J}$ and $\mathbf{A} \in \mathbb{K}^{J \times K}$, the product of $\mathbf{B}$ by $\mathbf{A}$ gives a matrix $\mathbf{C} = \mathbf{BA} \in \mathbb{K}^{I \times K}$ such that $c_{ik} = \sum_{j=1}^{J} b_{ij} a_{jk}$, for $i \in \langle I \rangle; k \in \langle K \rangle$. It is said that $\mathbf{A}$ is pre-multiplied by $\mathbf{B}$ and $\mathbf{B}$ is post-multiplied by $\mathbf{A}$.

This product can be written in terms of the outer products of column vectors of $\mathbf{B}$ with transposed row vectors of $\mathbf{A}$:

$$\mathbf{C} = \sum_{j=1}^{J} \mathbf{B}_{.j} \circ \mathbf{A}_{j.}^{T} = \sum_{j=1}^{J} \mathbf{B}_{.j} \mathbf{A}_{j.}. \qquad [4.3]$$

If $\mathbb{K} = \mathbb{R}$, then the elements of $\mathbf{C}$ are given by the Euclidian inner product of transposed row vectors of $\mathbf{B}$ and column vectors of $\mathbf{A}$, according to $c_{ik} = \langle \mathbf{A}_{.k}, \mathbf{B}_{i.}^{T} \rangle = \mathbf{B}_{i.} \mathbf{A}_{.k}$. Similarly, for $\mathbb{K} = \mathbb{C}$, the Hermitian inner product is used to give: $c_{ik} = \langle \mathbf{A}_{.k}, \mathbf{B}_{i.}^{H} \rangle = \mathbf{B}_{i.} \mathbf{A}_{.k}$.

FACT 4.8.– From equation [4.3], it can be deduced that row $i$ and column $j$ of $\mathbf{C}$ are given by $\mathbf{C}_{i.} = [\mathbf{BA}]_{i.} = \sum_{j=1}^{J} b_{ij} \mathbf{A}_{j.}$ and $\mathbf{C}_{.k} = [\mathbf{BA}]_{.k} = \sum_{j=1}^{J} a_{jk} \mathbf{B}_{.j}$, which shows that the rows of $\mathbf{BA}$ are linear combinations of the rows of $\mathbf{A}$, while the columns of $\mathbf{BA}$ are linear combinations of the columns of $\mathbf{B}$.

We also have:

$$\mathbf{C}_{i.} = \mathbf{B}_{i.} \mathbf{A} \ , \ \mathbf{C}_{.k} = \mathbf{B} \mathbf{A}_{.k}. \qquad [4.4]$$

PROPOSITION 4.9.– *The matrix product satisfies the following properties:*

– *The product is not commutative, since in general*[5] $\mathbf{BA} \neq \mathbf{AB}$.

– *The product is associative:*

$$(\mathbf{BA})\mathbf{C} = \mathbf{B}(\mathbf{AC}) = \mathbf{BAC}. \qquad [4.5a]$$

– *The product is distributive over the addition:*

$$\mathbf{B}(\mathbf{A} + \mathbf{C}) = \mathbf{BA} + \mathbf{BC} \ , \ (\mathbf{B} + \mathbf{D})\mathbf{C} = \mathbf{BC} + \mathbf{DC}. \qquad [4.5b]$$

– *The product is associative over scalar multiplication:*

$$\mathbf{B}(\lambda \mathbf{A}) = (\lambda \mathbf{B})\mathbf{A} = \lambda(\mathbf{BA}), \quad \forall \lambda \in \mathbb{K}. \qquad [4.5c]$$

– *The transpose and conjugate transpose of a product of matrices are such that:*

$$(\mathbf{BA})^{T} = \mathbf{A}^{T} \mathbf{B}^{T}, \quad (\mathbf{BA})^{H} = \mathbf{A}^{H} \mathbf{B}^{H}. \qquad [4.5d]$$

It should be noted that the notation $\mathbf{b}^{T}\mathbf{a}$ (or $\mathbf{b}^{H}\mathbf{a}$ in $\mathbb{C}$) used for the inner product $\langle \mathbf{a}, \mathbf{b} \rangle$ derives from the aforementioned definition of the matrix product. Indeed, when it is applied to the case where $I = K = 1$, that is, the product between a row vector $\mathbf{b}^{T}$

5 When matrices $\mathbf{A}$ and $\mathbf{B}$ are not square (but have a dimension in common), one of the expressions $\mathbf{BA}$ and $\mathbf{AB}$ is not defined. For the product $\mathbf{AB}$, it is needed that the number of columns of $\mathbf{A}$ be equal to the number of rows of $\mathbf{B}$. Then, $\mathbf{A}$ and $\mathbf{B}$ are said to be conformable for multiplication.

(or $\mathbf{b}^H$ in $\mathbb{C}$) and a column vector $\mathbf{a}$ such that $\mathbf{a}, \mathbf{b} \in \mathbb{R}^J$ (or $\mathbb{C}^J$) , we get $\mathbf{b}^T\mathbf{a} = \sum_{j=1}^J a_j b_j = \langle \mathbf{a}, \mathbf{b} \rangle$ in the real case, and $\mathbf{b}^H\mathbf{a} = \sum_{j=1}^J a_j b_j^* = \langle \mathbf{a}, \mathbf{b} \rangle$ in the complex case.

In a similar way, if the definition is applied to the case where $J = 1$, that is, to the product $\mathbf{b}\mathbf{a}^T$ between a column vector $\mathbf{b}$ and a row vector $\mathbf{a}^T$ such that $\mathbf{b} \in \mathbb{K}^I$ and $\mathbf{a} \in \mathbb{K}^K$, we then obtain:

$$\left(\mathbf{b}\mathbf{a}^T\right)_{ik} = b_i a_k = (\mathbf{b} \circ \mathbf{a})_{ik} \,,$$

which justifies the notation $\mathbf{b}\mathbf{a}^T$ often used for the outer product of $\mathbf{b}$ and $\mathbf{a}$.

EXAMPLE 4.10.– We provide here two examples of trilinear and bilinear expressions in the form of matrix products. For $\mathbf{A} \in \mathbb{K}^{I \times J}$, $\mathbf{B} \in \mathbb{K}^{J \times K}$, and $\mathbf{C} \in \mathbb{K}^{K \times L}$, we have:

$$\sum_{j=1}^{J} \sum_{k=1}^{K} a_{ij} b_{jk} c_{kl} = [\mathbf{ABC}]_{il},$$

$$\sum_{j=1}^{J} \sum_{k=1}^{K} a_{ij} c_{kl}^* = [\mathbf{ABC}^*]_{il}, \quad \text{with} \quad \mathbf{B} = \mathbf{1}_J \mathbf{1}_K^T.$$

### 4.7.2. *Powers of a matrix*

The $n$th power of a square matrix is defined for every natural integer $n$ as follows:

$$\mathbf{A}^n = \begin{cases} \mathbf{I}, & n = 0, \\ \mathbf{A}, & n = 1, \\ \mathbf{A}\mathbf{A}^{n-1}, & n > 1, \end{cases}$$

thus satisfying $\mathbf{A}^n\mathbf{A}^p = \mathbf{A}^{n+p}$ and $(\mathbf{A}^n)^p = \mathbf{A}^{np}$ for all $n, p \in \mathbb{N}$. It is easy to also show that the $n$th power of a transposed matrix is equal to the transpose of the $n$th power of the matrix: $\left(\mathbf{A}^T\right)^n = \left(\mathbf{A}^n\right)^T$.

PROPOSITION 4.11 (Binomial theorem).– *For $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{I \times I}$, we have:*

$$(\mathbf{A} + \mathbf{B})^n = \sum_{q=0}^{n} C_n^q \, \mathbf{A}^{n-q} \mathbf{B}^q, \quad \textit{where} \quad C_n^q = \frac{n!}{q!(n-q)!}.$$

In Table 4.3, we define three particular classes of matrices from the properties satisfied by $\mathbf{A}^n$.

| Matrices | | Conditions |
|---|---|---|
| $\mathbf{A}$ periodic of period $n \geq 1$ | If | $\mathbf{A}^{n+1} = \mathbf{A}$ |
| $\mathbf{A}$ nilpotent of degree (or index) $n \geq 1$ | If | $\mathbf{A}^{n-1} \neq \mathbf{0}$ and $\mathbf{A}^n = \mathbf{0}$ |
| $\mathbf{A}$ idempotent | If | $\mathbf{A}^2 = \mathbf{A}$ |

**Table 4.3.** *Periodic/Nilpotent/Idempotent matrices*

PROPOSITION 4.12.– *If $\mathbf{A}$ is nilpotent, then the trace[6] of $\mathbf{A}$ is zero, that is,* $\mathrm{tr}(\mathbf{A}) = 0$.

EXAMPLE 4.13.– The matrix $\mathbf{x}\mathbf{y}^T$, with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$, is nilpotent of degree 2 if and only if vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. Indeed, we have: $(\mathbf{x}\mathbf{y}^T)^2 = \mathbf{x}(\mathbf{y}^T\mathbf{x})\mathbf{y}^T = \mathbf{0}$ if and only if $\mathbf{y}^T\mathbf{x} = 0$. The same property is true for $\mathbf{x}\mathbf{y}^H$, with $\mathbf{x}, \mathbf{y} \in \mathbb{C}^I$.

EXAMPLE 4.14.– Any strictly upper (or lower) triangular matrix of order $n$ is nilpotent of degree $k \leq n$. Thus, upper ($\mathbf{M}_n$) and lower ($\mathbf{D}_n$) shift matrices[7] of order $n$:

$$\mathbf{M}_n = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}, \quad \mathbf{D}_n = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ 1 & \ddots & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} = \mathbf{M}_n^T,$$

are nilpotent of degree $n$, because $\mathbf{M}_n^k \neq \mathbf{0}$ and $\mathbf{D}_n^k \neq \mathbf{0}$, $\forall k \in \langle n-1 \rangle$, and $\mathbf{M}_n^n = \mathbf{D}_n^n = \mathbf{0}$. The shift matrices can also be defined as $m_{ij} = [\delta_{i+1,j}]$ and $d_{ij} = [\delta_{i,j+1}]$, where $\delta_{i,j}$ is the Kronecker delta.

Pre-multiplication of a vector by $\mathbf{M}_n$ corresponds to an upper shift of its components:

$$\mathbf{M}_n^k \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{k+1} \\ \vdots \\ x_n \\ \mathbf{0}_k \end{bmatrix}, \quad \forall k \in \langle n-1 \rangle \; ; \quad \mathbf{M}_n^k \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{0}_n, \quad \forall k \geq n.$$

Similarly, pre-multiplication of a vector by $\mathbf{D}_n$ corresponds to a lower shift of its components.

6 The matrix trace will be defined in section 4.8.1.

7 A shift matrix is a binary matrix with 1s on the superdiagonal (upper shift) or subdiagonal (lower shift) and 0s elsewhere.

EXAMPLE 4.15.– Computation of powers of a unit triangular matrix $\mathbf{A} = \mathbf{I}_n + \mathbf{B}$, where $\mathbf{B}$ is a strictly upper (or lower) triangular square matrix of order $n$.

$\mathbf{B}$ being a nilpotent matrix of degree $k \leq n$, we have $\mathbf{B}^i = \mathbf{0}$ for all $i \geq k$, and therefore the application of the binomial theorem gives:

$$\mathbf{A}^p = (\mathbf{I}_n + \mathbf{B})^p = \sum_{i=0}^{k-1} \frac{p!}{(p-i)!\, i!} \mathbf{B}^i.$$

For instance:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \Rightarrow \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

We have $\mathbf{B}^i = \mathbf{0}, \forall i \geq 3$, and therefore:

$$\mathbf{A}^p = \mathbf{I}_3 + p\mathbf{B} + \frac{p(p-1)}{2}\mathbf{B}^2 = \begin{bmatrix} 1 & 0 & 0 \\ p & 1 & 0 \\ \frac{p(p+1)}{2} & p & 1 \end{bmatrix}.$$

EXAMPLE 4.16.– Application to the decomposition of a lower triangular Toeplitz matrix in the form of a matrix polynomial:

$$\mathbf{A} = \begin{bmatrix} t_0 & 0 & \cdots & \cdots & 0 \\ t_1 & t_0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ t_{n-1} & \cdots & \cdots & t_1 & t_0 \end{bmatrix}.$$

Such a matrix is completely specified by its first column:

$$\mathbf{A}_{.1} = [t_0, \cdots, t_{n-1}]^T,$$

the other columns are obtained by simple down shifts of $\mathbf{A}_{.1}$, namely, $\mathbf{A}_{.j} = \mathbf{D}_n^{j-1}\mathbf{A}_{.1}$, $j \in \{2, \cdots, n\}$. It can also be written as:

$$\mathbf{A} = t(\mathbf{D}_n) = \sum_{j=0}^{n-1} t_j \mathbf{D}_n^j.$$

This writing in the form of a matrix polynomial of degree $n-1$ in the variable $\mathbf{D}_n$, the $n$th-order lower shift matrix , allows to demonstrate the property stated as follows (Pollock 1999).

PROPOSITION 4.17.– *If $\mathbf{A}$ and $\mathbf{B}$ are lower triangular Toeplitz matrices of order $n$, admitting as respective polynomial representations $t(\mathbf{D}_n)$ and $r(\mathbf{D}_n)$, their product $\mathbf{AB}$ is commutative and gives a lower triangular Toeplitz matrix.*

This property derives from the commutativity of the product of two polynomials and of the nilpotence of $\mathbf{D}_n$.

This result can be used to calculate the product of two polynomials $r(z)$ and $t(z)$, of respective degrees $m$ and $n$, through the product of two lower triangular Toeplitz matrices $\mathbf{A}$ and $\mathbf{B}$, of order $m+n+1$. The first column of $\mathbf{C}=\mathbf{AB}$ contains the coefficients of the polynomial $p(z)=r(z)t(z)$. Explicitly, the product of polynomials $p(z)=r(z)t(z)=(\sum_{k=0}^{m} r_k z^k)(\sum_{k=0}^{n} t_k z^k)=\sum_{k=0}^{m+n} p_k z^k$ can be achieved by means of a matrix product, or simply a matrix - column vector product.

EXAMPLE 4.18.– For $m=1, n=2$, we have:

$$\begin{bmatrix} p_0 & 0 & 0 & 0 \\ p_1 & p_0 & 0 & 0 \\ p_2 & p_1 & p_0 & 0 \\ p_3 & p_2 & p_1 & p_0 \end{bmatrix} = \begin{bmatrix} r_0 & 0 & 0 & 0 \\ r_1 & r_0 & 0 & 0 \\ 0 & r_1 & r_0 & 0 \\ 0 & 0 & r_1 & r_0 \end{bmatrix} \begin{bmatrix} t_0 & 0 & 0 & 0 \\ t_1 & t_0 & 0 & 0 \\ t_2 & t_1 & t_0 & 0 \\ 0 & t_2 & t_1 & t_0 \end{bmatrix},$$

which gives $p_0=r_0t_0$, $p_1=r_0t_1+r_1t_0$, $p_2=r_0t_2+r_1t_1$, and $p_3=r_1t_2$.

## 4.8. Matrix trace, inner product and Frobenius norm

### 4.8.1. *Definition and properties of the trace*

The trace of a square matrix $\mathbf{A}$ of order $I$ is defined as the sum of its diagonal elements: $\mathrm{tr}(\mathbf{A})=\sum_{i=1}^{I} a_{ii}$.

PROPOSITION 4.19.– *The trace satisfies the following properties:*

$$\mathrm{tr}(\alpha\mathbf{A}+\beta\mathbf{B}) = \alpha\ \mathrm{tr}(\mathbf{A}) + \beta\ \mathrm{tr}(\mathbf{B}), \quad [4.6a]$$

$$\mathrm{tr}(\mathbf{A}^T) = \mathrm{tr}(\mathbf{A}), \quad [4.6b]$$

$$\mathrm{tr}(\mathbf{A}^*) = \mathrm{tr}(\mathbf{A}^H) = [\mathrm{tr}(\mathbf{A})]^*, \quad [4.6c]$$

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}) = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}b_{ji}, \quad [4.6d]$$

$$\mathrm{tr}(\mathbf{AC}^T) = \mathrm{tr}(\mathbf{C}^T\mathbf{A}) = \mathrm{tr}(\mathbf{A}^T\mathbf{C}) = \mathrm{tr}(\mathbf{CA}^T) = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}c_{ij}, \quad [4.6e]$$

$$\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB}) = \mathrm{tr}(\mathbf{BCA}). \quad [4.6f]$$

Property [4.6f] is called the cyclic invariance of the trace, and it can be generalized to the product of any number of matrices. Note that, in general, $\mathrm{tr}(\mathbf{ABC}) \neq \mathrm{tr}(\mathbf{ACB})$ and $\mathrm{tr}(\mathbf{ABC}) \neq \mathrm{tr}(\mathbf{BAC})$.

EXAMPLE 4.20.– Expression of a bilinear form and a quadratic form in terms of a matrix trace: from the relation $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ and from the fact that the trace of a scalar is the scalar itself, it can be deduced that:

$$\mathbf{y}^H\mathbf{x} = \sum_{i=1}^{I} y_i^* x_i = \mathrm{tr}(\mathbf{y}^H\mathbf{x}) = \mathrm{tr}(\mathbf{x}\mathbf{y}^H),$$

$$\mathbf{y}^H\mathbf{A}\mathbf{x} = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij} y_i^* x_j = \mathrm{tr}(\mathbf{y}^H\mathbf{A}\mathbf{x}) = \mathrm{tr}(\mathbf{x}\mathbf{y}^H\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{x}\mathbf{y}^H).$$

### 4.8.2. *Matrix inner product*

Similarly to the v.s. $\mathbb{K}^I$, the v.s. $\mathbb{K}^{I\times J}$ can be equipped with an inner product in order to give it a Euclidean space structure if $\mathbb{K} = \mathbb{R}$, and a Hermitian one if $\mathbb{K} = \mathbb{C}$. This inner product is defined by:

$$\mathbb{R}^{I\times J} \times \mathbb{R}^{I\times J} \to \mathbb{R}$$

$$(\mathbf{A}, \mathbf{B}) \mapsto \langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{tr}\left(\mathbf{B}^T\mathbf{A}\right) = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij} b_{ij}$$

in the real case and by:

$$\mathbb{C}^{I\times J} \times \mathbb{C}^{I\times J} \to \mathbb{C}$$

$$(\mathbf{A}, \mathbf{B}) \mapsto \langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{tr}\left(\mathbf{B}^H\mathbf{A}\right) = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij} b_{ij}^*$$

in the complex case. The following identities can be easily verified:

$$\mathrm{tr}(\mathbf{B}^T\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{B}^T) = \langle \mathrm{vec}(\mathbf{A}), \mathrm{vec}(\mathbf{B}) \rangle = \mathrm{vec}^T(\mathbf{B})\,\mathrm{vec}(\mathbf{A})$$

in $\mathbb{R}$ and their analogues in $\mathbb{C}$:

$$\mathrm{tr}(\mathbf{B}^H\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{B}^H) = \langle \mathrm{vec}(\mathbf{A}), \mathrm{vec}(\mathbf{B}) \rangle = \mathrm{vec}^H(\mathbf{B})\,\mathrm{vec}(\mathbf{A}).$$

### 4.8.3. *Frobenius norm*

The previous definition of the matrix inner product induces the Euclidean/Hermitian matrix norm as:

$$\|\cdot\|_F : \mathbb{K}^{I\times J} \to \mathbb{R}^+$$

$$\mathbf{A} \mapsto \|\mathbf{A}\|_F = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J} |a_{ij}|^2}.$$

This norm is called the Frobenius norm, hence the letter "F" in its notation. It is also called the Hilbert–Schmidt norm or the Schur norm.

A natural link can be found between this norm and the Euclidean/Hermitian norm of a vector:

$$\mathrm{tr}(\mathbf{A}\mathbf{A}^H) = \mathrm{tr}(\mathbf{A}^H\mathbf{A}) = \sum_{i,j=1}^{I} |a_{ij}|^2 = \sum_{i=1}^{I} \|\mathbf{A}_{i.}\|_2^2 = \sum_{j=1}^{J} \|\mathbf{A}_{.j}\|_2^2 = \|\,\mathrm{vec}(\mathbf{A})\|_2^2.$$

## 4.9. Subspaces associated with a matrix

Given the matrix $\mathbf{A} \in \mathbb{K}^{I\times J}$, its column space, denoted $C(\mathbf{A})$, is the subspace of $\mathbb{K}^I$ spanned[8] by the column vectors of $\mathbf{A}$:

$$C(\mathbf{A}) = \mathrm{lc}(\mathbf{A}_{.1}, \cdots, \mathbf{A}_{.J}).$$

The dimension of the space $C(\mathbf{A})$, that is, the maximal number of linearly independent column vectors of $\mathbf{A}$, is called the column rank of $\mathbf{A}$.

Similarly, the column space of $\mathbf{A}^T$, denoted $C(\mathbf{A}^T)$, is the subspace of $\mathbb{K}^J$ spanned by the column vectors of $\mathbf{A}^T$, in other words, the row vectors of $\mathbf{A}$. This space is called the row space of $\mathbf{A}$ and is denoted by $L(\mathbf{A})$. It is such that:

$$\begin{aligned} L(\mathbf{A}) &= \mathrm{lc}(\mathbf{A}_{1.}, \cdots, \mathbf{A}_{I.}) \\ &= C(\mathbf{A}^T). \end{aligned}$$

The dimension of this space, that is, the maximal number of linearly independent rows of $\mathbf{A}$, is called the row rank of $\mathbf{A}$.

As discussed in section 4.13, $\mathbf{A}$ can be seen as the matrix of the linear map $\mathcal{L} : \mathbb{K}^J \to \mathbb{K}^I$ such that $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$. The image space of the map, denoted by $\mathrm{Im}(\mathbf{A})$, can then be defined as:

$$\mathrm{Im}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{K}^I : \exists \mathbf{x} \in \mathbb{K}^J \text{ such that } \mathbf{y} = \mathbf{A}\mathbf{x}\},$$

that is, the space formed of all linear combinations of the column vectors of $\mathbf{A}$ or equivalently the subspace of $\mathbb{K}^I$ spanned by the columns of $\mathbf{A}$. This subspace is also called the range of $\mathbf{A}$. We thus have:

$$C(\mathbf{A}) = \mathrm{Im}(\mathbf{A}) \subseteq \mathbb{K}^I.$$

8 The definition of a space spanned by a set of vectors, as well as the notation lc, is given in section 2.5.12.2.

Similarly, we have:

$$C(\mathbf{A}^T) = \mathrm{Im}(\mathbf{A}^T) = \{\mathbf{x} \in \mathbb{K}^J : \exists \mathbf{y} \in \mathbb{K}^I \text{ such that } \mathbf{x} = \mathbf{A}^T\mathbf{y}\} \subseteq \mathbb{K}^J.$$

The kernel of $\mathbf{A}$, also called the null space of $\mathbf{A}$ and denoted by $\mathcal{N}(\mathbf{A})$, is defined as:

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{K}^J : \mathbf{A}\mathbf{x} = \mathbf{0}_I\} \subseteq \mathbb{K}^J.$$

This is the subspace of $\mathbb{K}^J$ formed of elements whose image is equal to the null vector of $\mathbb{K}^I$. The kernel $\mathcal{N}(\mathbf{A})$ can therefore be viewed as the set of all solutions of the homogeneous system of equations $\mathbf{A}\mathbf{x} = \mathbf{0}_I$. Similarly, the kernel of $\mathbf{A}^T$ can be defined as:

$$\mathcal{N}(\mathbf{A}^T) = \{\mathbf{y} \in \mathbb{K}^I : \mathbf{A}^T\mathbf{y} = \mathbf{0}_J\} \subseteq \mathbb{K}^I.$$

$\mathcal{N}(\mathbf{A}^T)$ is also called the left-hand kernel (or left nullspace) of $\mathbf{A}$ because it is the set of all solutions to the left-hand homogeneous system $\mathbf{y}^T\mathbf{A} = \mathbf{0}_J^T$. It is worth noting that, in the case of complex matrices ($\mathbb{K} = \mathbb{C}$), $\mathbf{A}^T$ should be replaced by $\mathbf{A}^H$ in the previous definitions.

In summary, with any matrix $\mathbf{A}$ we can associate the four fundamental subspaces given in Table 4.4.

| Subspaces | Notations |
|---|---|
| Column space | $C(\mathbf{A}) = \mathrm{Im}(\mathbf{A}) \subseteq \mathbb{K}^I$ |
| Row space | $L(\mathbf{A}) = C(\mathbf{A}^T) \subseteq \mathbb{K}^J$ |
| Kernel | $\mathcal{N}(\mathbf{A}) \subseteq \mathbb{K}^J$ |
| Left-hand kernel | $\mathcal{N}(\mathbf{A}^T) \subseteq \mathbb{K}^I$ |

**Table 4.4.** *Subspaces associated with a matrix*

The spaces $C(\mathbf{A})$ and $\mathcal{N}(\mathbf{A})$ are linked by the relationships that constitute the rank theorem, also called fundamental theorem of linear algebra, as stated in the following.

THEOREM 4.21.– *Any matrix* $\mathbf{A} \in \mathbb{K}^{I \times J}$ *satisfies:*

$$\dim[C(\mathbf{A})] + \dim[\mathcal{N}(\mathbf{A})] = J, \qquad [4.7a]$$

*or equivalently:*

$$\dim[L(\mathbf{A})] + \dim[\mathcal{N}(\mathbf{A}^T)] = I. \qquad [4.7b]$$

In Table 4.5, we present the relations linking the subspaces $C(\mathbf{A})$, $L(\mathbf{A})$, $\mathcal{N}(\mathbf{A})$, and $\mathrm{Im}(\mathbf{A})$. The spaces $[C(\mathbf{A})]^{\perp}$, $[L(\mathbf{A})]^{\perp}$, and $[\mathcal{N}(\mathbf{A})]^{\perp}$ are the orthogonal complements of $C(\mathbf{A})$, $L(\mathbf{A})$, and $\mathcal{N}(\mathbf{A})$, respectively.

| | |
|---|---|
| $C(\mathbf{A}) = [\mathcal{N}(\mathbf{A}^T)]^{\perp} = \mathrm{Im}(\mathbf{A})$ | $C(\mathbf{A}^T) = [\mathcal{N}(\mathbf{A})]^{\perp} = L(\mathbf{A})$ |
| $C(\mathbf{A}) \cap \mathcal{N}(\mathbf{A}^T) = \{\mathbf{0}_I\}$ | $\mathcal{N}(\mathbf{A}) \cap C(\mathbf{A}^T) = \{\mathbf{0}_J\}$ |
| $\dim[C(\mathbf{A})] + \dim[\mathcal{N}(\mathbf{A})] = J$ | |
| $\dim[L(\mathbf{A})] + \dim[\mathcal{N}(\mathbf{A}^T)] = I$ | |

**Table 4.5.** *Relations linking subspaces $C(\mathbf{A})$, $L(\mathbf{A})$, and $\mathcal{N}(\mathbf{A})$*

The equalities $C(\mathbf{A}) \cap \mathcal{N}(\mathbf{A}^T) = \{\mathbf{0}_I\}$ and $\mathcal{N}(\mathbf{A}) \cap C(\mathbf{A}^T) = \{\mathbf{0}_J\}$ arise from the fact that the intersection of two orthogonal subspaces is made up of the null vector.

From the equalities of this table, it can be deduced that any matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ provides an orthogonal decomposition of the spaces $\mathbb{K}^I$ and $\mathbb{K}^J$ such as:

$$\mathbb{K}^I = C(\mathbf{A}) \oplus [C(\mathbf{A})]^{\perp} = C(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^T) \quad [4.8a]$$

$$\mathbb{K}^J = \mathcal{N}(\mathbf{A}) \oplus [N(\mathbf{A})]^{\perp} = \mathcal{N}(\mathbf{A}) \oplus C(\mathbf{A}^T). \quad [4.8b]$$

## 4.10. Matrix rank

### 4.10.1. *Definition and properties*

A basic result of linear algebra states that the column rank and row rank of a matrix are equal, which leads to the following proposition.

PROPOSITION 4.22.– *Any matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ is such that:*

$$\dim[\mathrm{C}(\mathbf{A})] = \dim[\mathrm{L}(\mathbf{A})].$$

From this result, the rank of a matrix can be defined as:

$$\mathrm{r}(\mathbf{A}) = \dim[C(\mathbf{A})] = \dim[L(\mathbf{A})], \quad [4.9]$$

implying that the rank is equal to the maximal number of linearly independent columns or rows, such that for any matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$, we have:

$$\mathrm{r}(\mathbf{A}) \leq \min(I, J)$$

The properties of a matrix related to its rank are summarized in Table 4.6.

| Definitions | Conditions |
|---|---|
| Deficient rank | $\mathrm{r}(\mathbf{A}) < \min(I, J)$ |
| Full column rank | $\mathrm{r}(\mathbf{A}) = J \leq I$ |
| Full row rank | $\mathrm{r}(\mathbf{A}) = I \leq J$ |
| Full rank | $\mathrm{r}(\mathbf{A}) = \min(I, J)$ |

**Table 4.6.** *Properties related to the rank of* $\mathbf{A} \in \mathbb{K}^{I \times J}$

In the case of a square matrix $\mathbf{A} \in \mathbb{K}^{I \times I}$ of full rank (i.e. of rank equal to $I$), $\mathbf{A}$ is said to be non-singular or regular. If it is of deficient rank (i.e. when $\mathrm{r}(A) < I$), then it is said to be singular.

Different ways to define the rank are described[9] in Table 4.7.

| |
|---|
| Maximal number of independent columns |
| Maximal number of independent rows |
| Order of the largest non-singular submatrix |
| $\mathrm{r}(\mathbf{A}) = \dim[C(\mathbf{A})] = \dim[\mathrm{Im}(\mathbf{A})]$ |
| $\mathrm{r}(\mathbf{A}) = \dim[L(\mathbf{A})] = \dim[C(\mathbf{A}^T)]$ |
| Rank theorem: $\begin{cases} \mathrm{r}(\mathbf{A}) = J - \dim[\mathcal{N}(\mathbf{A})] \\ \mathrm{r}(\mathbf{A}) = I - \dim[\mathcal{N}(\mathbf{A}^T)] \end{cases}$ |

**Table 4.7.** *Different ways of defining the rank of* $\mathbf{A} \in \mathbb{K}^{I \times J}$

The rank theorem implies that the rank of a matrix is equal to the number of its columns (rows) subtracted by the maximal number of linearly independent vectors of its kernel (left-hand kernel). This result is similar to the rank theorem of a linear map $f \in \mathcal{L}(E, F)$ with the following correspondences (see equation [2.18]):

$$(f, \mathrm{r}(f), \dim(E), \mathrm{Ker}(f)) \leftrightarrow (\mathbf{A}, \mathrm{r}(\mathbf{A}), J, \mathcal{N}(\mathbf{A})).$$

FACT 4.23.– From the rank theorem, we can conclude that $\mathbf{A}$ is full column rank if and only if $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}_J\}$ and full row rank if $\mathcal{N}(\mathbf{A}^T) = \{\mathbf{0}_I\}$.

PROPOSITION 4.24.– *For* $\mathbf{A} \in \mathbb{C}^{I \times J}$*, we have the following properties:*

$$\mathrm{r}(\mathbf{A}^T) = \mathrm{r}(\mathbf{A}^*) = \mathrm{r}(\mathbf{A}^H) = \mathrm{r}(\mathbf{A}).$$

9 The notion of submatrix is defined in section 5.2.

### 4.10.2. *Sum and difference rank*

Some relations linking the rank of the sum $\mathbf{A}+\mathbf{B}$ and of the difference $\mathbf{A}-\mathbf{B}$ of two matrices to the ranks of $\mathbf{A}$ and $\mathbf{B}$ are given in Table 4.8.

| Operations | Ranks |
|---|---|
| Sum | $\mathrm{r}(\mathbf{A}+\mathbf{B}) \leq \mathrm{r}(\mathbf{A})+\mathrm{r}(\mathbf{B})$ |
| Difference | $\mathrm{r}(\mathbf{A}-\mathbf{B}) \leq \lvert \mathrm{r}(\mathbf{A})-\mathrm{r}(\mathbf{B}) \rvert$ |

**Table 4.8.** *Rank of a sum and difference of matrices*

### 4.10.3. *Subspaces associated with a matrix product*

Some properties related to the subspaces associated with a matrix product are summarized in Table 4.9.

| Matrices | Relations |
|---|---|
| $\mathbf{A}_1 \in \mathbb{K}^{I\times J}, \mathbf{A}_2 \in \mathbb{K}^{J\times K}$ | $C(\mathbf{A}_1\mathbf{A}_2) \subseteq C(\mathbf{A}_1)$ |
| $\mathbf{A}_1 \in \mathbb{K}^{I\times J}, \mathbf{A}_2 \in \mathbb{K}^{J\times K}$ | $L(\mathbf{A}_1\mathbf{A}_2) \subseteq L(\mathbf{A}_2)$ |
| $\mathbf{A}_1 \in \mathbb{K}^{I\times J}, \mathbf{A}_2 \in \mathbb{K}^{J\times K}$ | $\mathcal{N}(\mathbf{A}_2) \subseteq \mathcal{N}(\mathbf{A}_1\mathbf{A}_2)$ |
| $\mathbf{A}_1 \in \mathbb{K}^{I_1\times I_2}, \cdots, \mathbf{A}_N \in \mathbb{K}^{I_N\times I_{N+1}}$ | $C(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_N) \subseteq C(\mathbf{A}_1)$ |
| $\mathbf{A}_1 \in \mathbb{K}^{I_1\times I_2}, \cdots, \mathbf{A}_N \in \mathbb{K}^{I_N\times I_{N+1}}$ | $L(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_N) \subseteq L(\mathbf{A}_N)$ |
| $\mathbf{A}_1 \in \mathbb{K}^{I_1\times I_2}, \cdots, \mathbf{A}_N \in \mathbb{K}^{I_N\times I_{N+1}}$ | $\mathcal{N}(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_N) = \mathrm{Im}(\mathbf{A}_2\mathbf{A}_3\cdots\mathbf{A}_N) \bigcap \mathcal{N}(\mathbf{A}_1)$ |
| $\mathbf{A}_1 \in \mathbb{K}^{I_1\times I_2}, \cdots, \mathbf{A}_N \in \mathbb{K}^{I_N\times I_{N+1}}$ | $\mathcal{N}(\mathbf{A}_N) \subseteq \mathcal{N}(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_N)$ |
| $\mathbf{A} \in \mathbb{K}^{I\times I}$ | $C(\mathbf{A}^k) \subseteq C(\mathbf{A})$ |
| $\mathbf{A} \in \mathbb{K}^{I\times I}$ | $\mathcal{N}(\mathbf{A}) \subseteq \mathcal{N}(\mathbf{A}^k)$ |
| $\mathbf{A} \in \mathbb{K}^{I\times J}$ | $C(\mathbf{A}^H\mathbf{A}) = C(\mathbf{A}^H)\ ,\ \mathcal{N}(\mathbf{A}^H\mathbf{A}) = \mathcal{N}(\mathbf{A})$ |
| $\mathbf{A} \in \mathbb{K}^{I\times J}$ | $C(\mathbf{A}\mathbf{A}^H) = C(\mathbf{A})\ ,\ \mathcal{N}(\mathbf{A}\mathbf{A}^H) = \mathcal{N}(\mathbf{A}^H)$ |

**Table 4.9.** *Subspaces associated with a matrix product*

In the real case ($\mathbf{A} \in \mathbb{R}^{I\times J}$), the conjugate transpose operation must be replaced by the transpose operation.

From this table, it can be concluded that the Gram matrix $\mathbf{A}^H\mathbf{A}$ has the same column space as $\mathbf{A}^H$ and the same kernel as $\mathbf{A} \in \mathbb{C}^{I\times J}$, while the Gram matrix $\mathbf{A}\mathbf{A}^H$ has the same column space as $\mathbf{A}$ and the same kernel as $\mathbf{A}^H$. Consequently, the Gram matrices $\mathbf{A}^H\mathbf{A}$ and $\mathbf{A}\mathbf{A}^H$ have the same rank as $\mathbf{A}$.

### 4.10.4. *Product rank*

In the following proposition, we present a useful upper bound on the rank of the product of two matrices.

PROPOSITION 4.25.– *The rank of the product of two matrices is less than or equal to the minimum between the ranks of the two matrices:*

$$\mathrm{r}(\mathbf{A}\,\mathbf{B}) \leq min\Big(\mathrm{r}(\mathbf{A}),\, \mathrm{r}(\mathbf{B})\Big). \qquad [4.10]$$

*From this result, one can deduce that the rank of the outer product of two non-zero vectors is equal to 1:*

$$\mathrm{r}(\mathbf{a} \circ \mathbf{b}) = \mathrm{r}(\mathbf{a}\mathbf{b}^T) = 1. \qquad [4.11]$$

Some results concerning the rank of a matrix product are given in Table 4.10. We highlight the property of invariance of the rank of $\mathbf{A} \in \mathbb{K}^{I\times J}$ under pre- and/or post-multiplication by a non-singular matrix, that is, for all matrices $\mathbf{B} \in \mathbb{K}^{J\times J}$ and $\mathbf{C} \in \mathbb{K}^{I\times I}$, we have:

$$\left.\begin{array}{l}\mathrm{r}(\mathbf{B}) = J \\ \mathrm{r}(\mathbf{C}) = I\end{array}\right\} \quad \Rightarrow \quad \mathrm{r}(\mathbf{C}\mathbf{A}) = \mathrm{r}(\mathbf{A}\mathbf{B}) = \mathrm{r}(\mathbf{C}\mathbf{A}\mathbf{B}) = \mathrm{r}(\mathbf{A}).$$

This property results from the fact that pre-multiplication (post-multiplication) by a non-singular matrix does not change the number of linearly independent columns (rows) of $\mathbf{A}$. In a more general way, the rank of $\mathbf{A} \in \mathbb{K}^{I\times J}$ does not change when pre-multiplying it by a matrix $\mathbf{C} \in \mathbb{K}^{M\times I}$, with $M \geq I$, of full column rank or when it is post-multiplied by a matrix $\mathbf{B} \in \mathbb{K}^{J\times N}$, with $N \geq J$, of full row rank.

| Properties | Ranks |
|---|---|
| $\mathbf{B}$ and $\mathbf{C}$ non-singular | $\mathrm{r}(\mathbf{C}\mathbf{A}) = \mathrm{r}(\mathbf{A}\mathbf{B}) = \mathrm{r}(\mathbf{C}\mathbf{A}\mathbf{B}) = \mathrm{r}(\mathbf{A})$ |
| $\mathbf{A} \in \mathbb{K}^{I\times J}, \mathbf{B} \in \mathbb{K}^{J\times K}$ | $\mathrm{r}(\mathbf{A}\mathbf{B}) = \mathrm{r}(\mathbf{B}) - \dim\Big(\mathrm{Im}(\mathbf{B}) \bigcap \mathcal{N}(\mathbf{A})\Big)$<br>$\mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) - J \leq \mathrm{r}(\mathbf{A}\mathbf{B}) \leq \min\{\mathrm{r}(\mathbf{A}), \mathrm{r}(\mathbf{B})\}$ |
| | $\mathrm{r}(\mathbf{A}\mathbf{B}\mathbf{C}) \geq \mathrm{r}(\mathbf{A}\mathbf{B}) + \mathrm{r}(\mathbf{B}\mathbf{C}) - \mathrm{r}(\mathbf{B})$ |
| Gram matrices | $\mathrm{r}(\mathbf{A}^H\mathbf{A}) = \mathrm{r}(\mathbf{A}\mathbf{A}^H) = \mathrm{r}(\mathbf{A})$. |

**Table 4.10.** *Rank of a matrix product*

NOTE 4.26.– Notice that the double inequality satisfied by $\mathrm{r}(\mathbf{A}\mathbf{B})$ in Table 4.10 provides lower and upper bounds for $\mathrm{r}(\mathbf{A}\mathbf{B})$ depending only on $\mathrm{r}(\mathbf{A})$ and $\mathrm{r}(\mathbf{B})$.

## 4.11. Determinant, inverses and generalized inverses

### 4.11.1. *Determinant*

The study of determinants has played an important role in the development of matrix algebra. The determinant of a square matrix:

– provides a criterion for invertibility of a matrix;

– can be employed for the calculation of the inverse and thereby for solving linear systems of equations;

– enables the measurement of areas and volumes;

– is directly related to the notion of characteristic polynomial and thus of eigenvalue.

Different approaches can be utilized to define the determinant of a square matrix of order $I$. Hereafter, we adopt the one relying on the notion of permutation relative to the set of integers $\{1, 2, \cdots, I\}$.

The determinant of the square matrix $\mathbf{A}$ of order $I$ is the scalar given by:

$$\det(\mathbf{A}) = \sum_{p} \sigma(p) a_{1,p_1} a_{2,p_2} \cdots a_{I,p_I} \qquad [4.12]$$

where the summation involves the $I!$ permutations $p = (p_1, p_2, \cdots, p_I)$ of the set $\{1, 2, \cdots, I\}$, and $\sigma(p)$ refers to the signature of permutation $p$ defined as:

$$\sigma(p) = \begin{cases} +1, & \text{if the initial order } 1, \cdots, I \text{ can be found from } p \\ & \text{using an even number of elementary permutations;} \\ -1, & \text{if the initial order } 1, \cdots, I \text{ can be found from } p \\ & \text{using an odd number of elementary permutations,} \end{cases}$$

an elementary permutation corresponding to the inversion of two elements of the $I$-tuple $(p_1, p_2, \cdots, p_I)$, see section 2.5.5.4.

The expression [4.12] of the determinant is called the Leibniz formula.

FACT 4.27.– It should be noted that:

– The parity of a permutation is unique, that is, if the initial order $(1, 2, \cdots, I)$ is found from $p$ using an even (odd) number of elementary permutations, then any other sequence of elementary permutations for finding $(1, 2, \cdots, I)$ from $p$ will consist of an even (odd) number of permutations.

– Every term $a_{1,p_1}a_{2,p_2},\cdots,a_{I,p_I}$ contains only one element of each row and each column of $\mathbf{A}$.

EXAMPLE 4.28.– For $\mathbf{A} \in \mathbb{K}^{3\times 3}$ and $(p_1,p_2,p_3) = (1,3,2)$, we have $\sigma(p)a_{1,p_1}a_{2,p_2}a_{3,p_3} = -a_{11}a_{23}a_{32}$. The application of formula [4.12] gives:

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

This formula corresponds to the Sarrus' rule[10] for the calculation of the determinant of a square matrix $\mathbf{A}$ of order three. This rule consists in repeating the first two columns to the right of $\mathbf{A}$, adding the products of the three terms located on each diagonal parallel to the main diagonal of $\mathbf{A}$, and subtracting the products of the three terms located on each diagonal parallel to the secondary diagonal of $\mathbf{A}$.

The determinant can also be computed using the following formula called the Laplace[11] expansion with respect to row $i$:

$$\det(\mathbf{A}) = \sum_{j=1}^{I}(-1)^{i+j}a_{ij}\det(\mathbf{A}_i^j), \quad i \in \langle I\rangle, \qquad [4.13]$$

where $\mathbf{A}_i^j$ is a square submatrix of $\mathbf{A}$, of order $I-1$, obtained by removing the $i$th row and the $j$th column in $\mathbf{A}$. The determinant $\det(\mathbf{A}_i^j)$ is called the minor associated with coefficient $a_{ij}$, and the minor preceded by the signature depending on the element's position $(i,j)$, namely, $(-1)^{i+j}\det(\mathbf{A}_i^j)$, is called the cofactor of $a_{ij}$. A similar formula based on the expansion of the determinant with respect to column $j$ can also be used.

10 Pierre Frédéric Sarrus (1798–1861), a French mathematician who received the Grand Prize of the Académie des Sciences, in 1843, for his work on the calculus of variations. He is best known for the rule bearing his name for the calculation of the determinant of a $3 \times 3$ matrix.

11 Pierre-Simon Laplace (1749–1827), French mathematician and astronomer, member of the Académie des Sciences, known to be the French Newton and died a century after him, is one of the most prolific and influential scientists. He made important contributions on a wide range of topics including integral calculus, difference and differential equations, with applications to celestial mechanics, heat theory, metric system, and theory of probability. In 1795, he founded with Lagrange (1736–1813) the Bureau des longitudes. In 1796, he published *Exposition du système du monde*, composed of five books, where he introduced his nebular hypothesis for the birth of the solar system, then, his *Traité de Mécanique céleste* (1799–1825), also composed of five volumes. In 1809, he proved the central limit theorem, a fundamental result in probability theory stating that the distribution of the sum of a large number of independent random variables tends towards a Normal distribution. He employed the Normal distribution as a model of errors. In 1812, he published his *Théorie analytique des probabilités*.

EXAMPLE 4.29.– Considering the previous example, the Laplace expansion with respect to row 1 gives:

$$\begin{aligned}\det(\mathbf{A}) &= \sum_{j=1}^{3}(-1)^{j+1}a_{1j}\det(\mathbf{A}_1^j)\\ &= (-1)^2a_{11}(a_{22}a_{33}-a_{32}a_{23})+(-1)^3a_{12}(a_{21}a_{33}-a_{31}a_{23})\\ &\quad +(-1)^4a_{13}(a_{21}a_{32}-a_{31}a_{22})\\ &= a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{13}a_{22}a_{31}-a_{12}a_{21}a_{33}\\ &\quad -a_{11}a_{23}a_{32}.\end{aligned}$$

In general, the definition based on the notion of permutation is easier to apply than the Laplace formula for the calculation of the determinant of a high-order square matrix. However, for certain matrices, such as sparse matrices, the Laplace expansion may lead to faster computation.

PROPOSITION 4.30.– *The determinant satisfies the following properties:*

*– The determinant changes the sign if two rows (or columns) of* $\mathbf{A}$ *are swapped.*

*– If a row (or column) is multiplied by a scalar* $k$*, then the determinant is multiplied by* $k$*.*

*– The determinant remains unmodified if* $k$ *times row (or column)* $i$ *is added to row (or column)* $j$*.*

*– In addition, for a square matrix of order* $I$*, we have:*

$$\det(\mathbf{A}^T)=\det(\mathbf{A});\quad \det(\mathbf{A}^H)=\det(\mathbf{A}^*)=[\det(\mathbf{A})]^*, \tag{4.14a}$$

$$\det(\alpha\mathbf{A})=\alpha^I\det(\mathbf{A}),\quad \forall\alpha\in\mathbb{K}, \tag{4.14b}$$

$$\det(-\mathbf{A})=(-1)^I\det(\mathbf{A}). \tag{4.14c}$$

*– The determinant is a multiplicative mapping, that is, it satisfies the following properties.*

PROPOSITION 4.31.– *For all square matrices* $\mathbf{A}$*,* $\mathbf{B}$*, and* $\mathbf{C}$*, we have:*

$$\det(\mathbf{AB})=\det(\mathbf{A})\det(\mathbf{B}) \tag{4.15}$$

$$\det(\mathbf{ABC})=\det(\mathbf{A})\det(\mathbf{B})\det(\mathbf{C}). \tag{4.16}$$

### 4.11.2. *Matrix inversion*

The inverse of a square matrix $\mathbf{A}$ of order $I$ is a matrix $\mathbf{X}$ such that $\mathbf{AX} = \mathbf{XA} = \mathbf{I}_I$. When such a matrix exists, it is said that $\mathbf{A}$ is invertible or non-singular, or still regular, and it is denoted by $\mathbf{A}^{-1}$. Otherwise, it is said that $\mathbf{A}$ is singular.

The inverse $\mathbf{A}^{-1}$ is unique and given by:

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\,\mathbf{A}_C^T, \qquad [4.17]$$

where the square matrix $\mathbf{A}_C$, of order $I$, is the matrix formed by the cofactors of $\mathbf{A}$.

EXAMPLE 4.32.– Inverse of a square matrix of order two:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \qquad [4.18]$$

EXAMPLE 4.33.– Inversion of a unit triangular matrix $\mathbf{A} = \mathbf{I}_n + \mathbf{B}$, where $\mathbf{B}$ is nilpotent of degree $n$, that is, $\mathbf{B}^n = \mathbf{0}_n$ (see Example 4.15):

$$(\mathbf{I}_n + \mathbf{B})^{-1} = \mathbf{I}_n - \mathbf{B} + \mathbf{B}^2 + \cdots + (-1)^{n-1}\mathbf{B}^{n-1}. \qquad [4.19]$$

Indeed, taking the nilpotence property of $\mathbf{B}$ into account, we verify that:

$$(\mathbf{I}_n + \mathbf{B})(\mathbf{I}_n - \mathbf{B} + \mathbf{B}^2 + \cdots + (-1)^{n-1}\mathbf{B}^{n-1}) = \mathbf{I}_n + (-1)^{n-1}\mathbf{B}^n = \mathbf{I}_n.$$

For instance, for $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$, we obtain:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 5 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix}.$$

A direct consequence of [4.15] and of the definition of the inverse ($\mathbf{AA}^{-1} = \mathbf{I}$), with det($\mathbf{I}$) = 1, is the following property:

PROPOSITION 4.34.– *The determinant of the inverse of a square matrix is equal to the inverse of the determinant of the matrix:*

$$det(\mathbf{A}^{-1}) = (det(\mathbf{A}))^{-1}. \qquad [4.20]$$

From [4.17], it can be deduced that $\mathbf{A}^{-1}$ cannot exist when $\det(\mathbf{A}) = 0$. In fact, the condition $\det(\mathbf{A}) \neq 0$ is also sufficient: $\mathbf{A}^{-1}$ exists if and only if $\det(\mathbf{A}) \neq 0$.

For a square matrix $\mathbf{A}$ of order $I$, the properties listed in Table 4.11 are equivalent in terms of invertibility. One should bear in mind that $\dim[\mathcal{N}(\mathbf{A})] = 0$ is the result of the rank theorem applied to a square matrix $\mathbf{A}$ of full rank, that is, when $\mathrm{r}(\mathbf{A}) = I$. See Table 4.7.

| |
|---|
| Column vectors are linearly independent |
| Row vectors are linearly independent |
| $\mathrm{r}(\mathbf{A}) = I$ |
| $\det(\mathbf{A}) \neq 0$ |
| $\dim[\mathcal{N}(\mathbf{A})] = 0$ |

**Table 4.11.** *Several manners of characterizing the invertibility of a square matrix* $\mathbf{A} \in \mathbb{C}^{I \times I}$

The inverse satisfies the properties listed in the following proposition.

PROPOSITION 4.35.– *Let* $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{I \times I}$ *denote two invertible matrices. We have:*

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}, \qquad [4.21a]$$

$$(\mathbf{A}^{*})^{-1} = (\mathbf{A}^{-1})^{*}, \qquad [4.21b]$$

$$(\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}, \quad (\mathbf{A}^{H})^{-1} = (\mathbf{A}^{-1})^{H}, \qquad [4.21c]$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}, \qquad [4.21d]$$

$$(\mathbf{A}^{k})^{-1} = (\mathbf{A}^{-1})^{k}. \qquad [4.21e]$$

Due to [4.21c], the notations $\mathbf{A}^{-T}$ and $\mathbf{A}^{-H}$ are used, respectively, to denote $(\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}$ and $(\mathbf{A}^{H})^{-1} = (\mathbf{A}^{-1})^{H}$.

### 4.11.3. *Solution of a homogeneous system of linear equations*

As indicated in Table 4.11, for a non-singular matrix $\mathbf{A}$, we have $\dim[\mathcal{N}(\mathbf{A})] = 0$, in other words, $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$. Consequently, the homogeneous system $\mathbf{Ax} = \mathbf{0}$ then only admits the trivial solution $\mathbf{x} = \mathbf{0}$.

More generally, if $\mathbf{A}$ and $\mathbf{B}$ are square matrices and $\mathbf{AB} = \mathbf{0}$, then either $\mathbf{A}$ is singular, or $\mathbf{B} = \mathbf{0}$. By symmetry, the same reasoning applies to $\mathbf{B}$: indeed, $\mathbf{AB} = \mathbf{0}$ implies $(\mathbf{AB})^{T} = \mathbf{B}^{T}\mathbf{A}^{T} = \mathbf{0}$, and so either $\mathbf{B}^{T}$ is singular (which amounts to saying that $\mathbf{B}$ is singular), or $\mathbf{A}^{T} = \mathbf{0}$, and thus $\mathbf{A} = \mathbf{0}$. It can be, therefore, concluded that $\mathbf{AB} = \mathbf{0}$ when (at least) one of the following conditions is met: (i) $\mathbf{A} = \mathbf{0}$; (ii) $\mathbf{B} = \mathbf{0}$; (iii) $\mathbf{A}$ and $\mathbf{B}$ are both singular.

EXAMPLE 4.36.– The matrices

$$\mathbf{A} = \begin{bmatrix} a & 0 \\ b & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ c & d \end{bmatrix}$$

are such that $\mathbf{AB} = \mathbf{0}$ for all $a, b, c, d \in \mathbb{K}$. It can then be concluded that $\mathbf{A}$ and $\mathbf{B}$ are singular, as it can be easily verified. On the other hand, when

$$\mathbf{A} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix},$$

with $a \neq 0$ and $b \neq 0$, then $\mathbf{AB} = \mathbf{0}$ if and only if $\mathbf{B} = \mathbf{0}$, because, in this case, $\mathbf{A}$ is non-singular.

### 4.11.4. *Complex matrix inverse*

PROPOSITION 4.37.– *Let $\mathbf{A} \in \mathbb{C}^{I \times I}$ be a complex matrix decomposed using its real and imaginary parts as $\mathbf{A} = \mathbf{X} + j\mathbf{Y}$, with $\mathbf{X} \in \mathbb{R}^{I \times I}$ and $\mathbf{Y} \in \mathbb{R}^{I \times I}$.*

*– If $\mathbf{X}$ is invertible, we have:*

$$\mathbf{A}^{-1} = (\mathbf{I}_I - j\mathbf{X}^{-1}\mathbf{Y})(\mathbf{X} + \mathbf{Y}\mathbf{X}^{-1}\mathbf{Y})^{-1}. \qquad [4.22a]$$

*– Similarly, if $\mathbf{Y}$ is invertible, we can write:*

$$\mathbf{A}^{-1} = (\mathbf{Y}^{-1}\mathbf{X} - j\mathbf{I}_I)(\mathbf{Y} + \mathbf{X}\mathbf{Y}^{-1}\mathbf{X})^{-1}. \qquad [4.22b]$$

Formulae [4.22a] and [4.22b] can be demonstrated by defining $\mathbf{A}^{-1} = \mathbf{P} + j\mathbf{Q}$, by writing that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_I$ and by solving a system of two matrix equations with unknowns $\mathbf{P}$ and $\mathbf{Q}$.

EXAMPLE 4.38.– Given

$$\mathbf{A} = \begin{bmatrix} 1+j & 1 \\ j & 1+j \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + j \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \mathbf{X} + j\mathbf{Y},$$

we have:

$$\mathbf{X}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{Y}^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix},$$

and thus:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} + j \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1-j & j \\ -1 & 1-j \end{bmatrix}.$$

### 4.11.5. *Orthogonal and unitary matrices*

A square matrix $\mathbf{A} \in \mathbb{R}^{I \times I}$ is said to be orthogonal if it satisfies $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$. According to the results of section 4.7, we can conclude that a matrix is orthogonal if and only if one of the following conditions is satisfied:

1) its rows are orthonormal, that is, $\langle \mathbf{A}_{i.}^T, \mathbf{A}_{j.}^T \rangle = \delta_{i,j}$ for all $i, j \in \langle I \rangle$;

2) its columns are orthonormal, that is, $\langle \mathbf{A}_{.i}, \mathbf{A}_{.j} \rangle = \delta_{i,j}$ for all $i, j \in \langle I \rangle$.

Another way of characterizing orthogonal matrices consists of saying that $\mathbf{A}$ is orthogonal if and only if its inverse is equal to its transpose ($\mathbf{A}^{-1} = \mathbf{A}^T$), which implies that the inverse of an orthogonal matrix is itself orthogonal. This class of matrices plays a central role in linear algebra.

In the complex case, it is said that $\mathbf{A} \in \mathbb{C}^{I \times I}$ is unitary when $\mathbf{A}^H \mathbf{A} = \mathbf{A}\mathbf{A}^H = \mathbf{I}$. Therefore, the same conditions 1) and 2) described earlier can be applied to the complex case with the Hermitian inner product. The inverse of $\mathbf{A}$ is then equal to its conjugate transpose ($\mathbf{A}^{-1} = \mathbf{A}^H$).

As we have seen in section 3.5.4, the product of two orthogonal (unitary) matrices yields an orthogonal (unitary) matrix.

PROPOSITION 4.39.– *Using the definition of an orthogonal matrix ($\mathbf{A}\mathbf{A}^T = \mathbf{I}$) and of a unitary matrix ($\mathbf{A}\mathbf{A}^H = \mathbf{I}$), and properties [4.14a] and [4.15] of the determinant, the following results can be deduced:*

*– The determinant of an orthogonal matrix is equal to* $\pm 1$.

*– The determinant of a unitary matrix is of modulus 1.*

PROPOSITION 4.40.– *Another important property of an orthogonal/unitary matrix* $\mathbf{A}$ *is the invariance of the norm of a vector when it is multiplied by* $\mathbf{A}$*:*

$$\|\mathbf{A}\mathbf{x}\|_2^2 = \mathbf{x}^T \mathbf{A}^T \mathbf{A}\mathbf{x} = \|\mathbf{x}\|_2^2 \quad \forall \mathbf{x} \in \mathbb{R}^I$$

$$\|\mathbf{A}\mathbf{x}\|_2^2 = \mathbf{x}^H \mathbf{A}^H \mathbf{A}\mathbf{x} = \|\mathbf{x}\|_2^2 \quad \forall \mathbf{x} \in \mathbb{C}^I.$$

In Table 4.12, we summarize the definitions of orthogonal and unitary matrices, with those of normal and column-orthonormal matrices. From the definition given in Table 4.12, we deduce that a real matrix is normal if it commutes with its transpose, whereas a complex matrix is normal if it commutes with its conjugate transpose. Consequently, symmetric and Hermitian matrices are also normal.

### 4.11.6. *Involutory matrices and anti-involutory matrices*

A square matrix $\mathbf{A}$ of order $I$ is said to be involutory if it is equal to its inverse

$$\mathbf{A}^{-1} = \mathbf{A} \quad \Rightarrow \quad \mathbf{A}^2 = \mathbf{I}_I.$$

When $\mathbf{A}$ is equal to its inverse with the opposite sign, that is, $\mathbf{A}^{-1} = -\mathbf{A}$ (which implies $\mathbf{A}^2 = -\mathbf{I}_I$), it is said to be anti-involutory.

| Properties | Conditions |
|---|---|
| $\mathbf{A} \in \mathbb{R}^{I\times I}$ normal | $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T$ |
| $\mathbf{A} \in \mathbb{C}^{I\times I}$ normal | $\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H$ |
| $\mathbf{A} \in \mathbb{R}^{I\times I}$ column-orthonormal | $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ |
| $\mathbf{A} \in \mathbb{C}^{I\times I}$ column-orthonormal | $\mathbf{A}^H\mathbf{A} = \mathbf{I}$ |
| $\mathbf{A} \in \mathbb{R}^{I\times I}$ orthogonal | $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$ |
| $\mathbf{A} \in \mathbb{C}^{I\times I}$ unitary | $\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H = \mathbf{I}$ |

**Table 4.12.** *Normal, column-orthonormal, orthogonal and unitary matrices*

PROPOSITION 4.41.– *A real (complex) symmetric (Hermitian) matrix is involutory if and only if it is orthogonal (unitary).*

*Indeed, when* $\mathbf{A} \in \mathbb{R}^{I\times I}$ *and* $\mathbf{A}^T = \mathbf{A}$*, we have:*

$$\mathbf{A}^{-1} = \mathbf{A} \quad \Leftrightarrow \quad \mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}_I.$$

*Similarly, if* $\mathbf{A} \in \mathbb{C}^{I\times I}$ *and* $\mathbf{A}^H = \mathbf{A}$*, we get:*

$$\mathbf{A}^{-1} = \mathbf{A} \quad \Leftrightarrow \quad \mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A} = \mathbf{I}_I.$$

EXAMPLE 4.42.– The matrix $\mathbf{A} = \mathbf{I}_I - 2\mathbf{x}\mathbf{z}^T$ is involutory if $\mathbf{z}^T\mathbf{x} = 1$. In particular, if $\mathbf{z} = \|\mathbf{x}\|_2^{-2}\mathbf{x}$, then the matrix $\mathbf{A} = \mathbf{I}_I - \frac{2}{\|\mathbf{x}\|_2^2}\mathbf{x}\mathbf{x}^T$ is involutory and symmetric, thereby orthogonal (according to the aforementioned property). This matrix is called reflector, or Householder matrix, associated with the non-null vector $\mathbf{x} \in \mathbb{R}^I$.

This type of matrix is used in matrix factorization algorithms, such as $\mathbf{QR}$ factorization introduced in section 3.6.5, using the Gram–Schmidt orthonormalization method. The transformation associated with the aforementioned matrix $\mathbf{A}$, called Householder (1958) transformation, transforms a vector $\mathbf{u} = k\mathbf{x} + \mathbf{y}$, with $\mathbf{x} \perp \mathbf{y}$, into a vector $\mathbf{v} = \mathbf{A}\mathbf{u} = (\mathbf{I}_I - \frac{2}{\|\mathbf{x}\|_2^2}\mathbf{x}\mathbf{x}^T)(k\mathbf{x} + \mathbf{y}) = -k\mathbf{x} + \mathbf{y}$ which is the reflection of the vector $\mathbf{u}$ through the subspace generated by $\mathbf{y}$ and orthogonal to $\mathbf{x}$.

So, for example, in the plane $\mathbb{R}^2$ where $\mathbf{x} = \mathbf{e}_1$ and $\mathbf{y} = \mathbf{e}_2$ are the unit vectors according to the horizontal and vertical axes, respectively, the vector $\mathbf{u} = \begin{bmatrix} k \\ 1 \end{bmatrix}$ is transformed into $\mathbf{v} = \begin{bmatrix} -k \\ 1 \end{bmatrix}$, a vector symmetric to $\mathbf{u}$ with respect to the vertical axis.

EXAMPLE 4.43.– Permutation matrices, which are square binary matrices having one entry equal to 1 in each row and each column, are involutory.

### 4.11.7. *Left and right inverses of a rectangular matrix*

#### 4.11.7.1. *Definitions and existence theorem*

A matrix $\mathbf{A} \in \mathbb{K}^{I\times J}$ is said to be left (or right) invertible if there exists a matrix $\mathbf{A}_G^{-1}$(or $\mathbf{A}_D^{-1}$) $\in \mathbb{K}^{J\times I}$ such that: $\mathbf{A}_G^{-1}\mathbf{A} = \mathbf{I}_J$ (or $\mathbf{A}\mathbf{A}_D^{-1} = \mathbf{I}_I$). Matrices $\mathbf{A}_G^{-1}$ and $\mathbf{A}_D^{-1}$ are, respectively, called left-inverse and right-inverse of $\mathbf{A}$.

When $\mathbf{A}$ is square and non-singular, the two inverses are equal: $\mathbf{A}_G^{-1} = \mathbf{A}_D^{-1} = \mathbf{A}^{-1}$. For $I \neq J$, these inverses are not unique, and they can exist on one side only: on the left if $I > J$ or on the right if $I < J$.

THEOREM 4.44 [Existence theorem of left and right inverses (Perlis 1958)].– *Given a matrix $\mathbf{A} \in \mathbb{K}^{I\times J}$, we have:*

*1) $\mathbf{A}$ has a left-inverse if and only if $I \geq J$ and $\mathrm{r}(\mathbf{A}) = J$, that is, if $\mathbf{A}$ has full column rank.*

*2) $\mathbf{A}$ has a right-inverse if and only if $J \geq I$ and $\mathrm{r}(\mathbf{A}) = I$, that is, if $\mathbf{A}$ has full row rank.*

EXAMPLE 4.45.– Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}. \qquad [4.23]$$

We have $I = 3$, $J = 2$, and $\mathrm{r}(\mathbf{A}) = 2$. Subsequently, there exists a left inverse:

$$\mathbf{A}_G^{-1} = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \end{bmatrix}$$

such that $\mathbf{A}_G^{-1}\mathbf{A} = \mathbf{I}_2$, which gives the four following equations:

$$\begin{aligned} x_1 + x_2 + x_3 &= 1, \\ x_4 + x_5 + x_6 &= 0, \\ x_1 + x_3 &= 0, \\ x_4 + x_6 &= 1. \end{aligned}$$

Since there are four equations and six unknowns, there exists an infinite number of left inverses. Let $x_1 = x_3 = 0$ and $x_4 = \alpha$. A left inverse is then given by:

$$\mathbf{A}_G^{-1}(\alpha) = \begin{bmatrix} 0 & 1 & 0 \\ \alpha & -1 & 1-\alpha \end{bmatrix}, \quad \alpha \in \mathbb{K}, \qquad [4.24]$$

#### 4.11.7.2. *Formulae for left and right inverses*

PROPOSITION 4.46.– *Given a matrix* $\mathbf{A} \in \mathbb{K}^{I \times J}$ *of full rank, we have:*

*– If* $I \geq J$*, all left inverses are given by the partitioned formula:*

$$\mathbf{A}_G^{-1} = [\mathbf{A}_1^{-1} - \mathbf{Y}\mathbf{A}_2\mathbf{A}_1^{-1} \quad \mathbf{Y}]\mathbf{P}, \qquad [4.25]$$

*where* $\mathbf{P} \in \mathbb{K}^{I \times I}$ *is a permutation matrix such that:*

$$\mathbf{P}\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix},$$

*with* $\mathbf{A}_1 \in \mathbb{K}^{J \times J}$ *non-singular,* $\mathbf{A}_2 \in \mathbb{K}^{(I-J) \times J}$*, and* $\mathbf{Y} \in \mathbb{K}^{J \times (I-J)}$ *arbitrary.*

*– If* $J \geq I$*, all right-inverses are given by:*

$$\mathbf{A}_D^{-1} = \mathbf{P} \begin{bmatrix} \mathbf{A}_1^{-1} - \mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{Y} \\ \mathbf{Y} \end{bmatrix}, \qquad [4.26]$$

*where* $\mathbf{P} \in \mathbb{K}^{J \times J}$ *is a permutation matrix such that* $\mathbf{A}\mathbf{P} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix}$*, with* $\mathbf{A}_1 \in \mathbb{K}^{I \times I}$ *non-singular,* $\mathbf{A}_2 \in \mathbb{K}^{I \times (J-I)}$*, and* $\mathbf{Y} \in \mathbb{K}^{(J-I) \times I}$ *arbitrary.*

PROOF.– Consider the case $I \geq J$. Since matrix $\mathbf{A}$ is of full rank, it is possible to swap its rows so that the first $J$ rows are independent, that is, there exists a permutation matrix $\mathbf{P}$ such that:

$$\mathbf{P}\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix},$$

with $\mathbf{A}_1 \in \mathbb{K}^{J \times J}$ non-singular, and $\mathbf{A}_2 \in \mathbb{K}^{(I-J) \times J}$.

The permutation matrix being orthogonal (i.e. $\mathbf{P}\mathbf{P}^T = \mathbf{P}^T\mathbf{P} = \mathbf{I}_I$), we have $\mathbf{A}_G^{-1}\mathbf{A} = \mathbf{A}_G^{-1}\mathbf{P}^T\mathbf{P}\mathbf{A} = \mathbf{I}_J$ for any left inverse $\mathbf{A}_G^{-1}$. By decomposing $\mathbf{A}_G^{-1}\mathbf{P}^T$ into the partitioned form in two column-blocks

$$\mathbf{A}_G^{-1}\mathbf{P}^T = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \end{bmatrix}, \qquad [4.27]$$

with $\mathbf{X} \in \mathbb{K}^{J \times J}$ and $\mathbf{Y} \in \mathbb{K}^{J \times (I-J)}$, from $\begin{bmatrix} \mathbf{X} & \mathbf{Y} \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} = \mathbf{I}_J$, it follows that:

$$\mathbf{X}\mathbf{A}_1 + \mathbf{Y}\mathbf{A}_2 = \mathbf{I}_J \quad \Rightarrow \quad \mathbf{X} = \mathbf{A}_1^{-1} - \mathbf{Y}\mathbf{A}_2\mathbf{A}_1^{-1}. \qquad [4.28]$$

Consequently, from relations [4.27] and [4.28], with $(\mathbf{P}^T)^{-1} = \mathbf{P}$, it can be deduced that:

$$\mathbf{A}_G^{-1} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \end{bmatrix} \mathbf{P} = \begin{bmatrix} \mathbf{A}_1^{-1} - \mathbf{Y}\mathbf{A}_2\mathbf{A}_1^{-1} & \mathbf{Y} \end{bmatrix} \mathbf{P}.$$

Formula [4.26] can be similarly demonstrated. □

EXAMPLE 4.47.– For

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

we define:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

which gives $\mathbf{A}_1 = \mathbf{I}_2, \mathbf{A}_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}$. Thus, we have:

$$\mathbf{X} = \mathbf{I}_2 - \mathbf{Y}\begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1-a & 0 \\ -b & 1 \end{bmatrix}, \ \mathbf{Y} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\mathbf{A}_G^{-1} = [\mathbf{X} \ \ \mathbf{Y}]\mathbf{P} = \begin{bmatrix} 1-a & a & 0 \\ -b & b & 1 \end{bmatrix}.$$

### 4.11.8. *Generalized inverses*

In the previous section, we introduced the notion of left and right inverses of a rectangular matrix, such inverses existing if and only if the matrix has full rank. In the following, we will define the notion of generalized inverse for any matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$. Noting that when it exists, a left or right inverse of $\mathbf{A}$, denoted by $\mathbf{A}^\sharp$, satisfies the following equations:

$$\mathbf{A}\mathbf{A}^\sharp\mathbf{A} = \mathbf{A}, \quad \mathbf{A}^\sharp\mathbf{A}\mathbf{A}^\sharp = \mathbf{A}^\sharp, \qquad [4.29]$$

a generalized inverse of $\mathbf{A} \in \mathbb{K}^{I \times J}$ is defined as a matrix $\mathbf{A}^\sharp \in \mathbb{K}^{J \times I}$ that satisfies equations [4.29].

It can be shown that any matrix has a generalized inverse. Nonetheless, this generalized inverse is not unique in general. In fact, it is unique if and only if $\mathbf{A} = \mathbf{0}$ or if $\mathbf{A}$ is a regular square matrix.

Any generalized inverse of a singular matrix $\mathbf{A}$ can be obtained using a bordering technique that allows to remove the singularity of $\mathbf{A}$ by building a non-singular bordered matrix $\mathbf{M}$ such that (Bjerhammar 1973):

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{D} & \mathbf{B} \end{bmatrix}.$$

In Nomakuchi (1980), it is shown that the submatrix corresponding to $\mathbf{A}$ in $\mathbf{M}^{-1}$ is a generalized inverse of $\mathbf{A}$.

We apply this procedure hereafter for a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ of full column rank, and then of full row rank using an orthogonal bordering of $\mathbf{A}$, i.e. using columns and rows orthogonal to columns and rows of $\mathbf{A}$, respectively, so as to form a non-singular square matrix $\mathbf{M}$. When $\mathbb{K} = \mathbb{C}$, transposes are to be replaced by conjugate transposes in the formulae that follow.

– If $\mathbf{A}$ is of full column rank and $I > J$, then $\mathbf{A}^T\mathbf{A}$ is invertible (see Table 4.10), and the generalized inverse of $\mathbf{A}$ is given by:

$$\mathbf{A}^\sharp = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T. \qquad [4.30a]$$

It is indeed possible to find a matrix $\mathbf{C} \in \mathbb{K}^{I \times (I-J)}$ of full column rank ($\det(\mathbf{C}^T\mathbf{C}) \neq 0$) such that $\mathbf{M} = \begin{bmatrix}\mathbf{A} & \mathbf{C}\end{bmatrix}$ is regular, with $\mathbf{C}^T\mathbf{A} = \mathbf{0}$. Equivalently, we have $\mathbf{A}^T\mathbf{C} = \mathbf{0}$, which means that the columns of $\mathbf{C}$ belong to the kernel $\mathcal{N}(\mathbf{A}^T)$. As a result, according to [4.8a], it can be concluded that the columns of $\mathbf{C}$ form a basis of the orthogonal complement of the column space of $\mathbf{A}$. Since $\mathbf{M}^{-1} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$ for any regular matrix $\mathbf{M}$, we have:

$$\mathbf{M}^{-1} = \begin{bmatrix}\mathbf{A}^T\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^T\mathbf{C}\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{A}^T \\ \mathbf{C}^T\end{bmatrix} = \begin{bmatrix}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \\ (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\end{bmatrix} = \begin{bmatrix}\mathbf{A}^\sharp \\ \mathbf{C}^\sharp\end{bmatrix}.$$

From this equation, we deduce [4.30a].

– If $\mathbf{A}$ is of full row rank and $J > I$, then similarly $\mathbf{A}\mathbf{A}^T$ is regular and the generalized inverse of $\mathbf{A}$ is given by:

$$\mathbf{A}^\sharp = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \qquad [4.30b]$$

Indeed, we can find a matrix $\mathbf{D} \in \mathbb{K}^{(J-I) \times J}$ of full row rank ($\det(\mathbf{D}\mathbf{D}^T) \neq 0$) such that

$$\mathbf{M} = \begin{bmatrix}\mathbf{A} \\ \mathbf{D}\end{bmatrix}$$

is regular, with $\mathbf{A}\mathbf{D}^T = \mathbf{0}$. In this case, the rows of $\mathbf{D}$ belong to the kernel $\mathcal{N}(A)$ and, according to [4.8b], they form a basis for the orthogonal complement of the space $C(\mathbf{A}^T)$, that is, the row space of $\mathbf{A}$. By writing that $\mathbf{M}^{-1} = \mathbf{M}^T(\mathbf{M}\mathbf{M}^T)^{-1}$, a similar reasoning to the previous case leads to the following formula:

$$\mathbf{M}^{-1} = \begin{bmatrix}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} & \mathbf{D}^T(\mathbf{D}\mathbf{D}^T)^{-1}\end{bmatrix} = \begin{bmatrix}\mathbf{A}^\sharp & \mathbf{D}^\sharp\end{bmatrix},$$

from which the expression [4.30b] of the generalized inverse can be deduced. It is easy to verify that the generalized inverses [4.30a] and [4.30b] satisfy the conditions [4.29].

## 4.11.9. *Moore–Penrose pseudo-inverse*

### 4.11.9.1. *Definition and uniqueness*

The uniqueness of the generalized inverse can be obtained by imposing additional constraints. This is the case of the Moore–Penrose pseudo-inverse, denoted by $\mathbf{A}^\dagger$, which is defined using the four following relations:

$$\begin{array}{llll} \text{(i)} & \mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}, & \text{(ii)} & \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger, \\ \text{(iii)} & (\mathbf{A}^\dagger\mathbf{A})^T = \mathbf{A}^\dagger\mathbf{A}, & \text{(iv)} & (\mathbf{A}\mathbf{A}^\dagger)^T = \mathbf{A}\mathbf{A}^\dagger. \end{array} \quad [4.31]$$

When $\mathbb{K} = \mathbb{C}$, that is, in the case of a complex matrix, transposition must be replaced by conjugate transposition in [4.31]. Relations (iii) and (iv) denote that $\mathbf{A}^\dagger\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\dagger$ are symmetric if $\mathbb{K} = \mathbb{R}$ and Hermitian if $\mathbb{K} = \mathbb{C}$, respectively.

It can be verified that the generalized inverses [4.30a]–[4.30b] satisfy the relations of definition [4.31] of the Moore–Penrose pseudo-inverse.

PROPOSITION 4.48.– *The Moore–Penrose pseudo-inverse of a matrix* $\mathbf{A}$ *is unique.*

PROOF.– Let $\mathbf{X}$ and $\mathbf{Y}$ denote two pseudo-inverses of $\mathbf{A}$. The use of the relations of definition [4.31] for $\mathbf{X}$ and $\mathbf{Y}$, as indicated in the following equalities, gives (Rotella and Borne 1995):

$$\begin{aligned} \mathbf{X} &\overset{\text{(ii)}}{=} \mathbf{X}\mathbf{A}\mathbf{X} \overset{\text{(iv)}}{=} \mathbf{X}\mathbf{X}^H\mathbf{A}^H \overset{\text{(i)}}{=} \mathbf{X}\mathbf{X}^H(\mathbf{A}\mathbf{Y}\mathbf{A})^H = \mathbf{X}(\mathbf{A}\mathbf{X})^H(\mathbf{A}\mathbf{Y})^H \\ &\overset{\text{(iv)}}{=} (\mathbf{X}\mathbf{A}\mathbf{X})(\mathbf{A}\mathbf{Y}) \overset{\text{(i)}}{=} \mathbf{X}\mathbf{A}\mathbf{Y} \overset{\text{(iii)}}{=} \mathbf{A}^H\mathbf{X}^H\mathbf{Y} \overset{\text{(i)}}{=} (\mathbf{A}\mathbf{Y}\mathbf{A})^H\mathbf{X}^H\mathbf{Y} \\ &= (\mathbf{Y}\mathbf{A})^H(\mathbf{X}\mathbf{A})^H\mathbf{Y} \overset{\text{(iii)}}{=} \mathbf{Y}(\mathbf{A}\mathbf{X}\mathbf{A})\mathbf{Y} \overset{\text{(i)}}{=} \mathbf{Y}\mathbf{A}\mathbf{Y} \overset{\text{(ii)}}{=} \mathbf{Y}. \end{aligned}$$

□

EXAMPLE 4.49.– Let us again consider the matrix $\mathbf{A}$ defined by [4.23], which has full column rank. The Moore–Penrose pseudo-inverse is then:

$$\mathbf{A}^\dagger = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & -1 & 1/2 \end{bmatrix},$$

which corresponds to the left inverse [4.24] with $\alpha = 1/2$.

### 4.11.9.2. *Properties*

In Table 4.13, we summarize the key properties of the Moore–Penrose pseudo-inverse of a matrix $\mathbf{A} \in \mathbb{K}^{I\times J}$ (Lancaster and Tismenetsky 1985; Magnus and Neudecker 1988).

| | Properties | Conditions |
|---|---|---|
| (i) | $\mathbf{A}^\dagger = \mathbf{A}^{-1}$ | $\mathbf{A}$ non-singular |
| (ii) | $(\mathbf{A}^\dagger)^\dagger = \mathbf{A}$ | |
| (iii) | $(\mathbf{A}^H)^\dagger = (\mathbf{A}^\dagger)^H$ | |
| (iv) | $(\alpha\mathbf{A})^\dagger = \alpha^{-1}\mathbf{A}^\dagger, \forall\alpha \in \mathbb{K}, \alpha \neq 0$ | |
| (v) | $(\mathbf{A}\mathbf{A}^\dagger)^\dagger = \mathbf{A}\mathbf{A}^\dagger\ ,\ (\mathbf{A}^\dagger\mathbf{A})^\dagger = \mathbf{A}^\dagger\mathbf{A}$ | |
| (vi) | $(\mathbf{A}\mathbf{A}^H)^\dagger = (\mathbf{A}^\dagger)^H\mathbf{A}^\dagger\ ,\ (\mathbf{A}^H\mathbf{A})^\dagger = \mathbf{A}^\dagger(\mathbf{A}^\dagger)^H$ | |
| (vii) | $\mathbf{A}^H\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^H = \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^H$ | |
| (viii) | $\mathbf{A}^H(\mathbf{A}^\dagger)^H\mathbf{A}^\dagger = \mathbf{A}^\dagger = \mathbf{A}^\dagger(\mathbf{A}^\dagger)^H\mathbf{A}^H$ | |
| (ix) | $\mathbf{A}(\mathbf{A}^H\mathbf{A})^\dagger\mathbf{A}^H\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^\dagger\mathbf{A}$ | |
| (x) | $\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^\dagger\mathbf{A}^H = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^\dagger$ | |
| (xi) | $\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$ | $\mathbf{A}$ of full column rank |
| (xii) | $\mathbf{A}^\dagger = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}$ | $\mathbf{A}$ of full row rank |
| (xiii) | $(\mathbf{A}\mathbf{C}\mathbf{B})^\dagger = \mathbf{B}^\dagger\mathbf{C}^{-1}\mathbf{A}^\dagger$ | $\mathbf{A} \in \mathbb{K}^{I\times K}, \mathbf{B} \in \mathbb{K}^{K\times J}, \mathbf{C} \in \mathbb{K}^{K\times K}$ of full rank |
| (xiv) | $(\mathbf{A}\mathbf{C}\mathbf{B})^\dagger = \mathbf{B}^H\mathbf{C}^\dagger\mathbf{A}^H$ | $\mathbf{A}$ and $\mathbf{B}$ unitary , $\mathbf{C} \in \mathbb{K}^{I\times J}$ |
| (xv) | $\mathrm{r}(\mathbf{A}^\dagger) = \mathrm{r}(\mathbf{A})$ | |
| (xvi) | $(\mathbf{0}_{I\times J})^\dagger = \mathbf{0}_{J\times I}$ | |

**Table 4.13.** *Properties of the Moore–Penrose pseudo-inverse*

## 4.12. Multiplicative groups of matrices

In Table 4.14, we summarize the multiplicative groups of matrices.

For example, consider the set of invertible square matrices of order $n$, denoted by $GL_n$ (for General Linear Group). This set has a multiplicative group structure, that is, a group for the multiplication. Indeed, the set $GL_n$ is closed under the multiplication operation (i.e. if $\mathbf{A}$ and $\mathbf{B} \in GL_n$, then $\mathbf{AB} \in GL_n$), and the following properties are satisfied for the product operation:

| Groups | Sets of matrices | Structural constraints |
|---|---|---|
| $GL_n \subseteq \mathbb{K}^{n\times n}$ | Invertible | $\det(\mathbf{A}) \neq 0$ |
| $SL_n \subseteq \mathbb{K}^{n\times n}$ | Invertible with determinant 1 | $\det(\mathbf{A}) = 1$ |
| $O_n \subseteq \mathbb{R}^{n\times n}$ | Orthogonal | $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}_n$ |
| $U_n \subseteq \mathbb{C}^{n\times n}$ | Unitary | $\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H = \mathbf{I}_n$ |
| $SO_n \subseteq \mathbb{R}^{n\times n}$ | Orthogonal with determinant 1 | $\mathbf{A}^T\mathbf{A} = \mathbf{I}_n$ and $\det(\mathbf{A}) = 1$ |
| $SU_n \subseteq \mathbb{C}^{n\times n}$ | Unitary of determinant 1 | $\mathbf{A}^H\mathbf{A} = \mathbf{I}_n$ and $\det(\mathbf{A}) = 1$ |
| $SP_n \subseteq \mathbb{R}^{2n\times 2n}$ | Symplectic | $\mathbf{A}^T\widetilde{\mathbf{J}}\mathbf{A} = \widetilde{\mathbf{J}}$ (‡) |

(‡) $\widetilde{\mathbf{J}} = \begin{bmatrix} \mathbf{0}_n & \mathbf{I}_n \\ -\mathbf{I}_n & \mathbf{0}_n \end{bmatrix}$.

**Table 4.14.** *Multiplicative groups of matrices*

– Associativity:

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} \quad \forall \mathbf{A}, \mathbf{B}, \mathbf{C} \in GL_n.$$

– Existence of an identity element, the identity matrix $\mathbf{I}_n \in GL_n$:

$$\mathbf{A}\mathbf{I}_n = \mathbf{I}_n\mathbf{A} = \mathbf{A} \quad \forall \mathbf{A} \in GL_n.$$

– Existence of an inverse $\mathbf{A}^{-1} \in GL_n$ for any matrix $\mathbf{A} \in GL_n$:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n \quad \forall \mathbf{A} \in GL_n.$$

The sets $O_n = \{\mathbf{A} \in \mathbb{R}^{n\times n} : \mathbf{A} \text{ orthogonal}\}$ and $U_n = \{\mathbf{A} \in \mathbb{C}^{n\times n} : \mathbf{A} \text{ unitary}\}$ of orthogonal matrices and unitary matrices of order $n$ also have a group structure. In effect, the following properties are satisfied:

– $\mathbf{AB} \in U_n$, $\forall \mathbf{A}, \mathbf{B} \in U_n$, that is $U_n$ closed under the multiplication operation;

– $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} \quad \forall \mathbf{A}, \mathbf{B}, \mathbf{C} \in U_n$;

– $\mathbf{I}_n \in U_n$;

– $\mathbf{A}^{-1} \in U_n, \forall \mathbf{A} \in U_n$.

Recall that the conditions for a subset $M \subseteq GL_n$ to be a subgroup of $GL_n$ are:

– $\mathbf{AB} \in M$, $\forall \mathbf{A}, \mathbf{B} \in M$, that is, $M$ closed under the multiplication operation;

– $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} \quad \forall \mathbf{A}, \mathbf{B}, \mathbf{C} \in M$;

– $\mathbf{I}_n \in M$;

– for all $\mathbf{A} \in GL_n$, if $\mathbf{A} \in M$, then $\mathbf{A}^{-1} \in M$.

Then, it is easy to verify that $O_n$ and $U_n$ are subgroups of $GL_n$. Moreover, since the determinant of any matrix belonging to $O_n$ (respectively, $U_n$) is equal to $\pm 1$ (respectively, of modulus 1), the property $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ implies that the sets $SO_n$ and $SU_n$ of orthogonal matrices and unitary matrices having a determinant equal to 1 are subgroups of $O_n$ and $U_n$, respectively. Note that the same result does not apply to matrices of $O_n$ and of $U_n$ whose determinant is $-1$.

## 4.13. Matrix associated to a linear map

### 4.13.1. *Matrix representation of a linear map*

Any linear map $\mathcal{L} : \mathcal{U} \to \mathcal{V}$ from a $\mathbb{K}$-v.s. $\mathcal{U}$, of dimension $J$ with the ordered basis $\{\mathbf{u}\} = \{\mathbf{u}_1, \cdots, \mathbf{u}_J\}$, to a $\mathbb{K}$-v.s. $\mathcal{V}$, of dimension $I$ with the basis

$\{\mathbf{v}\} = \{\mathbf{v}_1, \cdots, \mathbf{v}_I\}$, is defined using the images $\mathcal{L}(\mathbf{u}_j)$, $j \in \langle J \rangle$, expressed in the basis $\{\mathbf{v}\}$, namely:

$$\mathcal{L}(\mathbf{u}_j) = \sum_{i=1}^{I} a_{ij}\mathbf{v}_i, \quad j \in \langle J \rangle.$$

The coefficients $a_{ij}$ are the coordinates of $\mathcal{L}(\mathbf{u}_j)$ in the basis $\{\mathbf{v}\}$. The matrix $\mathbf{A} = [a_{ij}]$, of dimensions $I \times J$, is called the matrix associated with the linear map $\mathcal{L}$, relative to the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$. This dependency of $\mathbf{A}$ with respect to bases will be highlighted by denoting $\mathbf{A}_{\mathbf{vu}}$ if necessary.

By defining the matrix whose columns are the images $\mathcal{L}(\mathbf{u}_j)$, the following equation is obtained:

$$\begin{aligned}\left[\mathcal{L}(\mathbf{u}_1) \cdots \mathcal{L}(\mathbf{u}_j) \cdots \mathcal{L}(\mathbf{u}_J)\right] &= \left[\mathbf{v}_1 \cdots \mathbf{v}_I\right] \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1J} \\ \vdots & & \vdots & & \vdots \\ a_{I1} & \cdots & a_{Ij} & \cdots & a_{IJ} \end{bmatrix} \\ &= \left[\mathbf{v}_1 \cdots \mathbf{v}_I\right] \mathbf{A}. \end{aligned} \quad [4.32]$$

PROPOSITION 4.50.– *The matrix* $\mathbf{A}$ *completely specifies the map* $\mathcal{L}$ *with respect to the bases* $\{\mathbf{u}\}$ *and* $\{\mathbf{v}\}$. *In other words, given any element* $\mathbf{x}$ *of* $\mathcal{U}$ *and its image* $\mathbf{y} = \mathcal{L}(\mathbf{x})$ *in* $\mathcal{V}$, *vectors* $\mathbf{x}_\mathbf{u}$ *and* $\mathbf{y}_\mathbf{v}$ *of the coordinates of* $\mathbf{x}$ *and* $\mathbf{y}$ *in the respective bases* $\{\mathbf{u}\}$ *and* $\{\mathbf{v}\}$ *are related by the following relationship*[12]*:*

$$\mathbf{y}_\mathbf{v} = \mathbf{A}\mathbf{x}_\mathbf{u} = \mathbf{A}_{\mathbf{vu}}\mathbf{x}_\mathbf{u}.$$

PROOF.– Let $\mathbf{x} = \sum_{j=1}^{J} x_j\mathbf{u}_j$ and $\mathbf{y} = \sum_{i=1}^{I} y_i\mathbf{v}_i$ be the expansions of $\mathbf{x}$ and $\mathbf{y}$ over the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$. Using the linearity property of $\mathcal{L}$, we have:

$$\mathbf{y} = \mathcal{L}(\mathbf{x}) = \sum_{j=1}^{J} x_j\mathcal{L}(\mathbf{u}_j) = \sum_{j=1}^{J} x_j \left(\sum_{i=1}^{I} a_{ij}\mathbf{v}_i\right) = \sum_{i=1}^{I} \left(\sum_{j=1}^{J} a_{ij}x_j\right) \mathbf{v}_i,$$

from which it can be deduced that:

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

By defining the vectors of coordinates:

$$\mathbf{x}_\mathbf{u} = \left[x_1 \cdots x_J\right]^T \quad \text{and} \quad \mathbf{y}_\mathbf{v} = \left[y_1 \cdots y_I\right]^T,$$

12 The vectors (in the broad sense) of the v.s. $\mathcal{U}$ and $\mathcal{V}$ are denoted in bold to distinguish them from their scalar coordinates.

we get the relation:

$$\mathbf{y_v} = \mathbf{A x_u} = \mathbf{A_{vu} x_u}. \qquad [4.33]$$

□

FACT 4.51.– Let us define the mappings $\mathcal{E}_{\mathbf{u}} : \mathcal{U} \to \mathbb{K}^J$ and $\mathcal{F}_{\mathbf{v}} : \mathcal{V} \to \mathbb{K}^I$ that transform an element $\mathbf{x}$ of $\mathcal{U}$ and an element $\mathbf{y}$ of $\mathcal{V}$ into their vectors of coordinates: $\mathcal{U} \ni \mathbf{x} \mapsto \mathbf{x_u} \in \mathbb{K}^J$ and $\mathcal{V} \ni \mathbf{y} \mapsto \mathbf{y_v} \in \mathbb{K}^I$, respectively. The linear map $\mathcal{L}$ can also be viewed as the composition $\mathcal{F}_{\mathbf{v}}^{-1} \circ \mathcal{L}_{\mathbf{vu}} \circ \mathcal{E}_{\mathbf{u}}$, where $\mathcal{L}_{\mathbf{vu}} : \mathbb{K}^J \to \mathbb{K}^I$ is a linear map such that $\mathbf{x_u} \mapsto \mathbf{y_v} = \mathcal{L}_{\mathbf{vu}}(\mathbf{x_u}) = \mathbf{A_{vu} x_u}$.

Indeed, for all $\mathbf{x} \in \mathcal{U}$, we have $\mathcal{F}_{\mathbf{v}}^{-1} \circ \mathcal{L}_{\mathbf{vu}} \circ \mathcal{E}_{\mathbf{u}}(\mathbf{x}) = \mathcal{F}_{\mathbf{v}}^{-1} \circ \mathcal{L}_{\mathbf{vu}}(\mathbf{x_u}) = \mathcal{F}_{\mathbf{v}}^{-1}(\mathbf{y_v}) = \mathbf{y} = \mathcal{L}(\mathbf{x})$, and therefore:

$$\mathcal{L} = \mathcal{F}_{\mathbf{v}}^{-1} \circ \mathcal{L}_{\mathbf{vu}} \circ \mathcal{E}_{\mathbf{u}}. \qquad [4.34]$$

This is illustrated in Figure 4.3, in section 4.13.2.

PROPOSITION 4.52.– *Let $\mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$ be the v.s. of the linear maps*[13] *from $\mathcal{U}$ to $\mathcal{V}$. The following properties can be demonstrated:*

*– If $\mathcal{U} = \mathbb{K}^J$ and $\mathcal{V} = \mathbb{K}^I$, and if the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$ are the canonical bases of these spaces, that is, $\{\mathbf{u}_j = \mathbf{e}_j^{(J)}\}$ and $\{\mathbf{v}_i = \mathbf{e}_i^{(I)}\}$, we then have $\mathbf{x} = \mathbf{x_u}$ and $\mathbf{y} = \mathbf{y_v}$, and thus $\mathbf{y} = \mathbf{A x}$.*

*– Given the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$ of $\mathcal{U}$ and $\mathcal{V}$, the mapping:*

$$\begin{aligned} &\varphi : \mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V}) \to \mathbb{K}^{I \times J} \\ &\qquad \mathcal{L} \mapsto \varphi(\mathcal{L}) = \mathbf{A_{vu}}, \end{aligned} \qquad [4.35]$$

*which makes the correspondence between an element $\mathcal{L} \in \mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$ and its associated matrix $\mathbf{A_{vu}}$, is bijective. In addition, it is easy to verify that for all $\mathcal{L}$ and $\mathcal{K} \in \mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$ such that $\varphi(\mathcal{L}) = \mathbf{A_{vu}}$ and $\varphi(\mathcal{K}) = \mathbf{B_{vu}}$, and for all $\alpha \in \mathbb{K}$, we have:*

$$\varphi(\alpha \mathcal{L}) = \alpha \mathbf{A_{vu}} = \alpha \varphi(\mathcal{L}) \ , \ \varphi(\mathcal{L} + \mathcal{K}) = \mathbf{A_{vu}} + \mathbf{B_{vu}} = \varphi(\mathcal{L}) + \varphi(\mathcal{K}),$$

*that is, the mapping $\varphi$ preserves the operations of addition and multiplication by a scalar in the vector space $\mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$. Therefore, $\varphi$ is an isomorphism between the vector spaces $\mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$ and $\mathbb{K}^{I \times J}$.*

13 This v.s. was introduced in section 2.5.10.

*– $\mathcal{U}$ and $\mathcal{V}$ being two v.s. of the same dimension, with bases $\{\mathbf{u}_1, \ldots, \mathbf{u}_I\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_I\}$, the linear map $\mathcal{L} \in \mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$, with the associated matrix $\mathbf{A}_{\mathbf{vu}}$, is bijective if and only if $\mathbf{A}_{\mathbf{vu}}$ is invertible, and we have $\varphi(\mathcal{L}^{-1}) = [\varphi(\mathcal{L})]^{-1} = \mathbf{A}_{\mathbf{vu}}^{-1}$.*

PROPOSITION 4.53.– *Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be three $\mathbb{K}$-v.s., of respective dimensions $J$, $I$, and $K$. Consider two linear maps $\mathcal{L} : \mathcal{U} \to \mathcal{V}$ and $\mathcal{K} : \mathcal{V} \to \mathcal{W}$ with the associated matrices $\mathbf{A}_{\mathbf{vu}} \in \mathbb{K}^{I \times J}$ and $\mathbf{B}_{\mathbf{wv}} \in \mathbb{K}^{K \times I}$, then the composite map $\mathcal{K} \circ \mathcal{L}$ is itself a linear map of $\mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{W})$ with the associated matrix $\mathbf{C}_{\mathbf{wu}} = \mathbf{B}_{\mathbf{wv}}\mathbf{A}_{\mathbf{vu}} \in \mathbb{K}^{K \times J}$.*

PROOF.– Consider the bases $\{\mathbf{u}_1, \ldots, \mathbf{u}_J\}$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_I\}$, and $\{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ of $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$, respectively, and define the matrices $\varphi(\mathcal{K}) = \mathbf{B}_{\mathbf{wv}} = [b_{ki}]$ and $\varphi(\mathcal{L}) = \mathbf{A}_{\mathbf{vu}} = [a_{ij}]$ associated with the linear maps $\mathcal{K}$ and $\mathcal{L}$. The linearity property of these maps allows us to write:

$$\mathcal{K} \circ \mathcal{L}(\mathbf{u}_j) = \mathcal{K}\left(\sum_{i=1}^{I} a_{ij}\mathbf{v}_i\right) = \sum_{i=1}^{I} a_{ij}\mathcal{K}(\mathbf{v}_i)$$
$$= \sum_{i=1}^{I} a_{ij}\left(\sum_{k=1}^{K} b_{ki}\mathbf{w}_k\right) = \sum_{k=1}^{K}\left(\sum_{i=1}^{I} b_{ki}a_{ij}\right)\mathbf{w}_k.$$

Accordingly, the matrix associated with $\mathcal{K} \circ \mathcal{L}$ is such that $c_{kj} = \sum_{i=1}^{I} b_{ki}a_{ij}$, from which it can be concluded that $\mathbf{C}_{\mathbf{wu}} = \mathbf{B}_{\mathbf{wv}}\mathbf{A}_{\mathbf{vu}}$. □

From this result, one can deduce the following corollary.

COROLLARY 4.54.– *Let $\mathcal{U}$ be a $\mathbb{K}$-v.s., and $f$ an endomorphism of $\mathcal{U}$, with $\mathbf{B}_{\mathbf{u}}$ as associated matrix in the basis $\{\mathbf{u}\}$ of $\mathcal{U}$. Then, the matrix associated with $f^n = \underbrace{f \circ \cdots \circ f}_{n \text{ times}}$ is the matrix $\mathbf{B}_{\mathbf{u}}^n$.*

### 4.13.2. *Change of basis*

Let $\{\mathbf{u}\}$ and $\{\underline{\mathbf{u}}\}$ be two bases of the vector space $\mathcal{U}$, linked by the following formula:

$$\underline{\mathbf{u}}_j = \sum_{i=1}^{J} p_{ij}\mathbf{u}_i, \quad j \in \langle J \rangle. \tag{4.36}$$

Define the matrix $\mathbf{P} = [p_{ij}]$, of dimensions $J \times J$, called the change of basis matrix from $\{\mathbf{u}\}$ to $\{\underline{\mathbf{u}}\}$. The components of its $j$th column are the coordinates of $\underline{\mathbf{u}}_j$ in the basis $\{\mathbf{u}\}$, and the equation for the change of basis is written as:

$$\underline{\mathbf{U}} = \left[\underline{\mathbf{u}}_1 \cdots \underline{\mathbf{u}}_J\right] = \left[\mathbf{u}_1 \cdots \mathbf{u}_J\right]\mathbf{P} = \mathbf{U}\mathbf{P}.$$

$\mathbf{P}$ is invertible, and its inverse $\mathbf{P}^{-1}$ is the transformation matrix from $\{\underline{\mathbf{u}}\}$ to $\{\mathbf{u}\}$ such that:

$$\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 \cdots \mathbf{u}_J\end{bmatrix} = \begin{bmatrix}\underline{\mathbf{u}}_1 \cdots \underline{\mathbf{u}}_J\end{bmatrix}\mathbf{P}^{-1} = \underline{\mathbf{U}}\mathbf{P}^{-1}.$$

Its columns contain the coordinates of the vectors of the basis $\{\mathbf{u}\}$ in the basis $\{\underline{\mathbf{u}}\}$.

PROPOSITION 4.55.– *Given a vector $\mathbf{x} \in \mathcal{U}$, the coordinate vectors $\mathbf{x}_\mathbf{u}$ and $\mathbf{x}_{\underline{\mathbf{u}}}$ in the bases $\{\mathbf{u}\}$ and $\{\underline{\mathbf{u}}\}$ are related by the following relationship:*

$$\mathbf{x}_\mathbf{u} = \mathbf{P}\mathbf{x}_{\underline{\mathbf{u}}}. \qquad [4.37]$$

PROOF.– This property can be demonstrated by essentially following the same approach as the one used in the proof of Proposition 4.50, with the following correspondences: $(\mathbf{u}, \underline{\mathbf{u}}, \mathbf{x}_\mathbf{u}, \mathbf{x}_{\underline{\mathbf{u}}}, \mathbf{P}) \leftrightarrow (\mathbf{v}, \mathbf{u}, \mathbf{y}_\mathbf{v}, \mathbf{x}_\mathbf{u}, \mathbf{A})$. □

PROPOSITION 4.56.– *Consider the linear map $\mathcal{L} : \mathcal{U} \to \mathcal{V}$ and the changes of basis in $\mathcal{U}$ and $\mathcal{V}$ characterized by matrices $\mathbf{P}$ and $\mathbf{Q}$, respectively, and designate by $\mathbf{A}_{\mathbf{vu}}$ and $\mathbf{A}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}}$ the matrices associated with $\mathcal{L}$ in the bases $(\{\mathbf{u}\}, \{\mathbf{v}\})$ on the one hand, and $(\{\underline{\mathbf{u}}\}, \{\underline{\mathbf{v}}\})$ on the other. We have the following relationship:*

$$\mathbf{A}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^{-1}\mathbf{A}_{\mathbf{vu}}\mathbf{P}, \qquad [4.38]$$

*and it is said that the matrices $\mathbf{A}_{\mathbf{vu}}$ and $\mathbf{A}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}}$ are equivalent.*

PROOF.– Let $\mathbf{y} \in \mathcal{V}$. Just as relation [4.37] links the coordinate vectors $\mathbf{x}_\mathbf{u}$ and $\mathbf{x}_{\underline{\mathbf{u}}}$, the one linking the coordinate vectors $\mathbf{y}_\mathbf{v}$ and $\mathbf{y}_{\underline{\mathbf{v}}}$, in bases $\mathbf{v}$ and $\underline{\mathbf{v}}$, respectively, is given by:

$$\mathbf{y}_\mathbf{v} = \mathbf{Q}\mathbf{y}_{\underline{\mathbf{v}}} \qquad [4.39]$$

with $\begin{bmatrix}\underline{\mathbf{v}}_1 \cdots \underline{\mathbf{v}}_I\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1 \cdots \mathbf{v}_I\end{bmatrix}\mathbf{Q}$. Assuming that $\mathbf{y}$ is the image $\mathcal{L}(\mathbf{x})$ of a vector $\mathbf{x} \in \mathcal{U}$, the relations [4.33] and [4.37] give:

$$\mathbf{y}_\mathbf{v} = \mathbf{A}_{\mathbf{vu}}\mathbf{x}_\mathbf{u} = \mathbf{A}_{\mathbf{vu}}\mathbf{P}\mathbf{x}_{\underline{\mathbf{u}}}, \qquad [4.40]$$

and after the change of basis in $\mathcal{V}$, we deduce from [4.39] and [4.40]:

$$\begin{aligned}\mathbf{y}_{\underline{\mathbf{v}}} &= \mathbf{Q}^{-1}\mathbf{y}_\mathbf{v} = \mathbf{Q}^{-1}\mathbf{A}_{\mathbf{vu}}\mathbf{P}\mathbf{x}_{\underline{\mathbf{u}}} \\ &= \mathbf{A}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}}\mathbf{x}_{\underline{\mathbf{u}}},\end{aligned}$$

with:

$$\mathbf{A}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^{-1}\mathbf{A}_{\mathbf{vu}}\mathbf{P}, \qquad [4.41]$$

which demonstrates [4.38], also called an equivalence relation. □

From this result, it can be concluded that a linear map may be represented by different matrices according to the bases of the input ($\mathcal{U}$) and output ($\mathcal{V}$) spaces.

NOTE 4.57.– The multiplication on the right- (left-) hand side of the matrix of a linear map by a non-singular matrix is equivalent to a change of basis in the input (output) space of the map.

The previous results can be summarized using the diagram shown in Figure 4.3.

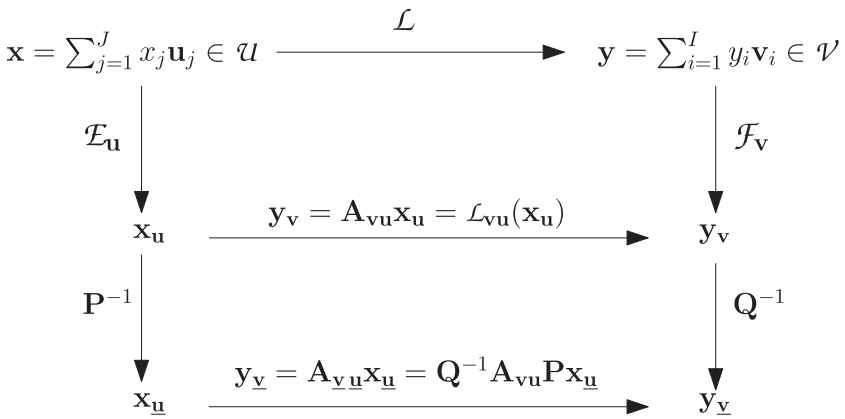



**Figure 4.3.** *Effect of changes of bases with a linear map* $\mathcal{L}$

### 4.13.3. *Endomorphisms*

As discussed in section 2.5.10.2, an endomorphism $f \in \mathcal{L}(\mathcal{U})$ is a linear map from $\mathcal{U}$ into itself. By choosing the basis $\{\mathbf{v}\}$ identical to $\{\mathbf{u}\}$, the matrix of an endomorphism in the basis $\{\mathbf{u}\}$ is a square matrix, denoted by $\mathbf{A}_{\mathbf{u}}$. Relation [4.41] then becomes:

$$\mathbf{A}_{\underline{\mathbf{u}}} = \mathbf{P}^{-1}\mathbf{A}_{\mathbf{u}}\mathbf{P}, \qquad [4.42]$$

where $\mathbf{P}$ is the change of basis matrix in $\mathcal{U}$. This relation is called a similarity transformation, and matrices $\mathbf{A}_{\mathbf{u}}$ and $\mathbf{A}_{\underline{\mathbf{u}}}$ are said to be similar.

An important problem in matrix calculus consists in finding the transformation matrix $\mathbf{P}$ allowing to reduce the matrix of a linear map to its simplest form possible, as for example the diagonal form. When there exists $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}_{\mathbf{u}}\mathbf{P}$ is diagonal, it is said that $\mathbf{A}_{\mathbf{u}}$ is diagonalizable. The matrix $\mathbf{P}$ is then obtained from

the eigendecomposition of $\mathbf{A}_{\mathbf{u}}$, its columns being formed of eigenvectors, while the diagonal elements of the diagonal matrix $\mathbf{A}_{\underline{\mathbf{u}}}$ are the eigenvalues. See section 4.16 for the definitions of eigenvalues and eigenvectors. The eigendecomposition of a matrix will be presented in Volume 2.

EXAMPLE 4.58.– Let the map $\mathcal{L} : \mathbb{K}^{J\times J} \to \mathbb{K}^{I\times J}$ be defined by $\mathbb{K}^{J\times J} \ni \mathbf{X} \mapsto \mathbf{Y} = \mathcal{L}(\mathbf{X}) = \mathbf{T}\mathbf{X} \in \mathbb{K}^{I\times J}$ where $\mathbf{T} \in \mathbb{K}^{I\times J}$ is a given matrix. $\mathcal{L}$ is a linear map because for all $\mathbf{X}_1$ and $\mathbf{X}_2$ of $\mathbb{K}^{J\times J}$ and for all $\alpha \in \mathbb{K}$, we have:

$$\mathcal{L}(\alpha\mathbf{X}_1 + \mathbf{X}_2) = \mathbf{T}(\alpha\mathbf{X}_1 + \mathbf{X}_2) = \alpha\mathbf{T}\mathbf{X}_1 + \mathbf{T}\mathbf{X}_2 = \alpha\mathcal{L}(\mathbf{X}_1) + \mathcal{L}(\mathbf{X}_2).$$

Consider the vectorized form of the equation $\mathbf{Y} = \mathbf{T}\mathbf{X}$, associated with the map $\mathbb{K}^{J^2} \to \mathbb{K}^{JI} : \mathbb{K}^{J^2} \ni \text{vec}(\mathbf{X}) \mapsto \text{vec}(\mathbf{Y}) = \text{vec}(\mathbf{T}\mathbf{X}) \in \mathbb{K}^{JI}$.

Considering the particular case $I = 3, J = 2$, and $\mathbf{T} = [t_{ij}] \in \mathbb{K}^{3\times 2}$, we are going to show that the matrix $\mathbf{A} \in \mathbb{K}^{JI\times J^2}$ which links $\mathbf{y}_{\mathbf{v}} = \text{vec}(\mathbf{Y})$ and $\mathbf{x}_{\mathbf{u}} = \text{vec}(\mathbf{X})$ is such that:

$$\text{vec}(\mathbf{Y}) = \mathbf{A}\,\text{vec}(\mathbf{X}) = (\mathbf{I}_2 \otimes \mathbf{T})\,\text{vec}(\mathbf{X}). \qquad [4.43]$$

The canonical bases of $\mathcal{U} = \mathbb{K}^{2\times 2}$ and of $\mathcal{V} = \mathbb{K}^{3\times 2}$ being composed of matrices $\{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_4\}$ and $\{\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4, \mathbf{V}_5, \mathbf{V}_6\}$ such that:

$$\mathbf{U}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ \mathbf{U}_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},\ \mathbf{U}_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\ \mathbf{U}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

and

$$\mathbf{V}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},\ \mathbf{V}_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix},\ \mathbf{V}_3 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix},\ \mathbf{V}_4 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\mathbf{V}_5 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},\ \mathbf{V}_6 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix},$$

the expansions of matrices $\mathbf{X}$ and $\mathbf{Y}$ over these bases are given by:

$$\mathbf{X} = \sum_{i,j=1}^{J} x_{ij}\mathbf{U}_{(j-1)J+i} \text{ and } \mathbf{Y} = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}\,\mathbf{V}_{(j-1)I+i}.$$

The operator vec applied to matrices $\mathbf{X}$ and $\mathbf{Y}$ corresponds to the mappings $\mathcal{E}_{\mathbf{u}}$ and $\mathcal{F}_{\mathbf{v}}$ of Figure 4.3, namely:

$$\mathbf{x}_{\mathbf{u}} = \mathcal{E}_{\mathbf{u}}(\mathbf{X}) = \text{vec}(\mathbf{X}) = [x_{11}, x_{21}, x_{12}, x_{22}]^T$$

$$\mathbf{y}_{\mathbf{v}} = \mathcal{F}_{\mathbf{v}}(\mathbf{Y}) = \text{vec}(\mathbf{Y}) = [y_{11}, y_{21}, y_{31}, y_{12}, y_{22}, y_{32}]^T.$$

The columns of $\mathbf{A}$ contain the coordinate vectors of the images $\mathcal{L}(\mathbf{U}_j)$, $j \in \langle 4 \rangle$, in the basis $\{\mathbf{V}\}$. For example, we have:

$$\mathcal{L}(\mathbf{U}_1) = \mathbf{T}\mathbf{U}_1 = \begin{bmatrix} t_{11} & 0 \\ t_{21} & 0 \\ t_{31} & 0 \end{bmatrix} = t_{11}\mathbf{V}_1 + t_{21}\mathbf{V}_2 + t_{31}\mathbf{V}_3$$

$$\Rightarrow \mathbf{A}_{\cdot 1} = [t_{11}, t_{21}, t_{31}, 0, 0, 0]^T.$$

We get:

$$\mathbf{A} = \begin{bmatrix} t_{11} & t_{12} & 0 & 0 \\ t_{21} & t_{22} & 0 & 0 \\ t_{31} & t_{32} & 0 & 0 \\ 0 & 0 & t_{11} & t_{12} \\ 0 & 0 & t_{21} & t_{22} \\ 0 & 0 & t_{31} & t_{32} \end{bmatrix} = \begin{bmatrix} \mathbf{T} & \mathbf{0}_{3\times 2} \\ \mathbf{0}_{3\times 2} & \mathbf{T} \end{bmatrix} = \mathbf{I}_2 \otimes \mathbf{T}. \qquad [4.44]$$

This matrix comes from the vectorization of $\mathbf{Y} = \mathbf{T}\mathbf{X}$ which gives $\mathbf{y}_\mathbf{v} = \mathrm{vec}(\mathbf{Y}) = (\mathbf{I}_2 \otimes \mathbf{T})\,\mathrm{vec}(\mathbf{X}) = (\mathbf{I}_2 \otimes \mathbf{T})\mathbf{x}_\mathbf{u}$, where the symbol $\otimes$ denotes the Kronecker product which is defined in section 5.4.3. This vectorization formula for $\mathbf{T}\mathbf{X}$ will be demonstrated in Volume 2.

It can be verified that for $\mathbf{X} = [x_{ij}] \in \mathbb{K}^{2\times 2}$, we have:

$$\mathbf{Y} = \mathbf{T}\mathbf{X} = \begin{bmatrix} t_{11}x_{11} + t_{12}x_{21} & t_{11}x_{12} + t_{12}x_{22} \\ t_{21}x_{11} + t_{22}x_{21} & t_{21}x_{12} + t_{22}x_{22} \\ t_{31}x_{11} + t_{32}x_{21} & t_{31}x_{12} + t_{32}x_{22} \end{bmatrix},$$

and consequently:

$$\mathbf{y}_\mathbf{v} = \mathrm{vec}(\mathbf{Y}) = \begin{bmatrix} t_{11}x_{11} + t_{12}x_{21} \\ t_{21}x_{11} + t_{22}x_{21} \\ t_{31}x_{11} + t_{32}x_{21} \\ t_{11}x_{12} + t_{12}x_{22} \\ t_{21}x_{12} + t_{22}x_{22} \\ t_{31}x_{12} + t_{32}x_{22} \end{bmatrix} = \mathbf{A} \begin{bmatrix} x_{11} \\ x_{21} \\ x_{12} \\ x_{22} \end{bmatrix} = \mathbf{A}\mathbf{x}_\mathbf{u},$$

with $\mathbf{A}$ defined in [4.44].

### 4.13.4. *Nilpotent endomorphisms*

It is said that $f \in \mathcal{L}(\mathcal{U})$ is a nilpotent endomorphism of index $n$ if there exists a smallest integer $n > 0$ such that $f^n = 0$, which amounts to saying that the composition of $f$ by itself $n$ times gives the zero morphism.

FACT 4.59.– The endomorphism $f \in \mathcal{L}(\mathcal{U})$ is nilpotent of index $n$ if and only if its associated matrix $\mathbf{A_u}$ is strictly upper triangular.

This results from the application of Corollary 4.54, which implies that the matrix associated with $f^n$ is equal to $(\mathbf{A_u})^n$. As a consequence, the matrix associated with a nilpotent endomorphism of index $n$ is a nilpotent matrix of index $n$ defined such that $(\mathbf{A_u})^n = \mathbf{0}$ (see the definition of a nilpotent matrix in Table 4.3).

### 4.13.5. *Equivalent, similar and congruent matrices*

We first recall the notions of equivalent and similar matrices, highlighted with a change of bases in the input and output spaces of a linear map, and a change of basis in the space of an endomorphism, respectively. Next, we define congruent matrices.

Two matrices $\mathbf{A}$ and $\mathbf{B} \in \mathbb{K}^{I \times J}$ are said to be equivalent when there exists two non-singular matrices $\mathbf{P} \in \mathbb{K}^{J \times J}$ and $\mathbf{Q} \in \mathbb{K}^{I \times I}$ such that:

$$\mathbf{B} = \mathbf{QAP}. \qquad [4.45]$$

Two matrices $\mathbf{A}$ and $\mathbf{B} \in \mathbb{K}^{I \times I}$ are called similar if there exists a non-singular matrix $\mathbf{P} \in \mathbb{K}^{I \times I}$ such that:

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP} \text{ or } \mathbf{PB} = \mathbf{AP}. \qquad [4.46]$$

PROPOSITION 4.60.– *It is easy to deduce the following properties:*

*– If* $\mathbf{A}$ *and* $\mathbf{B}$ *are equivalent, then* $r(\mathbf{A}) = r(\mathbf{B})$.

*– If* $\mathbf{A}$ *and* $\mathbf{B}$ *are similar, then* $r(\mathbf{A}) = r(\mathbf{B})$, *and* $tr(\mathbf{A}) = tr(\mathbf{B})$, *because* $tr(\mathbf{B}) = tr(\mathbf{P}^{-1}\mathbf{AP}) = tr(\mathbf{PP}^{-1}\mathbf{A}) = tr(\mathbf{A})$.

*In section 4.16.7, we shall see that two simular matrices have the same eigenvalues.*

$\mathbf{A}$ and $\mathbf{B}$ are said to be orthogonally similar if there exists an orthogonal matrix $\mathbf{P}$ such that:

$$\mathbf{B} = \mathbf{P}^T\mathbf{AP}, \quad \text{with} \quad \mathbf{P}^{-1} = \mathbf{P}^T.$$

Similarly, matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be unitarily similar if there exists a unitary matrix $\mathbf{P}$ such that:

$$\mathbf{B} = \mathbf{P}^H\mathbf{AP}, \quad \text{with} \quad \mathbf{P}^{-1} = \mathbf{P}^H.$$

$\mathbf{A}$ and $\mathbf{B}$ are said to be congruent if there exists a non-singular matrix $\mathbf{P}$ such that:

$$\mathbf{B} = \mathbf{P}^T\mathbf{A}\mathbf{P}.$$

This equation is called a congruence transformation.

FACT 4.61.– Equivalent, similar, and congruent matrices are such that:

– Two rectangular (square) matrices are equivalent (similar) if they can represent the same linear map (the same endomorphism) in different bases.

– Two matrices are congruent if they can represent the same symmetric bilinear form in two different bases (see section 4.14.5). They are also equivalent and therefore of the same rank.

NOTE 4.62.– Any regular matrix $\mathbf{P}$ can be written as the product of a finite number of matrices that represent elementary operations involving the rows (columns) of a matrix $\mathbf{A}$, that is, in the form $\mathbf{PA}$ ($\mathbf{AP}$). These matrices will be detailed in section 5.13. Accordingly, we can say that two matrices are equivalent if and only if one can be obtained from the other through a sequence of elementary operations acting on its rows and columns. Similarly, two matrices are called orthogonally (respectively, unitarily) similar if and only if one can be obtained from the other through a sequence of elementary operations on its rows and columns, under the constraint that matrices representing these two sequences of operations are transposed (respectively conjugate transposed) one relatively to the other.

## 4.14. Matrix associated with a bilinear/sesquilinear form

In this section, we focus on the matrix representations of bilinear and sesquilinear forms on finite-dimensional inner product spaces over $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$, respectively. We are going to first define a bilinear and a sesquilinear map, before defining bilinear and sesquilinear forms as special cases. How changes of bases in the v.s. transform the matrices associated to these forms will be particularly highlighted, along with the special cases of symmetric forms.

### 4.14.1. *Definition of a bilinear/sesquilinear map*

Let $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ be three $\mathbb{R}$-v.s., of respective dimensions $J$, $I$, and $K$. A bilinear map $f : \mathcal{U} \times \mathcal{V} \to \mathcal{W}$, such that $\mathcal{U} \times \mathcal{V} \ni (x, y) \mapsto f(x, y) \in \mathcal{W}$, is linear with respect to each of the variables $x$ and $y$, when the other variable ($y$ and $x$, respectively) is fixed.

Similarly, when $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ are three $\mathbb{C}$-v.s., the map $f : \mathcal{U} \times \mathcal{V} \to \mathcal{W}$ is said to be sesquilinear if it is linear with respect to the first variable $x \in \mathcal{U}$, and semilinear with respect to the second variable $y \in \mathcal{V}$.

We thus have the following properties:

– $f$ is additive with respect to the two variables, that is, for all $x, x_1, x_2 \in \mathcal{U}$ and $y, y_1, y_2 \in \mathcal{V}$, we have:

$$f(x_1 + x_2, y) = f(x_1, y) + f(x_2, y),$$
$$f(x, y_1 + y_2) = f(x, y_1) + f(x, y_2).$$

– $f$ is homogeneous in the first variable

$$f(\alpha\, x, y) = \alpha f(x, y), \quad \alpha \in \mathbb{K}.$$

– In the case of a bilinear map ($\mathbb{K} = \mathbb{R}$), $f$ is homogeneous in the second variable

$$f(x, \alpha\, y) = \alpha f(x, y), \quad \alpha \in \mathbb{R}.$$

– In the case of a sesquilinear map ($\mathbb{K} = \mathbb{C}$), $f$ is conjugate-homogeneous in the second variable

$$f(x, \alpha\, y) = \alpha^* f(x, y), \quad \alpha \in \mathbb{C}.$$

FACT 4.63.– The set of bilinear maps, denoted by $\mathcal{BL}(\mathcal{U}, \mathcal{V}; \mathcal{W})$, is a $\mathbb{K}$-v.s. of dimension: $\dim[\mathcal{BL}(\mathcal{U}, \mathcal{V}; \mathcal{W})] = \dim[\mathcal{U}] \dim[\mathcal{V}] \dim[\mathcal{W}]$.

In the following, we consider the case $\mathcal{W} = \mathbb{K}$, corresponding to a bilinear form if $\mathbb{K} = \mathbb{R}$, and a sesquilinear form if $\mathbb{K} = \mathbb{C}$. Vectors $x$ and $y$ of v.s. $\mathcal{U}$ and $\mathcal{V}$ will be denoted in bold to distinguish them from their scalar components in a given basis. The notation $(x, y)$ will be used instead of $(\mathbf{x}, \mathbf{y})$ when there is no ambiguity.

FACT 4.64.– The set of bilinear forms, denoted by $\mathcal{BL}(\mathcal{U}, \mathcal{V}; \mathbb{K})$, is a $\mathbb{K}$-v.s. of dimension: $\dim[\mathcal{BL}(\mathcal{U}, \mathcal{V}; \mathbb{K})] = \dim[\mathcal{U}] \dim[\mathcal{V}]$.

EXAMPLE 4.65.– Given a matrix $\mathbf{B} \in \mathbb{R}^{I \times I}$, the following expression:

$$\mathbf{y}^T \mathbf{B} \mathbf{x} = \sum_{i,j=1}^{I} b_{ij} y_i x_j \qquad [4.47]$$

is a bilinear form on the vector space $\mathbb{R}^I$, that is, $f : \mathbb{R}^I \times \mathbb{R}^I \to \mathbb{R}$, such that $\mathbb{R}^I \times \mathbb{R}^I \ni (\mathbf{x}, \mathbf{y}) \mapsto f(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T \mathbf{B} \mathbf{x} \in \mathbb{R}$.

In the next section, we shall see that all bilinear forms can be represented by means of an equation like [4.47].

### 4.14.2. *Matrix associated to a bilinear/sesquilinear form*

Consider the $\mathbb{R}$-v.s. $\mathcal{U}$ and $\mathcal{V}$, of dimensions $J$ and $I$, with the respective ordered bases $\{\mathbf{u}\} = \{\mathbf{u}_1, \cdots, \mathbf{u}_J\}$ and $\{\mathbf{v}\} = \{\mathbf{v}_1, \cdots, \mathbf{v}_I\}$. The vectors $\mathbf{x} \in \mathcal{U}$ and $\mathbf{y} \in \mathcal{V}$ can then be written as:

$$\mathbf{x} = \sum_{j=1}^{J} x_j \mathbf{u}_j, \quad \mathbf{y} = \sum_{i=1}^{I} y_i \mathbf{v}_i.$$

Using the bilinearity property of the bilinear form $f : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ gives:

$$f(\mathbf{x}, \mathbf{y}) = f\left(\sum_{j=1}^{J} x_j \mathbf{u}_j, \sum_{i=1}^{I} y_i \mathbf{v}_i\right) = \sum_{i=1}^{I} \sum_{j=1}^{J} x_j y_i f(\mathbf{u}_j, \mathbf{v}_i). \quad [4.48]$$

By defining the matrix $\mathbf{B_{vu}} \in \mathbb{R}^{I \times J}$ such that $(\mathbf{B_{vu}})_{ij} = f(\mathbf{u}_j, \mathbf{v}_i) = [b_{ij}]$, $f(\mathbf{x}, \mathbf{y})$ can be rewritten as:

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{I} \sum_{j=1}^{J} b_{ij} x_j y_i = \mathbf{y}_\mathbf{v}^T \mathbf{B_{vu}} \mathbf{x_u} = \mathbf{x}_\mathbf{u}^T \mathbf{B}_\mathbf{vu}^T \mathbf{y_v}, \quad [4.49]$$

where $\mathbf{x_u} = [x_1, \cdots, x_J]^T$ and $\mathbf{y_v} = [y_1, \cdots, y_I]^T$ are the vectors of coordinates of $\mathbf{x}$ and $\mathbf{y}$ in the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$, respectively. The matrix $\mathbf{B_{vu}}$ is called the matrix associated with the bilinear form $f$, relative to the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$.

FACT 4.66.– As previously seen for multilinear forms in section 2.5.11.2, the set $\mathcal{BL}(\mathcal{U}, \mathcal{V}; \mathbb{R})$ of bilinear forms from $\mathcal{U} \times \mathcal{V}$ to $\mathbb{R}$, with $\dim(\mathcal{U}) = J$ and $\dim(\mathcal{V}) = I$, is an $\mathbb{R}$-v.s. of dimension $IJ$. The mapping which to a bilinear form $f$ associates its matrix $\mathbf{B_{vu}} \in \mathbb{R}^{I \times J}$ in the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$ of $\mathcal{U}$ and $\mathcal{V}$, is an isomorphism of v.s.

NOTE 4.67.– If $\mathcal{U}$ and $\mathcal{V}$ are assumed to be of dimension $I$ and $J$, instead of $J$ and $I$, equation [4.49] then becomes:

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{I} \sum_{j=1}^{J} x_i y_j f(\mathbf{u}_i, \mathbf{v}_j) = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} x_i y_j = \mathbf{x}_\mathbf{u}^T \mathbf{A_{uv}} \mathbf{y_v} = \mathbf{y}_\mathbf{v}^T \mathbf{A}_\mathbf{uv}^T \mathbf{x_u}$$

$$\left(\mathbf{A_{uv}}\right)_{ij} = f(\mathbf{u}_i, \mathbf{v}_j) = [a_{ij}] \in \mathbb{R}^{I \times J}.$$

### 4.14.3. *Changes of bases with a bilinear form*

Let the changes of bases be defined as:

$$\underline{\mathbf{U}} = [\underline{\mathbf{u}}_1, \cdots, \underline{\mathbf{u}}_J] = [\mathbf{u}_1, \cdots, \mathbf{u}_J]\mathbf{P} = \mathbf{UP}, \quad [4.50]$$

$$\underline{\mathbf{V}} = [\underline{\mathbf{v}}_1, \cdots, \underline{\mathbf{v}}_I] = [\mathbf{v}_1, \cdots, \mathbf{v}_I]\mathbf{Q} = \mathbf{VQ}. \quad [4.51]$$

We have:

$$\mathbf{x} = \mathbf{U}\mathbf{x}_{\mathbf{u}} = \underline{\mathbf{U}}\,\mathbf{x}_{\underline{\mathbf{u}}} = \mathbf{U}\mathbf{P}\mathbf{x}_{\underline{\mathbf{u}}},$$
$$\mathbf{y} = \mathbf{V}\mathbf{y}_{\mathbf{v}} = \underline{\mathbf{V}}\,\mathbf{y}_{\underline{\mathbf{v}}} = \mathbf{V}\mathbf{Q}\mathbf{y}_{\underline{\mathbf{v}}},$$

from which are inferred the relations: $\mathbf{x}_{\mathbf{u}} = \mathbf{P}\mathbf{x}_{\underline{\mathbf{u}}}$, $\mathbf{y}_{\mathbf{v}} = \mathbf{Q}\mathbf{y}_{\underline{\mathbf{v}}}$, where $\mathbf{x}_{\underline{\mathbf{u}}}$ and $\mathbf{y}_{\underline{\mathbf{v}}}$ are the vectors of coordinates of $\mathbf{x}$ and $\mathbf{y}$ in the bases $\underline{\mathbf{u}}$ and $\underline{\mathbf{v}}$, and $\mathbf{P}$ and $\mathbf{Q}$ are the change of basis matrices from $\mathbf{u}$ to $\underline{\mathbf{u}}$ and from $\mathbf{v}$ to $\underline{\mathbf{v}}$, respectively.

After the changes of bases, expression [4.49] of the bilinear form becomes:

$$\begin{aligned} f(\mathbf{x},\mathbf{y}) &= \mathbf{y}_{\underline{\mathbf{v}}}^T \mathbf{Q}^T \mathbf{B}_{\mathbf{vu}} \mathbf{P} \mathbf{x}_{\underline{\mathbf{u}}} \\ &= \mathbf{y}_{\underline{\mathbf{v}}}^T \mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \end{aligned} \qquad [4.52]$$

where

$$\mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^T \mathbf{B}_{\mathbf{vu}} \mathbf{P} \qquad [4.53]$$

is the matrix associated to $f$ relatively to the bases $\{\underline{\mathbf{u}}\}$ and $\{\underline{\mathbf{v}}\}$. Matrices $\mathbf{B}_{\mathbf{vu}}$ and $\mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}}$, which represent the same bilinear form $f$ in two different sets of bases, are equivalent according to the definition [4.45], and therefore have the same rank. This rank is called the rank of $f$.

### 4.14.4. *Changes of bases with a sesquilinear form*

In the case where $\mathcal{U}$ and $\mathcal{V}$ are two $\mathbb{C}$-v.s., with the respective bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$, equations [4.48] and [4.49] for a sesquilinear form become:

$$\begin{aligned} f(\mathbf{x},\mathbf{y}) &= \sum_{i=1}^{I} \sum_{j=1}^{J} x_j y_i^* f(\mathbf{u}_j, \mathbf{v}_i) \\ &= \mathbf{y}_{\mathbf{v}}^H \mathbf{B}_{\mathbf{vu}} \mathbf{x}_{\mathbf{u}}, \end{aligned} \qquad [4.54]$$

and after the changes of bases [4.50] and [4.51], the relation [4.53] becomes:

$$\mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^H \mathbf{B}_{\mathbf{vu}} \mathbf{P}. \qquad [4.55]$$

In Table 4.15, we summarize the main results established for matrices associated with a linear map, a bilinear form, and a sesquilinear form.

| Linear map | Bilinear form | Sesquilinear form |
|---|---|---|
| $\mathcal{L} : \mathcal{U} \to \mathcal{V}$ | $f : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ | $f : \mathcal{U} \times \mathcal{V} \to \mathbb{C}$ |
| **Matrix associated to $\mathcal{L}$** | **Matrix associated to $f$** | **Matrix associated to $f$** |
| $\mathbf{A}_{\cdot j} = \mathcal{L}(\mathbf{u}_j)$<br>$\mathbf{y_v} = \sum_{j=1}^{J} x_j \mathcal{L}(\mathbf{u}_j) = \mathbf{A_{vu}x_u}$ | $(\mathbf{B_{vu}})_{ij} = f(\mathbf{u}_j, \mathbf{v}_i)$<br>$f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_\mathbf{v}^T \mathbf{B_{vu}x_u}$ | $(\mathbf{B_{vu}})_{ij} = f(\mathbf{u}_j, \mathbf{v}_i)$<br>$f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_\mathbf{v}^H \mathbf{B_{vu}x_u}$ |
| **Changes of bases** | | |
| | $\begin{cases} \underline{\mathbf{U}} = \mathbf{UP} \\ \underline{\mathbf{V}} = \mathbf{VQ} \end{cases}$ | |
| $\mathbf{A}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^{-1}\mathbf{A_{vu}P}$ | $\mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^T\mathbf{B_{vu}P}$ | $\mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^H\mathbf{B_{vu}P}$ |

**Table 4.15.** *Matrices associated with a linear map, a bilinear, and a sesquilinear form*

### 4.14.5. *Symmetric bilinear/sesquilinear forms*

From now on, we consider that both $\mathbb{R}$-v.s. $\mathcal{U}$ and $\mathcal{V}$ are identical and $\{\mathbf{v}\} = \{\mathbf{u}\}$. The matrix associated with the bilinear form $f : \mathcal{U}^2 \to \mathbb{R}$, denoted by $\mathbf{B_u}$, is then square, and equation [4.49] becomes:

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_\mathbf{u}^T \mathbf{B_u x_u}. \qquad [4.56]$$

After the basis change with matrix $\mathbf{P}$, equations [4.52] and [4.53] can be written as:

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_{\underline{\mathbf{u}}}^T \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \qquad [4.57]$$

$$\mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^T \mathbf{B_u P}. \qquad [4.58]$$

Matrices $\mathbf{B_u}$ and $\mathbf{B}_{\underline{\mathbf{u}}}$ which represent $f$ in two different bases are said to be congruent.

For a sesquilinear form $f : \mathcal{U}^2 \to \mathbb{C}$, where $\mathcal{U}$ is a $\mathbb{C}$-v.s., equations [4.56]–[4.58] become:

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_\mathbf{u}^H \mathbf{B_u x_u} \qquad [4.59]$$

$$= \mathbf{y}_{\underline{\mathbf{u}}}^H \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \qquad [4.60]$$

$$\mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^H \mathbf{B_u P}. \qquad [4.61]$$

Matrices $\mathbf{B_u}$ and $\mathbf{B}_{\underline{\mathbf{u}}}$ are again congruent.

We now consider symmetric/antisymmetric bilinear forms and conjugate symmetric/antisymmetric sesquilinear forms.

It is said that the bilinear form $f : \mathcal{U}^2 \to \mathbb{R}$ is:

– symmetric, if: $\forall(\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = f(\mathbf{x}, \mathbf{y})$;

– antisymmetric, if: $\forall(\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = -f(\mathbf{x}, \mathbf{y})$.

If $\mathbf{B_u}$ is the matrix associated with $f$ in the basis $\{\mathbf{u}\}$, it is then easy to verify from [4.56] that $\mathbf{B_u}$ is such that:

$$f \text{ symmetric} \iff \mathbf{B}_\mathbf{u}^T = \mathbf{B_u}$$
$$f \text{ antisymmetric} \iff \mathbf{B}_\mathbf{u}^T = -\mathbf{B_u}$$

in other words, $\mathbf{B_u}$ is respectively symmetric and antisymmetric.

NOTE 4.68.– According to definition [2.10] of an alternating multilinear form, it can be concluded that the bilinear form $f$ is alternating if: $\forall \mathbf{x} \in \mathcal{U}$, we have $f(\mathbf{x}, \mathbf{x}) = 0$.

PROPOSITION 4.69.– *Any alternating bilinear form is antisymmetric.*

Indeed, if $f$ is an alternating bilinear form, we have:

$$f(\mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}) = f(\mathbf{x}, \mathbf{x}) + f(\mathbf{y}, \mathbf{y}) + f(\mathbf{x}, \mathbf{y}) + f(\mathbf{y}, \mathbf{x}) = f(\mathbf{x}, \mathbf{y}) + f(\mathbf{y}, \mathbf{x}) = 0$$
$$\Downarrow$$
$$f(\mathbf{y}, \mathbf{x}) = -f(\mathbf{x}, \mathbf{y}),$$

which proves that $f$ is antisymmetric.

In the case of a sesquilinear form $f : \mathcal{U}^2 \to \mathbb{C}$, it is said that $f$ has:

– a Hermitian symmetry if: $\forall(\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = f^*(\mathbf{x}, \mathbf{y})$;

– a skew-Hermitian (also called antihermitian) symmetry if: $\forall(\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = -f^*(\mathbf{x}, \mathbf{y})$.

Based on [4.59], we can deduce that $\mathbf{B_u}$ is such that:

$$f \text{ of Hermitian symmetry} \iff \mathbf{B}_\mathbf{u}^H = \mathbf{B_u}$$
$$f \text{ of skew-Hermitian symmetry} \iff \mathbf{B}_\mathbf{u}^H = -\mathbf{B_u}.$$

PROOF.– In the case of the Hermitian symmetry, we have:

$$f(\mathbf{y}, \mathbf{x}) = f^*(\mathbf{x}, \mathbf{y}) \Rightarrow \mathbf{x}_\mathbf{u}^H \mathbf{B_u} \mathbf{y_u} = (\mathbf{y}_\mathbf{u}^H \mathbf{B_u} \mathbf{x_u})^* = \mathbf{y}_\mathbf{u}^T \mathbf{B}_\mathbf{u}^* \mathbf{x}_\mathbf{u}^*$$
$$= \mathbf{x}_\mathbf{u}^H \mathbf{B}_\mathbf{u}^H \mathbf{y_u} \text{ (by transposition of a scalar)}$$

from which it is deduced that $\mathbf{B}_\mathbf{u}^H = \mathbf{B_u}$.

In the case of antihermitian symmetry, we have:

$$f(\mathbf{y},\mathbf{x}) = -f^*(\mathbf{x},\mathbf{y}) \;\Rightarrow\; \mathbf{B}_\mathbf{u}^H = -\mathbf{B}_\mathbf{u} \qquad \square$$

It is said that $\mathbf{B}_\mathbf{u}$ is Hermitian ($\mathbf{B}_\mathbf{u}^H = \mathbf{B}_\mathbf{u}$) and antihermitian ($\mathbf{B}_\mathbf{u}^H = -\mathbf{B}_\mathbf{u}$).

PROPOSITION 4.70.– *Any bilinear form $f$ with the associated matrix $\mathbf{B}_\mathbf{u}$ can be decomposed into the sum of a symmetric bilinear form with an antisymmetric bilinear form:*

$$f(x,y) = \frac{1}{2}[f(x,y) + f(y,x)] + \frac{1}{2}[f(x,y) - f(y,x)]$$

*where the first term in square brackets is a symmetric bilinear form, the second being an antisymmetric bilinear form. From this decomposition, it results that any real square matrix $\mathbf{B}_\mathbf{u}$ can be decomposed into the sum of a symmetric matrix and an antisymmetric matrix such as:*

$$\mathbf{B}_\mathbf{u} = \frac{1}{2}[\mathbf{B}_\mathbf{u} + \mathbf{B}_\mathbf{u}^T] + \frac{1}{2}[\mathbf{B}_\mathbf{u} - \mathbf{B}_\mathbf{u}^T].$$

NOTE 4.71.– Similarly, any complex square matrix can be decomposed into the sum of a Hermitian matrix and an antihermitian matrix, by replacing the transposition by the conjugate transposition, in the previous equation.

## 4.15. Quadratic forms and Hermitian forms

In this section, we consider quadratic forms deduced from symmetric bilinear forms and Hermitian forms associated with symmetric sesquilinear forms.

### 4.15.1. *Quadratic forms*

Given a symmetric bilinear form $f$ over a $\mathbb{R}$-v.s. $\mathcal{U}$, the quadratic form associated with $f$ designates the mapping $q : \mathcal{U} \to \mathbb{R}$ defined by:

$$\begin{aligned} q(\mathbf{x}) &= f(\mathbf{x},\mathbf{x}) \\ &= x_\mathbf{u}^T \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}, \quad \text{according to} \quad [4.56] \end{aligned}$$

the symmetry property of $f$ implying that: $\mathbf{B}_\mathbf{u}^T = \mathbf{B}_\mathbf{u}$. The symmetric bilinear form $f$ associated with the quadratic form $q$ is called the polar form of $q$, and the matrix $\mathbf{B}_\mathbf{u}$ is called the matrix of the quadratic form $q$ in the basis $\{\mathbf{u}\}$. The rank of $\mathbf{B}_\mathbf{u}$ is the rank of both $f$ and $q$. By defining the coordinate vector $\mathbf{x}_\mathbf{u}^T = [x_1, \cdots, x_J]$,

with $\mathbf{B_u} = [b_{ij}]$, $i, j \in \langle J \rangle$, the quadratic form can be developed as:

$$q(\mathbf{x}) = \sum_{i=1}^{J} \sum_{j=1}^{J} b_{ij} x_i x_j, \quad \text{with} \quad b_{ji} = b_{ij} \qquad [4.62]$$

$$= \sum_{i=1}^{J} b_{ii} x_i^2 + 2 \sum_{1 \leq i < j \leq J} b_{ij} x_i x_j. \qquad [4.63]$$

$q(\mathbf{x})$ is thus a homogeneous polynomial of second degree in the coordinates $\{x_j, j \in \langle J \rangle\}$ of vector $\mathbf{x}$ in the basis $\{\mathbf{u}\}$.

NOTE 4.72.– As it can be easily verified, shifting from one quadratic form $q$ to its polar form $f$ can be performed using the following rule:

– the terms $x_i^2$ are replaced by $x_i y_i$;

– the terms $x_i x_j$ are replaced by $\frac{1}{2}[x_i y_j + x_j y_i]$,

which gives:

$$f(x, y) = \sum_{i=1}^{J} b_{ii} x_i y_i + \sum_{1 \leq i < j \leq J} b_{ij} (x_i y_j + x_j y_i). \qquad [4.64]$$

By exploiting the bilinearity and symmetry properties of $f$, it can be deduced that:

$$q(\lambda \mathbf{x}) = f(\lambda \mathbf{x}, \lambda \mathbf{x}) = \lambda^2 q(\mathbf{x})$$

$$q(\mathbf{x} + \mathbf{y}) = f(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = q(\mathbf{x}) + 2\, f(\mathbf{x}, \mathbf{y}) + q(\mathbf{y}) \qquad [4.65]$$

$$q(\mathbf{x} - \mathbf{y}) = f(\mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y}) = q(\mathbf{x}) - 2\, f(\mathbf{x}, \mathbf{y}) + q(\mathbf{y}). \qquad [4.66]$$

PROPOSITION 4.73.– *The symmetric bilinear form $f$ associated with the quadratic form $q$ can be determined using the following formulae:*

$$f(\mathbf{x}, \mathbf{y}) = \frac{1}{4}[q(\mathbf{x} + \mathbf{y}) - q(\mathbf{x} - \mathbf{y})] \qquad [4.67]$$

$$= \frac{1}{2}[q(\mathbf{x} + \mathbf{y}) - q(\mathbf{x}) - q(\mathbf{y})] \qquad [4.68]$$

$$= \frac{1}{2}[q(\mathbf{x}) + q(\mathbf{y}) - q(\mathbf{x} - \mathbf{y})]. \qquad [4.69]$$

PROOF.– These formulae can be derived from [4.65] and [4.66]. □

NOTE 4.74.– These formulae are related to polarization formulae [3.42] and [3.43] that correspond to the case of a positive definite quadratic form, interpreted as the square of a norm (see section 4.15.4).

### 4.15.2. *Hermitian forms*

In the case of a sesquilinear form with Hermitian symmetry, we call Hermitian form associated to $f$ the map $q : \mathcal{U} \to \mathbb{R}$ defined by:

$$\begin{aligned} q(\mathbf{x}) &= f(\mathbf{x}, \mathbf{x}) \\ &= \mathbf{x}_{\mathbf{u}}^H \mathbf{B}_{\mathbf{u}} \mathbf{x}_{\mathbf{u}}, \quad \text{(following [4.54])}. \end{aligned}$$

Hermitian symmetry induces that: $\mathbf{B}_{\mathbf{u}}^H = \mathbf{B}_{\mathbf{u}}$, which implies that the diagonal coefficients of $\mathbf{B}_{\mathbf{u}}$ are real. In addition, we have:

$$\forall \mathbf{x} \in \mathcal{U},\ f(\mathbf{x}, \mathbf{x}) = f^*(\mathbf{x}, \mathbf{x}) \implies q(\mathbf{x}) = f(\mathbf{x}, \mathbf{x}) \in \mathbb{R}. \qquad [4.70]$$

The matrix $\mathbf{B}_{\mathbf{u}}$ of $f$ in the basis $\{\mathbf{u}\}$ is called the matrix of the Hermitian form $q$, and the rank of $\mathbf{B}_{\mathbf{u}}$ is also the rank of $f$ and $q$.

The relation [4.63] then becomes:

$$q(\mathbf{x}) = \sum_{i=1}^{J} b_{ii} |x_i|^2 + 2 \sum_{1 \le i < j \le J} \mathrm{Re}\left(b_{ij} x_i x_j^*\right).$$

Using the symmetry property $f(\mathbf{y}, \mathbf{x}) = f^*(\mathbf{x}, \mathbf{y})$, we have, with $j^2 = -1$:

$$q(\lambda \mathbf{x}) = f(\lambda \mathbf{x}, \lambda \mathbf{x}) = |\lambda|^2 q(\mathbf{x})$$

$$q(\mathbf{x} + \mathbf{y}) = f(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = q(\mathbf{x}) + f(\mathbf{x}, \mathbf{y}) + f^*(\mathbf{x}, \mathbf{y}) + q(\mathbf{y}) \qquad [4.71]$$

$$q(\mathbf{x} - \mathbf{y}) = f(\mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y}) = q(\mathbf{x}) - f(\mathbf{x}, \mathbf{y}) - f^*(\mathbf{x}, \mathbf{y}) + q(\mathbf{y}) \qquad [4.72]$$

$$q(\mathbf{x} + j\mathbf{y}) = f(\mathbf{x} + j\mathbf{y}, \mathbf{x} + j\mathbf{y}) = q(\mathbf{x}) - jf(\mathbf{x}, \mathbf{y}) + jf^*(\mathbf{x}, \mathbf{y}) + q(\mathbf{y}) \qquad [4.73]$$

$$q(\mathbf{x} - j\mathbf{y}) = f(\mathbf{x} - j\mathbf{y}, \mathbf{x} - j\mathbf{y}) = q(\mathbf{x}) + jf(\mathbf{x}, \mathbf{y}) - jf^*(\mathbf{x}, \mathbf{y}) + q(\mathbf{y}). \qquad [4.74]$$

Taking property [4.70] into account, equations [4.71]–[4.74] give:

$$q(\mathbf{x} + \mathbf{y}) = q(\mathbf{x}) + 2\,\mathrm{Re}\Big[f(\mathbf{x}, \mathbf{y})\Big] + q(\mathbf{y}) \in \mathbb{R} \qquad [4.75]$$

$$q(\mathbf{x} - \mathbf{y}) = q(\mathbf{x}) - 2\,\mathrm{Re}\Big[f(\mathbf{x}, \mathbf{y})\Big] + q(\mathbf{y}) \in \mathbb{R} \qquad [4.76]$$

$$q(\mathbf{x} + j\mathbf{y}) = q(\mathbf{x}) + 2\,\mathrm{Im}\Big[f(\mathbf{x}, \mathbf{y})\Big] + q(\mathbf{y}) \in \mathbb{R} \qquad [4.77]$$

$$q(\mathbf{x} - j\mathbf{y}) = q(\mathbf{x}) - 2\,\mathrm{Im}\Big[f(\mathbf{x}, \mathbf{y})\Big] + q(\mathbf{y}) \in \mathbb{R}. \qquad [4.78]$$

From these identities, it is easy to deduce the following proposition.

PROPOSITION 4.75.– *The sesquilinear form $f$ associated with the Hermitian form $q$ can be determined using the following formula:*

$$f(\mathbf{x},\mathbf{y}) = \frac{1}{4}[q(\mathbf{x}+\mathbf{y}) - q(\mathbf{x}-\mathbf{y})] + \frac{j}{4}[q(\mathbf{x}+j\mathbf{y}) - q(\mathbf{x}-j\mathbf{y})]. \qquad [4.79]$$

In Table 4.16, we summarize the different formulae of change of bases highlighted for non-symmetric and symmetric bilinear/sesquilinear forms and for quadratic/Hermitian forms.

| Form | Bilinear/quadratic | Sesquilinear/Hermitian |
|---|---|---|
| non-sym. bil/sesq | $\begin{cases} f(\mathbf{x},\mathbf{y}) = \mathbf{y}_{\underline{\mathbf{v}}}^T \mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^T \mathbf{B}_{\mathbf{vu}} \mathbf{P} \end{cases}$ | $\begin{cases} f(\mathbf{x},\mathbf{y}) = \mathbf{y}_{\underline{\mathbf{v}}}^H \mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^H \mathbf{B}_{\mathbf{vu}} \mathbf{P} \end{cases}$ |
| sym. bil/sesq | $\begin{cases} f(\mathbf{x},\mathbf{y}) = \mathbf{y}_{\underline{\mathbf{u}}}^T \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^T \mathbf{B}_{\mathbf{u}} \mathbf{P} \end{cases}$ | $\begin{cases} f(\mathbf{x},\mathbf{y}) = \mathbf{y}_{\underline{\mathbf{u}}}^H \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^H \mathbf{B}_{\mathbf{u}} \mathbf{P} \end{cases}$ |
| quadrat/Hermit | $\begin{cases} q(\mathbf{x}) = \mathbf{x}_{\underline{\mathbf{u}}}^T \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^T \mathbf{B}_{\mathbf{u}} \mathbf{P} \end{cases}$ | $\begin{cases} q(\mathbf{x}) = \mathbf{x}_{\underline{\mathbf{u}}}^H \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^H \mathbf{B}_{\mathbf{u}} \mathbf{P} \end{cases}$ |

**Table 4.16.** *Change of basis formulae for bilinear/sesquilinear forms and quadratic/Hermitian forms*

### 4.15.3. *Positive/negative definite quadratic/Hermitian forms*

A quadratic form $q$ is said to be positive if $q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{U}$. It is said to be positive definite if $q(\mathbf{x}) = f(\mathbf{x},\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$. The associated matrix $\mathbf{B}_\mathbf{u}$ then satisfies the condition $\langle \mathbf{B}_\mathbf{u}\mathbf{x}, \mathbf{x}\rangle = \mathbf{x}^T\mathbf{B}_\mathbf{u}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$, and it is said that $\mathbf{B}_\mathbf{u}$ is positive definite, which is written as $\mathbf{B}_\mathbf{u} > 0$.

In the same way, it is said that a quadratic form $q$ is positive semi-definite if $\mathbf{x}^T\mathbf{B}_\mathbf{u}\mathbf{x} \geq 0$, negative definite if $\mathbf{x}^T\mathbf{B}_\mathbf{u}\mathbf{x} < 0$, for all $\mathbf{x} \neq \mathbf{0}$ and negative semi-definite if $\mathbf{x}^T\mathbf{B}_\mathbf{u}\mathbf{x} \leq 0$. It is written $\mathbf{B}_\mathbf{u} \geq 0$, $\mathbf{B}_\mathbf{u} < 0$, and $\mathbf{B}_\mathbf{u} \leq 0$, respectively.

In the case of a Hermitian form, the previous definitions remain valid by replacing $\mathbf{x}^T$ by $\mathbf{x}^H$.

In Table 4.17, we summarize some special cases of bilinear/sesquilinear forms and quadratic/Hermitian forms with the properties satisfied by the associated matrices.

| Types | Definitions | Properties |
|---|---|---|
| | **Bilinear form:** $f : \mathcal{U}^2 \to \mathbb{R}$ | |
| Symmetric | $f(\mathbf{y},\mathbf{x}) = f(\mathbf{x},\mathbf{y}) = \mathbf{y}_\mathbf{u}^T \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}$ | $\mathbf{B}_\mathbf{u}^T = \mathbf{B}_\mathbf{u}$ |
| Antisymmetric | $f(\mathbf{y},\mathbf{x}) = -f(\mathbf{x},\mathbf{y})$ | $\mathbf{B}_\mathbf{u}^T = -\mathbf{B}_\mathbf{u}$. |
| | **Sesquilinear form:** $f : \mathcal{U}^2 \to \mathbb{C}$ | |
| Hermitian symmetry | $f(\mathbf{y},\mathbf{x}) = f^*(\mathbf{x},\mathbf{y}) = \mathbf{y}_\mathbf{u}^H \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}$ | $\mathbf{B}_\mathbf{u}^H = \mathbf{B}_\mathbf{u}$ |
| Antihermitian symmetry | $f(\mathbf{y},\mathbf{x}) = -f^*(\mathbf{x},\mathbf{y})$ | $\mathbf{B}_\mathbf{u}^H = -\mathbf{B}_\mathbf{u}$ |
| | **Quadratic form** : $q : \mathcal{U} \to \mathbb{R}$ | |
| Symmetric | $q(\mathbf{x}) = f(\mathbf{x},\mathbf{x}) = \mathbf{x}_\mathbf{u}^T \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}$ | $\mathbf{B}_\mathbf{u}^T = \mathbf{B}_\mathbf{u}$ |
| Positive definite | $q(\mathbf{x}) = \mathbf{x}_\mathbf{u}^T \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u} > 0 \ , \ \forall \mathbf{x}_\mathbf{u} \in \mathcal{U} \ , \ \mathbf{x}_\mathbf{u} \neq \mathbf{0}$ | $\mathbf{B}_\mathbf{u} > 0$ |
| | **Hermitian form** : $q : \mathcal{U} \to \mathbb{R}$ | |
| Hermitian symmetry | $q(\mathbf{x}) = f(\mathbf{x},\mathbf{x}) = \mathbf{x}_\mathbf{u}^H \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}$ | $\mathbf{B}_\mathbf{u}^H = \mathbf{B}_\mathbf{u}$ |
| Positive definite | $q(\mathbf{x}) = \mathbf{x}_\mathbf{u}^H \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u} > 0 \ , \ \forall \mathbf{x}_\mathbf{u} \in \mathcal{U} \ , \ \mathbf{x}_\mathbf{u} \neq \mathbf{0}$ | $\mathbf{B}_\mathbf{u} > 0$ |

**Table 4.17.** *Special cases of bilinear/sesquilinear forms and quadratic/Hermitian forms*

### 4.15.4. *Examples of positive definite quadratic forms*

As we saw in section 3.4.1.1, by definition, an inner product in a real pre-Hilbertian space is a positive definite bilinear form. The square of the norm induced from the inner product is the associated quadratic form which is positive definite.

For instance:

– In $\mathbb{R}^I$:

$$f(\mathbf{x},\mathbf{y}) = \langle \mathbf{x},\mathbf{y}\rangle = \sum_{i=1}^{I} x_i y_i \;\Rightarrow\; q(\mathbf{x}) = \|\mathbf{x}\|_2^2 = \sum_{i=1}^{I} x_i^2.$$

– In $\mathcal{C}^0([a,b],\mathbb{R})$, with $[a,b]\subset\mathbb{R}$:

$$f(x,y) = \langle x,y\rangle = \int_a^b x(t)y(t)dt \;\Rightarrow\; q(x) = \|x\|_2^2 = \int_a^b x^2(t)dt.$$

– In $\mathbb{R}^{I\times I}$:

$$f(\mathbf{A},\mathbf{B}) = \mathrm{tr}(\mathbf{B}^T\mathbf{A}) = \sum_{i,j=1}^{I} a_{ij}b_{ij} \;\Rightarrow\; q(\mathbf{A}) = \|\mathbf{A}\|_F^2 = \mathrm{tr}(\mathbf{A}^T\mathbf{A}) = \sum_{i,j=1}^{I} a_{ij}^2.$$

### 4.15.5. *Cauchy–Schwarz and Minkowski inequalities*

We have the following inequalities satisfied by a symmetric bilinear form and its associated quadratic form:

PROPOSITION 4.76.– *Cauchy–Schwarz inequality: Given a* $\mathbb{R}$*-v.s.* $\mathcal{U}$ *and a positive definite quadratic form* $q$ *over* $\mathcal{U}$*, we have:*

$$\forall (x,y)\in\mathcal{U}^2\ ,\ [f(x,y)]^2 \leq q(x)+q(y),$$

*with the equality if and only if* $x$ *and* $y$ *are linearly dependent in the v.s.* $\mathcal{U}$*.*

In the case where the quadratic form $q$ is positive definite, it can be interpreted as a norm ($\|x\| = \sqrt{q(x)}$), and the associated polar form $f$ as an inner product. The above inequality is to be compared with the Cauchy–Schwarz inequality [3.36].

PROPOSITION 4.77.– *Minkowski inequality: If the quadratic form* $q$ *is positive definite, we have:*

$$\forall (x,y)\in\mathcal{U}^2\ ,\ \sqrt{q(x+y)} \leq \sqrt{q(x)}+\sqrt{q(y)},$$

*with the equality if and only if there exists* $\lambda\geq 0$ *such that* $y=\lambda x$*, or if* $x=0$*.*

This inequality is to be compared to the Minkovski inequality [3.38] for $p=2$.

To close this section on quadratic and Hermitian forms, we are going to present two important results which are the Gauss reduction method, and Sylvester's law of inertia, leading to the notion of signature of a quadratic/Hermitian form. These results are

directly linked to the diagonalization of a quadratic/sesquilinear form. Although the field $\mathbb{K}$ can be arbitrary, in the next two sections, we specialize to the real case ($\mathbb{K} = \mathbb{R}$). We first introduce some basic definitions relating to $f$-orthogonality which allows to define a $f$-orthogonal basis. The kernel, the rank, and the degeneration property of a symmetric bilinear form are defined. Then in section 4.15.7, we will present the Gauss reduction method and Sylvester's law of inertia.

### 4.15.6. *Orthogonality, rank, kernel and degeneration of a bilinear form*

Consider a $\mathbb{R}$-v.s. $\mathcal{U}$ of dimension $J$, with the basis $\{u\}$, a symmetric bilinear form $f : \mathcal{U}^2 \to \mathbb{R}$, and its associated quadratic form $q$.

– Two vectors $x, y \in \mathcal{U}$ are called orthogonal relatively to $f$ if $f(x, y) = 0$.

It is also said that the vectors are $f$-orthogonal, and it is written $x \perp_f y$.

– When $q(x) = f(x, x) = 0$, that is, when $x$ is orthogonal to itself, it is then said that $x$ is isotropic, and the set of isotropic vectors is called the isotropic cone of $f$, denoted by $C_f$ and such that $C_f = \{x \in \mathcal{U} : q(x) = 0\}$.

– If $x$ is $f$-orthogonal to a set of vectors $\{y_1, \cdots, y_p\}$, then it is $f$-orthogonal to any linear combination of these vectors.

– Two subspaces $\mathcal{V}, \mathcal{W} \subset \mathcal{U}$ are said to be $f$-orthogonal, and denoted $\mathcal{V} \perp_f \mathcal{W}$, if:

$$\forall x \in \mathcal{V} \text{ and } \forall y \in \mathcal{W}\,,\ f(x, y) = 0.$$

– Let $\mathcal{W} \subset \mathcal{U}$ be a subset of $\mathcal{U}$. The $f$-orthogonal complement of $\mathcal{W}$, denoted by $\mathcal{W}^{\perp_f}$, is defined as the set:

$$\mathcal{W}^{\perp_f} = \{y \in \mathcal{U} : \forall x \in \mathcal{W}, x \perp_f y\}.$$

This set $\mathcal{W}^{\perp_f}$ is a subspace of $\mathcal{U}$. Indeed, given $u, v \in \mathcal{W}^{\perp_f}$, and $\lambda \in \mathbb{R}$, then for any $x \in \mathcal{W}$, we have $f(x, u + \lambda v) = f(x, u) + \lambda f(x, v) = 0$, and therefore $u + \lambda v \in \mathcal{W}^{\perp_f}$, implying that $\mathcal{W}^{\perp_f}$ is a subspace of $\mathcal{U}$.

– A basis $\{u\} = \{u_1, \cdots, u_J\}$ of $\mathcal{U}$ is said to be $f$-orthogonal if its vectors are pair-wise $f$-orthogonal:

$$f(u_i, u_j) = 0\ \ ,\ \ \forall i, j \in \langle J \rangle, i \neq j.$$

We then have:

$$q(x) = q(\sum_{j=1}^{J} x_j u_j) = \sum_{j=1}^{J} x_j^2\, q(u_j). \qquad [4.80]$$

The matrix associated with $f$, in the basis $\{u\}$, is such that:

$$(\mathbf{B}_u)_{ij} = f(u_i, u_j) = q(u_j)\delta_{ij},$$

that is, the matrix $\mathbf{B}_u$ is diagonal, having as diagonal elements the coefficients $q(u_j), j \in \langle J \rangle$. Conversely, if $\mathbf{B}_u$ is diagonal, then $f(u_i, u_j) = 0$ for $i \neq j$, which implies the $f$-orthogonality of the basis.

It can be concluded that a basis of $\mathcal{U}$ is orthogonal for $q$ if and only if the matrix of $q$ in this basis is diagonal. It is then said that this basis of $\mathcal{U}$ is $q$-orthogonal. The quadratic form associated with a symmetric bilinear form, with an $f$-orthogonal basis, can then be written as:

$$q(x) = \sum_{j=1}^{J} \lambda_j x_j^2, \qquad [4.81]$$

where the coefficients $\lambda_j$ are the diagonal elements of the matrix associated with $f$, and the $x_j$s are the coordinates of $x$ in the $f$-orthogonal basis.

– The rank of $q$ is equal to the number of vectors of the basis such that $q(u_j) \neq 0$, that is, the number of non-isotropic vectors of the basis. It is also the rank of the associated matrix $\mathbf{B}_u$.

– The kernel of $f$ is the subspace of $\mathcal{U}$, denoted by $\text{Ker}(f)$ and defined as:

$$\text{Ker}(f) = \{x \in \mathcal{U} : \forall y \in \mathcal{U}, f(x, y) = 0\}.$$

– The bilinear form $f$ is said to be non-degenerate if $\text{Ker}(f) = \{0\}$, that is, $\{0\}$ is the unique vector of $\mathcal{U}$ which is $f$-orthogonal to all other vectors of $\mathcal{U}$. Equivalently, $f$ is non-degenerate if for every non-null $x \in \mathcal{U}$, there exists $y \in \mathcal{U}$ such that $f(x, y) \neq 0$. Otherwise, $f$ is said to be degenerate.

– According to the rank theorem applied to $f$, the rank of $f$, denoted by $\text{r}(f)$, is the integer defined by:

$$\text{r}(f) = \dim(\mathcal{U}) - \dim[\text{Ker}(f)],$$

and if $f$ is non-degenerate:

$$\text{r}(f) = \dim(\mathcal{U}) = \text{r}(\mathbf{B}_u).$$

Therefore, $f$ is non-degenerate if and only if its matrix $\mathbf{B}_u$, in any basis of $\mathcal{U}$, is regular.

### 4.15.7. *Gauss reduction method and Sylvester's inertia law*

In the following, we assume that $\{u_1, \cdots u_J\}$ is an $f$-orthogonal basis of the $\mathbb{R}$-v.s. $\mathcal{U}$ of dimension $J$.

PROPOSITION 4.78.– *(Gauss[14] reduction): Any quadratic form can be written as a linear combination of squares of independent linear forms, the number of squares being equal to the rank of the quadratic form.*

Indeed, considering the dual basis $\{u_1^*, \cdots u_J^*\}$ associated with the basis $\{u_1, \cdots u_J\}$, such that $u_i^*(x) = x_i$ (see section 2.6.3.4), we can rewrite [4.80] as:

$$\forall\, x \in \mathcal{U},\ \ q(x) = \sum_{j=1}^{J} q(u_j) x_j^2$$
$$= \sum_{j=1}^{J} \lambda_j u_j^*(x)^2\,,\ \ \lambda_j = q(u_j) \in \mathbb{R}. \qquad [4.82]$$

Since vectors $\{u_1^*, \cdots u_J^*\}$ form a basis of the dual space $\mathcal{U}^*$, the linear forms $u_j^*(x)$ are linearly independent in $\mathcal{U}^*$. The rank of $q$ is given by the number of non-zero coefficients $\lambda_j$. The Gauss method proceeds iteratively on $j$ to diagonalize the matrix of $q$.

PROPOSITION 4.79.– *(Sylvester's law of inertia[15]): For any $f$-orthogonal basis of a $\mathbb{R}$-v.s. $\mathcal{U}$ of dimension $J$, the quadratic form associated with $f$ can be written as a linear combination of squares of independent linear forms such as [4.82], with a sum of $m$ squares (for the coefficients $\lambda_j > 0$), a difference of $n$ squares (for the coefficients $\lambda_j < 0$), and $J - m - n$ zero terms corresponding to zero coefficients $\lambda_j$. The rank of $q$ is equal to $m + n$. Then, there exists a basis of $\mathcal{U}$ in which the matrix of*

14 Johann Carl Friedrich Gauss (1777–1855), German mathematician, astronomer and physicist, who was one of the most famous scientists of the 19th century, not only for the large range of his discoveries but also for his very rigorous spirit. His works covered all the branches of mathematics. In 1799, he proved the fundamental theorem of algebra stated by d'Alembert (1717–1783), also called the d'Alembert–Gauss theorem. He developed the method of least-squares (LS) that he applied for calculating orbits of planets, in particular of Ceres. In homage, an asteroid was named Gaussia. In 1801, he published the book *Disquisitiones Arithmeticae*. The LS method was independently discovered by Legendre, in 1806, before its publication by Gauss, in 1809. In probability theory, he employed the Normal distribution, also called the Gaussian or Laplace–Gauss distribution, first introduced by de Moivre (1667–1754), for analyzing astronomical data. This is the most used distribution for modeling physical, biological or financial phenomena, among many others, because of the central limit theorem.

15 James Joseph Sylvester (1814–1897), English mathematician who made important contributions to matrix theory, in particular the theory of canonical forms, and to invariant theory, through a long collaboration with Arthur Cayley (1821–1895). In 1889, he rediscovered the singular value decomposition (SVD) which was discovered independently by Eugenio Beltrami (1835–1899) in 1873 and Camille Jordan (1838–1922) in 1874.

*the quadratic form $q$ is block-diagonal of the form:*

$$\begin{bmatrix} \mathbf{I}_m & & \mathbf{0} \\ & -\mathbf{I}_n & \\ \mathbf{0} & & \mathbf{0}_{k\times k} \end{bmatrix}, \qquad [4.83]$$

*where $k = J - m - n$. This matrix is called the canonical form of $q$ over $\mathbb{R}$.*

The pair $(m, n)$ is called the signature of the (real) quadratic form $q$. This signature can be determined by applying the Gauss method, or by calculating the eigendecomposition of $\mathbf{B_u}$. The numbers $m$ and $n$ are then equal to the numbers of positive and negative eigenvalues, respectively.

NOTE 4.80.– The signature of the (real) quadratic form $q$ is independent of the basis chosen to represent $q$.

PROPOSITION 4.81.– *From the signature $(m, n)$, the quadratic form can be classified as follows:*

*1) $q$ is of rank $m + n$, and thereby nondegenerate if $m + n = J$;*

*2) $q$ is degenerate if $m + n < J$;*

*3) $q$ is positive definite if $(m, n) = (J, 0)$; the canonical form (4.83) is then the identity matrix of order $J$;*

*4) $q$ is negative definite if $(m, n) = (0, J)$.*

EXAMPLE 4.82.– To illustrate the previous results, consider the quadratic form $q : \mathbb{R}^2 \ni (x_1, x_2) \mapsto x_1x_2 \in \mathbb{R}$. It is obvious that this quadratic form can be rewritten as:

$$(x_1, x_2) \mapsto \left(\frac{x_1 + x_2}{2}\right)^2 - \left(\frac{x_1 - x_2}{2}\right)^2, \qquad [4.84]$$

namely, a linear combination, in the form of a difference of squares of the two following linear forms:

$$(x_1, x_2) \mapsto \frac{x_1 + x_2}{2} \text{ and } (x_1, x_2) \mapsto \frac{x_1 - x_2}{2}.$$

This result can be recovered from a diagonalization of the matrix of the original quadratic form which can be written as:

$$(x_1, x_2) \mapsto [x_1x_2] \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{x}^T\mathbf{B}\mathbf{x}. \qquad [4.85]$$

This diagonalization is the result of an eigendecomposition[16] $\mathbf{B} = \mathbf{PDP}^{-1}$, where the diagonal matrix $\mathbf{D}$ has eigenvalues of $\mathbf{B}$ as diagonal elements, and $\mathbf{P}$ has for columns the eigenvectors (see section 4.16.2).

The characteristic equation being $\det(\lambda\mathbf{I} - \mathbf{B}) = \lambda^2 - 1/4 = 0$, the eigenvalues are equal to $\pm 1/2$, from which $\mathbf{D} = \begin{bmatrix} 1/2 & 0 \\ 0 & -1/2 \end{bmatrix}$, and $\mathbf{P} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

After a change of basis, the quadratic form is written as $(x_1, x_2) \mapsto \underline{\mathbf{x}}^T \underline{\mathbf{B}} \underline{\mathbf{x}}$, with:

$$\underline{\mathbf{x}} = \mathbf{P}^{-1}\mathbf{x} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (x_1 + x_2)/\sqrt{2} \\ (x_1 - x_2)/\sqrt{2} \end{bmatrix}$$

$$\underline{\mathbf{B}} = \mathbf{P}^T\mathbf{BP} = \mathbf{D} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

from which it can be deduced that $\underline{\mathbf{x}}^T \underline{\mathbf{B}} \underline{\mathbf{x}} = \left(\frac{x_1+x_2}{2}\right)^2 - \left(\frac{x_1-x_2}{2}\right)^2$.

## 4.16. Eigenvalues and eigenvectors

### 4.16.1. *Characteristic polynomial and Cayley–Hamilton theorem*

#### 4.16.1.1. *Matrix polynomials*

Given the square matrix $\mathbf{A}$ of order $n$, and a polynomial $p(\lambda)$ over the field $\mathbb{K}$ such that $p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$, we define the polynomial $p(\mathbf{A}) = \mathbf{A}^n + a_{n-1}\mathbf{A}^{n-1} + \cdots + a_1\mathbf{A} + a_0\mathbf{I}_n$. It is said that $p(\mathbf{A})$ is a polynomial in the matrix $\mathbf{A}$, and that $\mathbf{A}$ is a root or a zero of polynomial $p(\lambda)$ if $p(\mathbf{A}) = 0$.

A polynomial $p$ such that $p(\mathbf{A}) = \mathbf{0}$ is called an annihilating polynomial of $\mathbf{A}$, and the annihilating polynomial of lowest degree is called the minimal polynomial of $\mathbf{A}$, denoted by $p_{\min}$. Every annihilating polynomial is a multiple of $p_{\min}$. We shall see that the Cayley–Hamilton theorem provides an annihilating polynomial, called the characteristic polynomial.

#### 4.16.1.2. *Characteristic polynomial and characteristic equation*

The characteristic polynomial of the square matrix $\mathbf{A}$, of order $n$, is the polynomial defined as:

$$p_{\mathbf{A}}(\lambda) = \det(\lambda\mathbf{I}_n - \mathbf{A}) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0.$$

16 The eigendecomposition of a matrix will be presented in Volume 2.

The equation $p_{\mathbf{A}}(\lambda) = 0$ is called the characteristic equation of $\mathbf{A}$, and its roots are the eigenvalues of $\mathbf{A}$. These roots may be all distinct, or some of them may be repeated, as discussed in section 4.16.2.

FACT 4.83.– The following relations can be shown:

$$\det(\mathbf{A}) = (-1)^n a_0, \quad \operatorname{tr}(\mathbf{A}) = -a_{n-1}.$$

#### 4.16.1.3. *Cayley–Hamilton theorem*

The Cayley–Hamilton theorem establishes that any matrix $\mathbf{A}$ is a zero of its characteristic polynomial, in other words the matrix $p(\mathbf{A}) = \mathbf{A}^n + a_{n-1}\mathbf{A}^{n-1} + \cdots + a_1\mathbf{A} + a_0\mathbf{I}_n$ is zero. The characteristic polynomial is therefore an annihilating polynomial for $\mathbf{A}$.

It should be noted that this theorem must not be stated using the trivial formula $\det(\mathbf{A}\mathbf{I}_n - \mathbf{A}) = 0$, but as $\det(\mathbf{M}) = \mathbf{0}$, with:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} - a_{11}\mathbf{I}_n & -a_{12}\mathbf{I}_n & \dots & -a_{1n}\mathbf{I}_n \\ -a_{21}\mathbf{I}_n & \mathbf{A} - a_{22}\mathbf{I}_n & \dots & -a_{2n}\mathbf{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1}\mathbf{I}_n & -a_{n2}\mathbf{I}_n & \dots & \mathbf{A} - a_{nn}\mathbf{I}_n \end{bmatrix} = \mathbf{I}_n \otimes \mathbf{A} - \mathbf{A} \otimes \mathbf{I}_n,$$

where the symbol $\otimes$ denotes the Kronecker product. The elements of $\mathbf{M}$ are matrices of $\mathbb{K}^{n\times n}$, and the determinant of $\mathbf{M}$ is itself a matrix of $\mathbb{K}^{n\times n}$ and not a scalar.

From the Cayley–Hamilton theorem, it is easy to show the following proposition.

PROPOSITION 4.84.– *For all $p \geq n$, $\mathbf{A}^p$ can be expressed as a linear combination of $\mathbf{I}_n, \mathbf{A}, \cdots, \mathbf{A}^{n-1}$.*

EXAMPLE 4.85.– For

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

we have $p(\lambda) = \det(\lambda\mathbf{I}_2 - \mathbf{A}) = \lambda^2 + a_1\lambda + a_0$, with $a_1 = -\operatorname{tr}(\mathbf{A}) = -(a_{11} + a_{22})$ and $a_0 = \det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$. By defining the matrix:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} - a_{11}\mathbf{I}_2 & -a_{12}\mathbf{I}_2 \\ -a_{21}\mathbf{I}_2 & \mathbf{A} - a_{22}\mathbf{I}_2 \end{bmatrix},$$

the Cayley–Hamilton theorem can be written as:

$$\begin{aligned} \det(\mathbf{M}) &= \mathbf{A}^2 - (a_{11} + a_{22})\,\mathbf{A} + (a_{11}a_{22} - a_{12}a_{21})\,\mathbf{I}_2 \\ &= \mathbf{A}^2 - \operatorname{tr}(\mathbf{A})\mathbf{A} + \det(\mathbf{A})\,\mathbf{I}_2 \\ &= p(\mathbf{A}) = \mathbf{0}, \end{aligned}$$

from which the following is deduced:

$$\mathbf{A}^2 = \mathrm{tr}(\mathbf{A})\mathbf{A} - \det(\mathbf{A})\mathbf{I}_2,$$

$$\mathbf{A}^3 = \mathbf{A}^2\mathbf{A} = \mathrm{tr}(\mathbf{A})\mathbf{A}^2 - \det(\mathbf{A})\mathbf{A} = [\mathrm{tr}^2(\mathbf{A}) - \det(\mathbf{A})]\mathbf{A} - \mathrm{tr}(\mathbf{A})\det(\mathbf{A})\mathbf{I}_2,$$

and so on.

### 4.16.2. *Right eigenvectors*

Given a square matrix $\mathbf{A} \in \mathbb{K}^{I\times I}$, a right eigenvector (or simply eigenvector) is a non-zero vector $\mathbf{v}_k \in \mathbb{K}^I$ satisfying:

$$\mathbf{A}\mathbf{v}_k = \lambda_k \mathbf{v}_k, \qquad [4.86]$$

where the scalar $\lambda_k$ is called the eigenvalue of $\mathbf{A}$ associated with the eigenvector $\mathbf{v}_k$.

NOTE 4.86.– It is important to point out that an eigenvector may not be the zero vector.

The definition [4.86] can also be written as:

$$(\mathbf{A} - \lambda_k \mathbf{I}_I)\mathbf{v}_k = 0, \quad \text{with} \quad \mathbf{v}_k \neq \mathbf{0},$$

which implies that the matrix $\mathbf{A} - \lambda_k\mathbf{I}_I$ is singular, and therefore its determinant is zero (see section 4.11.3). Consequently, we can conclude that the eigenvector $\mathbf{v}_k$ is in the nullspace of $\mathbf{A}-\lambda_k\mathbf{I}_I$ , and the eigenvalues of $\mathbf{A}$ are the $I$ roots of the characteristic polynomial, or more specifically, the solutions of the characteristic equation:

$$p_{\mathbf{A}}(\lambda) = \det(\lambda\mathbf{I}_I - \mathbf{A}) = 0. \qquad [4.87]$$

They will be denoted by $\lambda_i(\mathbf{A})$, or just $\lambda_i, i \in \langle I\rangle$. If $\lambda_i$ is a zero of order $n_i$ of the characteristic equation (i.e. if $n_i$ roots of the characteristic polynomial are equal to $\lambda_i$), it is said that $\lambda_i$ is a multiple eigenvalue, with an (algebraic) multiplicity $n_i$. Otherwise, it is said that $\lambda_i$ is a simple eigenvalue.

One defines the geometric multiplicity $m_i$ of $\lambda_i$ as the dimension of the eigensubspace associated with $\lambda_i$, namely, the maximal number of linearly independent eigenvectors associated with $\lambda_i$.

It can be shown that the geometric multiplicity of $\lambda_i$ corresponds to the dimension of the kernel of $\mathbf{A} - \lambda_i\mathbf{I}$. Subsequently, using the rank theorem (see Table 4.7), it can be deduced that $m_i = \dim(\mathbf{A} - \lambda_i\mathbf{I}) = I - \mathrm{r}(\mathbf{A} - \lambda_i\mathbf{I})$.

Hereafter, we summarize a few properties of eigenvalues.

– If $\mathrm{r}(\mathbf{A}) = k$, then $\lambda = 0$ is an eigenvalue with a geometric multiplicity $I - \mathrm{r}(\mathbf{A}) = I - k$.

– If $\mathbf{A}$ has $k$ (distinct) non-zero eigenvalues, then $\mathrm{r}(\mathbf{A}) \geq k$. Therefore, $\lambda = 0$ is an eigenvalue with an algebraic multiplicity $I - k$, and a geometric multiplicity $I - \mathrm{r}(\mathbf{A}) \leq I - k$.

– The algebraic multiplicity of an eigenvalue is therefore at least equal to its geometric multiplicity, that is, $n_i \geq m_i$.

It can be demonstrated that a matrix is diagonalizable if and only if $n_i = m_i$ for all $\lambda_i$, which is the case, in particular, when all eigenvalues are simple.

– In the case where $\mathbf{A} \in \mathbb{K}^{I \times I}$ has $J \leq I$ distinct eigenvalues $\lambda_1, \cdots, \lambda_J$, of respective algebraic multiplicities $n_1, \cdots, n_J$, with $\sum_{j=1}^{J} n_j = I$, the characteristic polynomial can be written as $p(\lambda) = \prod_{j=1}^{J} (\lambda - \lambda_j)^{n_j}$.

EXAMPLE 4.87.– Let the matrix $\mathbf{A}$ be defined as:

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ 0 & 0 & d \\ 0 & 0 & 0 \end{bmatrix}.$$

with $a \neq 0$ and $d \neq 0$. We have $\mathrm{r}(\mathbf{A}) = 2$ and $\det(\lambda \mathbf{I} - \mathbf{A}) = \lambda^2(\lambda - a)$, and therefore $\mathbf{A}$ has a non-zero eigenvalue $\lambda = a$ and two zero eigenvalues. The geometric multiplicity of $\lambda = 0$ is equal to $I - \mathrm{r}(\mathbf{A}) = 1$, whereas its algebraic multiplicity is 2.

### 4.16.3. *Spectrum and regularity/singularity conditions*

The spectrum of $\mathbf{A}$ is the set of eigenvalues of $\mathbf{A}$. It is denoted as $\mathrm{sp}(\mathbf{A}) = \{\lambda_i(\mathbf{A}), i \in \langle I \rangle\}$. The spectral radius of $\mathbf{A}$ is the maximal value of the moduli of the eigenvalues:

$$\rho(\mathbf{A}) = \max_{i \in \langle I \rangle} |\lambda_i(\mathbf{A})|. \qquad [4.88]$$

FACT 4.88.– The spectral radius is such that $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$ for any matrix norm.

PROPOSITION 4.89.– *The matrix $\mathbf{A}$ is singular if and only if $0 \in \mathrm{sp}(\mathbf{A})$, that is, $\lambda = 0$ is an eigenvalue of $\mathbf{A}$.*

PROOF.– $\mathbf{A}$ is singular if and only if $\mathbf{A}\mathbf{x} = \mathbf{0}_I$ for $\mathbf{x} \neq \mathbf{0}_I$, or equivalently if and only if $\mathbf{A}\mathbf{x} = 0\,\mathbf{x}$ for $\mathbf{x} \neq \mathbf{0}_I$, that is, if and only if $\lambda = 0$ is an eigenvalue. □

COROLLARY 4.90.– *From Proposition 4.89, it can be concluded that $\mathbf{A}$ is regular if and only if $0 \notin \mathrm{sp}(\mathbf{A})$.*

### 4.16.4. *Left eigenvectors*

A left eigenvector of $\mathbf{A}$ associated with the eigenvalue $\mu_i$, is a vector $\mathbf{u}_i \neq \mathbf{0}_I$ satisfying:

$$\mathbf{u}_i^H \mathbf{A} = \mu_i \mathbf{u}_i^H \qquad [4.89]$$

or equivalently:

$$\mathbf{u}_i^H (\mu_i \mathbf{I}_I - \mathbf{A}) = 0,$$

which implies $\det(\mu_i \mathbf{I}_I - \mathbf{A}) = 0$, therefore the characteristic equation [4.87], demonstrating that left and right eigenvectors are associated with the same eigenvalues. However, it should be noted that right and left eigenvectors associated with the same eigenvalue are different in general.

### 4.16.5. *Properties of eigenvectors*

The eigenvectors of a matrix $\mathbf{A}$ satisfy the following properties.

PROPOSITION 4.91.– *If $\mathbf{v}_k$ is an eigenvector, then any multiple $\alpha \mathbf{v}_k$ with $\alpha \neq 0$ is also an eigenvector. The set of all eigenvectors $\{\mathbf{v}_k\}$ associated with $\lambda_k$ is a subspace of $\mathbb{K}^I$, of dimension at least equal to 1, called eigensubspace of $\mathbf{A}$ associated with $\lambda_k$.*

FACT 4.92.– In order to make eigenvectors unique, they can be normalized, namely, by choosing $\|\mathbf{v}_k\|_2^2 = \mathbf{v}_k^H \mathbf{v}_k = 1$.

PROPOSITION 4.93.– *The $p$ eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_p$ associated with $p$ distinct eigenvalues $\lambda_1, \cdots, \lambda_p$ are linearly independent.*

PROPOSITION 4.94.– *Given a right eigenvector $\mathbf{v}_k$ and a left eigenvector $\mathbf{u}_i$ associated with two distinct eigenvalues $\lambda_k$ and $\lambda_i$, respectively, we have the following relation of orthogonality according to the Hermitian inner product:*

$$\langle \mathbf{v}_k, \mathbf{u}_i \rangle = \mathbf{u}_i^H \mathbf{v}_k = 0. \qquad [4.90]$$

*This relation means that any left eigenvector is orthogonal to any right eigenvector when the two vectors are associated with distinct eigenvalues.*

PROOF.– Pre-multiplying the two members of the definition equation $\mathbf{A}\mathbf{v}_k = \lambda_k \mathbf{v}_k$ by $\mathbf{u}_i^H$ gives:

$$\mathbf{u}_i^H (\mathbf{A}\mathbf{v}_k) = \lambda_k \mathbf{u}_i^H \mathbf{v}_k.$$

Similarly, post-multiplying the two members of relation [4.89] by $\mathbf{v}_k$ gives:

$$(\mathbf{u}_i^H \mathbf{A})\mathbf{v}_k = \lambda_i \mathbf{u}_i^H \mathbf{v}_k.$$

Subtracting these last two equations memberwise gives $(\lambda_i - \lambda_k)\mathbf{u}_i^H \mathbf{v}_k = 0$, and taking into account the assumption that $\lambda_i \neq \lambda_k$ allows to deduce the orthogonality relation [4.90]. □

From the previous proposition, the following one is easy to derive.

PROPOSITION 4.95.– *Let $\mathbf{A} \in \mathbb{K}^{I \times I}$ be such that all its eigenvalues are distinct. Let us define matrices $\mathbf{P}, \mathbf{Q} \in \mathbb{K}^{I \times I}$ formed by the (normalized) right and left eigenvectors of $\mathbf{A}$, respectively:*

$$\mathbf{P} = \left[\mathbf{v}_1 \cdots \mathbf{v}_I\right], \quad \mathbf{Q} = \left[\mathbf{u}_1 \cdots \mathbf{u}_I\right]. \qquad [4.91]$$

*Then, from the orthogonality property [4.90], it can be inferred that $\mathbf{P}$ and $\mathbf{Q}$ satisfy the following orthogonality relation:*

$$\mathbf{Q}^H \mathbf{P} = \mathbf{I}_I, \qquad [4.92]$$

*that is, $\mathbf{P}$ and $\mathbf{Q}$ are bi-unitary, and their inverses are such that $\mathbf{P}^{-1} = \mathbf{Q}^H$ and $\mathbf{Q}^{-1} = \mathbf{P}^H$.*

For the demonstration of the next result, we use the lemma below, which will be demonstrated in Proposition 5.22.

LEMMA 4.96.– *For $\mathbf{A} \in \mathbb{K}^{I \times J}$ and $\mathbf{B} \in \mathbb{K}^{J \times I}$, we have the following property:*

$$\lambda^J \det(\lambda \mathbf{I}_I - \mathbf{AB}) = \lambda^I \det(\lambda \mathbf{I}_J - \mathbf{BA}). \qquad [4.93]$$

PROPOSITION 4.97.– *Given $\mathbf{A} \in \mathbb{K}^{I \times J}$ and $\mathbf{B} \in \mathbb{K}^{J \times I}$, with $J \geq I$, the spectrum of $\mathbf{BA}$ consists of eigenvalues of $\mathbf{AB}$ and $J - I$ zeros.*

PROOF.– According to [4.93], we have $\det(\lambda \mathbf{I}_J - \mathbf{BA}) = \lambda^{J-I} \det(\lambda \mathbf{I}_I - \mathbf{AB})$, which implies:

$$\det(\lambda \mathbf{I}_J - \mathbf{BA}) = 0 \quad \Leftrightarrow \quad \begin{cases} \det(\lambda \mathbf{I}_I - \mathbf{AB}) = 0, \\ \lambda = 0 \text{ (with multiplicity } J - I), \end{cases}$$

that is, the non-zero eigenvalues of $\mathbf{AB}$ and $\mathbf{BA}$ are identical, and $\mathbf{BA}$ has in addition $J - I$ zero eigenvalues. □

### 4.16.6. *Eigenvalues and eigenvectors of a regularized matrix*

In many applications, we have to invert an ill-conditioned matrix $\mathbf{A}$, namely, such that the ratio $|\lambda_{\max}(\mathbf{A})|/|\lambda_{\min}(\mathbf{A})|$ is large, where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$, respectively, denote the largest and smallest eigenvalue (in modulus). The result of the numerical computation of $\mathbf{A}^{-1}$ is then very inaccurate. That can occur when some eigenvalues are close to zero, that is, when the matrix is almost singular. To improve this computation, a so-called regularization method consists in adding a constant $\alpha \neq 0$ to each diagonal term of $\mathbf{A}$. As a result, a so-called regularized matrix is obtained of the form $\mathbf{A} + \alpha \mathbf{I}_I$, which satisfies the following property.

PROPOSITION 4.98.– *Given the matrix* $\mathbf{A}$ *having eigenpairs* $(\lambda_i, \mathbf{v}_i), i \in \langle I \rangle$*, then the matrix* $\mathbf{A} + \alpha \mathbf{I}_I$ *has the eigenpairs* $(\lambda_i + \alpha, \mathbf{v}_i), i \in \langle I \rangle$*, which means that the regularization operation has the effect of leaving eigenvectors unchanged, whereas eigenvalues are modified by adding the constant* $\alpha$*.*

PROOF.– Taking into account the definition of the pair $(\lambda_i, \mathbf{v}_i)$, we have $(\mathbf{A} + \alpha \mathbf{I}_I)\mathbf{v}_i = \mathbf{A}\mathbf{v}_i + \alpha \mathbf{v}_i = (\lambda_i + \alpha)\mathbf{v}_i$, which allows us to conclude that $\mathbf{v}_i$ is also an eigenvector of $\mathbf{A} + \alpha \mathbf{I}_I$, associated with the eigenvalue $\lambda_i + \alpha$. Therefore, $\mathbf{A}$ and $\mathbf{A} + \alpha \mathbf{I}_I$ have the same eigenvectors. □

### 4.16.7. *Other properties of eigenvalues*

For all $\mathbf{A} \in \mathbb{K}^{I \times I}$, we have the following relations:

$$\begin{aligned}
\lambda_i(\mathbf{A}^T) &= \lambda_i(\mathbf{A}), \quad \lambda_i(\mathbf{A}^*) = \lambda_i^*(\mathbf{A}) \quad \Rightarrow \quad \lambda_i(\mathbf{A}^H) = \lambda_i^*(\mathbf{A}), \\
\lambda_i(-\mathbf{A}) &= -\lambda_i(\mathbf{A}), \quad \lambda_i(k\mathbf{A}) = k\lambda_i(\mathbf{A}), \quad k \in \mathbb{K}, \\
\lambda_i(\mathbf{A}^{-1}) &= [\lambda_i(\mathbf{A})]^{-1}, \\
\lambda_i(\mathbf{A}^k) &= [\lambda_i(\mathbf{A})]^k, \; k \geq 2 \\
\mathrm{tr}(\mathbf{A}) &= \textstyle\sum_{i=1}^{I} \lambda_i(\mathbf{A}), \\
\mathrm{tr}(\mathbf{A}^k) &= \textstyle\sum_{i=1}^{I} \lambda_i^k(\mathbf{A}), \quad \text{with } k \geq 2, \\
\det(\mathbf{A}) &= \prod_{i=1}^{I} \lambda_i(\mathbf{A}).
\end{aligned}$$

Thus, the trace of a matrix is equal to the sum of its eigenvalues, whereas its determinant equals their product.

FACT 4.99.– From this last property, one can deduce that a matrix is invertible if and only if all its eigenvalues are different from zero.

Invertibility of a matrix thus depends on its eigenvalues, whereas its diagonalizability depends on the eigenvectors.

PROPOSITION 4.100.– *We have the following additional properties:*

*– If* $\mathbf{A}$ *is a lower or upper triangular matrix, its eigenvalues are equal to its diagonal elements.*

*– Two similar matrices have the same characteristic polynomial, and therefore the same eigenvalues with the same multiplicities. It is said that similarity preserves eigenvalues.*

PROOF.– According to [4.46], $\mathbf{A}$ and $\mathbf{B}$ are similar if they are linked by the relation $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. As a result, we have:

$$\begin{aligned}\det(\lambda\mathbf{I} - \mathbf{B}) &= \det(\lambda\mathbf{I} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P}) \\ &= \det[\mathbf{P}^{-1}(\lambda\mathbf{I} - \mathbf{A})\mathbf{P}] = \det[\mathbf{P}\mathbf{P}^{-1}(\lambda\mathbf{I} - \mathbf{A})] = \det(\lambda\mathbf{I} - \mathbf{A}),\end{aligned}$$

which allows us to conclude that $\mathbf{A}$ and $\mathbf{B}$ have the same characteristic polynomial, and therefore the same eigenvalues. □

*– A matrix* $\mathbf{A} \in \mathbb{K}^{I\times I}$ *is diagonalizable if there exists a regular matrix* $\mathbf{P} \in \mathbb{K}^{I\times I}$ *such that* $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ *is a diagonal matrix.*

Matrix diagonalization which plays a very important role in matrix calculus, will be studied in Volume 2. We provide here a result that is easy to use.

PROPOSITION 4.101.– *Any square matrix of order* $I$ *that has* $I$ *distinct eigenvalues is diagonalizable.*

### 4.16.8. *Symmetric/Hermitian matrices*

PROPOSITION 4.102.– *For a Hermitian matrix* $\mathbf{A} \in \mathbb{C}^{I\times I}$*, the eigenvalues are real-valued, and any right eigenvector of* $\mathbf{A}$ *is also a left eigenvector, associated with the same eigenvalue.*

PROOF.– Consider a Hermitian matrix $\mathbf{A}$, namely such that $\mathbf{A}^H = \mathbf{A}$. By transconjugating the two members of relation [4.89] associated with the eigenvalue $\mu_i$, we get:

$$\mathbf{A}\mathbf{u}_i = \mu_i^*\mathbf{u}_i. \qquad [4.94]$$

Subsequently, by pre-multiplying [4.94] by $\mathbf{u}_i^H$ and post-multiplying [4.89] by $\mathbf{u}_i$, it can be deduced that:

$$\mu_i^* = \frac{\mathbf{u}_i^H\mathbf{A}\mathbf{u}_i}{\mathbf{u}_i^H\mathbf{u}_i} = \mu_i, \qquad [4.95]$$

which allows us to conclude that the eigenvalues of a Hermitian matrix are real-valued. In addition, from relations [4.86], [4.94], and [4.95], we deduce that:

$$\mathbf{u}_i = \mathbf{v}_i, \quad \forall i \in \langle I \rangle,$$

that is, for a Hermitian matrix, any right eigenvector is also a left eigenvector, associated with the same eigenvalue. □

COROLLARY 4.103.– *The orthogonality relation* [4.90] *then implies that two eigenvectors of a Hermitian matrix associated with two distinct eigenvalues are orthogonal according to the Hermitian inner product. In the case where the $I$ eigenvalues are distinct, it is thus possible to construct an orthonormal basis of eigenvectors such that $\mathbf{u}_i^H \mathbf{u}_j = \delta_{ij}$.*

PROPOSITION 4.104.– *The eigenvalues of an anti-hermitian matrix are either zero, or pure imaginary numbers, and left and right eigenvectors associated with a same eigenvalue are identical.*

PROOF.– By post-multiplying the two members of [4.89] by $\mathbf{u}_i$, we have:

$$\mathbf{u}_i^H \mathbf{A} \mathbf{u}_i = \mu_i \mathbf{u}_i^H \mathbf{u}_i \quad \Rightarrow \quad \mu_i = \frac{\mathbf{u}_i^H \mathbf{A} \mathbf{u}_i}{\mathbf{u}_i^H \mathbf{u}_i}.$$

By taking the conjugate transpose of this scalar ratio and taking into account the assumption that $\mathbf{A}$ is antihermitian ($\mathbf{A}^H = -\mathbf{A}$), one obtains:

$$\mu_i^* = -\frac{\mathbf{u}_i^H \mathbf{A} \mathbf{u}_i}{\mathbf{u}_i^H \mathbf{u}_i} = -\mu_i, \qquad [4.96]$$

which allows us to conclude that the eigenvalues are either zero, or pure imaginary numbers.

In addition, by transconjugating the two members of [4.86], and using the properties $\mathbf{A}^H = -\mathbf{A}$ and [4.96], we get:

$$(\mathbf{A}\mathbf{v}_i)^H = -\mathbf{v}_i^H \mathbf{A} = \lambda_i^* \mathbf{v}_i^H = -\lambda_i \mathbf{v}_i^H \quad \Rightarrow \quad \mathbf{v}_i^H \mathbf{A} = \lambda_i \mathbf{v}_i^H.$$

Comparing this equation with the definition equation [4.89], it can be concluded that right and left eigenvectors, associated with a same eigenvalue, are identical. □

Similarly, it can be shown that the eigenvalues of a real symmetric matrix ($\mathbf{A}^T = \mathbf{A}$) are real-valued, which implies that the eigenvectors are real and orthogonal, that is, such that $\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$. As for a Hermitian matrix, in the case of a real symmetric matrix, it is thus possible to construct an orthonormal basis of eigenvectors.

Moreover, the eigenvalues of a symmetric positive definite (positive semi-definite) matrix are positive (positive or zero), and those of a real antisymmetric matrix ($\mathbf{A}^T = -\mathbf{A}$) are either zero, or pure imaginary numbers.

These results are summarized in Table 4.18. This table also includes unitary and orthogonal matrices, which are discussed in the following section.

| Classes of matrices | Properties of eigenvalues |
|---|---|
| $\mathbf{A} \in \mathbb{C}^{I\times I}$ Hermitian | $\lambda_i \in \mathbb{R}$ |
| $\mathbf{A} \in \mathbb{C}^{I\times I}$ antihermitian | $\lambda_i^* = -\lambda_i$ |
| $\mathbf{A} \in \mathbb{R}^{I\times I}$ symmetric | $\lambda_i \in \mathbb{R}$ |
| $\mathbf{A} \in \mathbb{R}^{I\times I}$ antisymmetric | $\lambda_i^* = -\lambda_i$ |
| $\mathbf{A} \in \mathbb{R}^{I\times I}$ positive definite | $\lambda_i > 0$ |
| $\mathbf{A} \in \mathbb{R}^{I\times I}$ positive semi-definite | $\lambda_i \geq 0$ |
| $\mathbf{A} \in \mathbb{R}^{I\times I}$ orthogonal | $\lvert\lambda_i\rvert = 1$ ( $\lambda_i = \pm 1$ if $\lambda_i \in \mathbb{R}$) |
| $\mathbf{A} \in \mathbb{C}^{I\times I}$ unitary | $\lvert\lambda_i\rvert = 1$ |

**Table 4.18.** *Properties of eigenvalues of some special matrices*

### 4.16.9. *Orthogonal/unitary matrices*

PROPOSITION 4.105.– *The eigenvalues of a real orthogonal matrix are of modulus* 1.

PROOF.– Let $\mathbf{A} \in \mathbb{R}^{I\times I}$ be an orthogonal matrix and $\mathbf{v}$ an eigenvector of $\mathbf{A}$ associated with the eigenvalue $\lambda$. Since the columns of $\mathbf{A}$ are linearly independent, we have $\det(\mathbf{A}) \neq 0$, and thus $\lambda \neq 0$. If $\lambda \in \mathbb{R}$, then $\mathbf{v}$ is also real. By pre-multiplying the two members of [4.86] by their transpose, and taking into account the orthogonality of $\mathbf{A}$, that is, $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, we get:

$$\mathbf{v}^T\mathbf{A}^T\mathbf{A}\mathbf{v} = \|\mathbf{v}\|^2 = \lambda^2\|\mathbf{v}\|^2,$$

which leads to $\lambda = \pm 1$. On the other hand, if $\lambda \in \mathbb{C}$, by multiplying on the left-hand side the two members of [4.86] by their conjugate transpose, with $\mathbf{A}^H = \mathbf{A}^T$ since $\mathbf{A}$ is real, we obtain:

$$\mathbf{v}^H\mathbf{A}^H\mathbf{A}\mathbf{v} = \mathbf{v}^H\mathbf{A}^T\mathbf{A}\mathbf{v} = \|\mathbf{v}\|^2 = |\lambda|^2\|\mathbf{v}\|^2, \qquad [4.97]$$

which means that $|\lambda| = 1$. □

Using relation [4.97], the following result can also be shown.

PROPOSITION 4.106.– *The eigenvalues of a unitary matrix are all of unit modulus.*

COROLLARY 4.107.– *By combining the last two propositions with the properties listed in the previous section, we deduce the following results:*

– *The eigenvalues of a real symmetric and orthogonal matrix are equal to* $1$ *or* $-1$.

– *The eigenvalues of a Hermitian and unitary matrix are equal to* $1$ *or* $-1$.

### 4.16.10. *Eigenvalues and extrema of the Rayleigh quotient*

We now consider the characterization of minimal and maximal eigenvalues of a positive (semi)-definite symmetric matrix as the extrema of the Rayleigh quotient. This result can be stated as follows.

PROPOSITION 4.108.– *Let* $\mathbf{A} \in \mathbb{R}^{I \times I}$ *be a real positive definite (or semi-definite) symmetric matrix. The extrema of the Rayleigh quotient* $\lambda$ *of vector* $\mathbf{x}$ *with respect to* $\mathbf{A}$*, defined as:*

$$\lambda = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \quad \textit{with } \mathbf{x} \neq \mathbf{0}, \tag{4.98}$$

*are eigenvalues of* $\mathbf{A}$*. The extrema of this ratio are therefore given by the largest and smallest eigenvalues of* $\mathbf{A}$.

PROOF.– Re-expressing the quotient as $\lambda = (\mathbf{x}^T \mathbf{A} \mathbf{x})(\mathbf{x}^T \mathbf{x})^{-1}$, the gradient of $\lambda$ with respect to $\mathbf{x}$ is:

$$\frac{\partial \lambda}{\partial \mathbf{x}} = 2\frac{\mathbf{A}\mathbf{x}}{(\mathbf{x}^T \mathbf{x})} - 2\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{(\mathbf{x}^T \mathbf{x})^2}\mathbf{x} = \frac{2}{\mathbf{x}^T \mathbf{x}}(\mathbf{A} - \lambda \mathbf{I}_I)\mathbf{x}.$$

The necessary condition to have an extremum of $\lambda$ is obtained by canceling the gradient, which gives $(\mathbf{A} - \lambda \mathbf{I}_I)\mathbf{x} = \mathbf{0}_I$. Therefore, any extremum (global minimum or global maximum) of [4.98] is a root of the characteristic equation of $\mathbf{A}$, in other words, an eigenvalue of $\mathbf{A}$. Concerning the second part of the proposition, it should be observed that a global maximum and a global minimum of the Rayleigh ratio are obtained by replacing in [4.98] the vector $\mathbf{x}$ by eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_I$ associated with the largest and smallest eigenvalues, $\lambda_1 \geq 0$ and $\lambda_I \geq 0$, respectively, that is:

$$0 \leq \lambda_I \leq \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \leq \lambda_1, \quad \text{with} \quad \lambda_1 = \frac{\mathbf{x}_1^T \mathbf{A} \mathbf{x}_1}{\mathbf{x}_1^T \mathbf{x}_1}, \quad \lambda_I = \frac{\mathbf{x}_I^T \mathbf{A} \mathbf{x}_I}{\mathbf{x}_I^T \mathbf{x}_I}. \qquad \square$$

The eigenvalues can also be interpreted in terms of extrema of the quadratic form $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, under the constraint $\|\mathbf{x}\|_2^2 = 1$. Indeed, this constrained optimization problem of the equality type can be solved by using the Lagrangian

$L(\mathbf{x}, \lambda) = q(\mathbf{x}) - \lambda(\mathbf{x}^T\mathbf{x} - 1)$, where $\lambda$ is the Lagrange multiplier. The Karush–Kuhn–Tucker optimality conditions give:

$$\frac{\partial L(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} - 2\lambda\mathbf{x} = 0 \quad , \quad \frac{\partial L(\mathbf{x}, \lambda)}{\partial \lambda} = 1 - \mathbf{x}^T\mathbf{x} = 0,$$

from which we deduce that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, with $\mathbf{x}^T\mathbf{x} = 1$. By carrying over these expressions into $q(\mathbf{x})$, we obtain the following value of the optimized cost function: $q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \lambda\mathbf{x}^T\mathbf{x} = \lambda$.

Therefore, optimizing the quadratic form $q(\mathbf{x})$, under the constraint $\|\mathbf{x}\|_2^2 = 1$, is equivalent to calculating the eigenvalues of $\mathbf{A}$, and the value of the optimized criterion is equal to the largest eigenvalue ($\lambda_1$) or to the smallest eigenvalue ($\lambda_I$), depending on whether the optimization corresponds to a maximization or minimization of $q(\mathbf{x})$.

## 4.17. Generalized eigenvalues

Given $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{I\times I}$, the notion of generalized eigenvalue is defined by means of the following equation:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{B}\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}. \tag{4.99}$$

The generalized eigenvalues $\lambda$ are determined by solving the equation:

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

Using a similar method as for the interpretation of eigenvalues as extrema of the Rayleigh quotient, it is easy to show that the calculation of the generalized eigenvalues defined by equation [4.99] is equivalent to searching for the extrema of the following ratio of quadratic forms:

$$\lambda = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} \tag{4.100}$$

where $\mathbf{A}$ and $\mathbf{B}$ are, respectively, a positive definite or semi-definite symmetric matrix, and a positive definite symmetric matrix. The absolute extrema of the ratio [4.100] are solutions to the equation:

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

or equivalently:

$$(\lambda\mathbf{I}_n - \mathbf{B}^{-1}\mathbf{A})\mathbf{x} = \mathbf{0},$$

that is, $\lambda$ and $\mathbf{x}$ are an eigenvalue and an eigenvector of $\mathbf{B}^{-1}\mathbf{A}$.

Consequently, the maximum and minimum of the ratio $\lambda$ defined in [4.100] are, respectively, given by the largest ($\lambda_1$) and smallest ($\lambda_I$) eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$. The values of this maximum and minimum are obtained by replacing $\mathbf{x}$ in [4.100] by the generalized eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_I$ associated with the generalized eigenvalues, that is, $\lambda_I \leq \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} \leq \lambda_1$, with $\lambda_1 = \frac{\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1}{\mathbf{x}_1^T\mathbf{B}\mathbf{x}_1}$ and $\lambda_I = \frac{\mathbf{x}_I^T\mathbf{A}\mathbf{x}_I}{\mathbf{x}_I^T\mathbf{B}\mathbf{x}_I}$.

EXAMPLE 4.109 (Application to matched filtering).– Given a stochastic process $s(k)$ observed over a finite time interval $k \in [0, K-1]$, with additive noise $v(k)$, according to equation $x(k) = s(k) + v(k)$, the matched filter is a transverse filter of impulse response $h(k)$ acting on the signal observed according to the convolution equation:

$$\begin{aligned} y(k) = h(k) * x(k) &= \sum_{i=0}^{n} h(i)x(k-i) \\ &= \sum_{i=0}^{n} h(i)s(k-i) + \sum_{i=0}^{n} h(i)v(k-i) \\ &= s_f(k) + v_f(k), \end{aligned}$$

where $s_f(k) = \mathbf{h}^T\mathbf{s}(k)$ and $v_f(k) = \mathbf{h}^T\mathbf{v}(k)$, respectively, represent the signals $s(k)$ and $v(k)$ filtered by the matched filter, with $\mathbf{h} = [h(0)\cdots h(n)]^T$, $\mathbf{s}(k) = [s(k)\cdots s(k-n)]^T$, and $\mathbf{v}(k) = [v(k)\cdots v(k-n)]^T$.

The matched filter is optimized by maximizing the signal-to-noise ratio (SNR) defined as:

$$\text{SNR} = \frac{E[s_f^2(k)]}{E[v_f^2(k)]},$$

where $E$ is the mathematical expectation. Assuming that processes $s(k)$ and $v(k)$ are second-order stationary, we have:

$$\begin{aligned} E[s_f^2(k)] &= E[\mathbf{h}^T\mathbf{s}(k)\mathbf{s}^T(k)\mathbf{h}] = \mathbf{h}^T\mathbf{R}_{ss}\mathbf{h} \\ E[v_f^2(k)] &= E[\mathbf{h}^T\mathbf{v}(k)\mathbf{v}^T(k)\mathbf{h}] = \mathbf{h}^T\mathbf{R}_{vv}\mathbf{h} \end{aligned}$$

where $\mathbf{R}_{ss}$ and $\mathbf{R}_{vv}$ are the $n+1$th-order autocorrelation matrices of signals $s(k)$ and $v(k)$, respectively. The SNR ratio can then be written as:

$$\text{SNR}(\mathbf{h}) = \frac{\mathbf{h}^T\mathbf{R}_{ss}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{vv}\mathbf{h}}.$$

Subsequently, to maximize the SNR with respect to the coefficients of the impulse response of the matched filter, that is, with respect to the vector $\mathbf{h}$, amounts to

determining the largest generalized eigenvalue $\lambda_{\max}$, or equivalently the largest eigenvalue of $\mathbf{R}_{vv}^{-1}\mathbf{R}_{ss}$, the optimal solution $\mathbf{h}_{\text{opt}}$ being the eigenvector associated with $\lambda_{\max}$:

$$\max_{\mathbf{h}} \text{SNR}(\mathbf{h}) = \max_{\mathbf{h}} \frac{\mathbf{h}^T\mathbf{R}_{ss}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{vv}\mathbf{h}} = \lambda_{\max} \quad \Leftrightarrow \quad \mathbf{R}_{ss}\mathbf{h}_{\text{opt}} = \lambda_{\max}\mathbf{R}_{vv}\mathbf{h}_{\text{opt}}.$$

When the additive noise $v(k)$ is white, of variance $\sigma^2$, then $\mathbf{R}_{vv} = \sigma^2\mathbf{I}_{n+1}$, and the matched filter is obtained from a simple calculation of eigenvalues:

$$\frac{1}{\sigma^2}\mathbf{R}_{ss}\mathbf{h}_{\text{opt}} = \lambda_{\max}\mathbf{h}_{\text{opt}}.$$

$\mathbf{h}_{\text{opt}}$ is thus the eigenvector associated with the largest eigenvalue of $\frac{1}{\sigma^2}\mathbf{R}_{ss}$.

# 5

# Partitioned Matrices

## 5.1. Introduction

Partitioned matrices, also called block matrices, play an important role in matrix and tensor calculus. They are usually employed for the computation of matrix products, especially for Khatri–Rao and Kronecker products, whose properties will be examined in Volume 2. Partitioned matrices are also underlying the definition of certain structured matrices such as Hamiltonian, Hadamard, Fourier, Toeplitz, and Hankel matrices, and subsequently block-Toeplitz and block-Hankel matrices. In Volume 2, these structured matrices will be presented in more detail. As we shall see in this second volume, quaternionic matrices can also be written as partitioned matrices.

This chapter has several objectives:

– to define the notions of submatrices (section 5.2) and partitioned matrices (section 5.3);

– to describe examples of partitioned matrices for the computation of matrix products (sections 5.4, 5.11, and 5.12);

– to present a few special cases such as block-diagonal matrices, Jordan forms, block-triangular matrices, block-Toeplitz and Hankel matrices (section 5.5);

– to define block operations such as transposition (section 5.6), trace (section 5.7), addition (section 5.9) and multiplication (section 5.10), as well as the determinants (section 5.16), and the ranks (section 5.17) of certain partitioned matrices;

– to introduce elementary operations and associated matrices (section 5.13), used for block triangularization, block-diagonalization, block-factorization, block-inversion, and generalized inversion of $2 \times 2$ block matrices (sections 5.14 and 5.15);

*From Algebraic Structures to Tensors,*
First Edition. Edited by Gérard Favier.

– to use inversion formulae of block matrices to deduce several fundamental results such as the matrix inversion lemma, the inversion of a partitioned Gram matrix, and recursive inversion with respect to the order of a square partitioned matrix (section 5.14);

– to provide an example of application of the recursive inversion formula of a $2 \times 2$ block matrix, for demonstrating the Levinson algorithm which is an algorithm widely used in signal processing for the estimation of the parameters of an autoregressive (AR) model, and for the linear prediction problem (section 5.18).

## 5.2. Submatrices

A submatrix $\mathbf{B}(m_i, n_j)$ of a matrix $\mathbf{A}(m, n)$, with $m_i \leq m$ and $n_j \leq n$, is a matrix whose elements are positioned at the intersections of the $m_i$ rows and $n_j$ columns of $\mathbf{A}$ defined by the sets of indices:

$$\alpha_{m_i} = \{i_k, k \in \langle m_i \rangle\} \subseteq \langle m \rangle \ , \ \beta_{n_j} = \{j_l, l \in \langle n_j \rangle\} \subseteq \langle n \rangle.$$

Thus, the element $a_{i_k j_l}$ of $\mathbf{A}$ is given by $a_{i_k j_l} = (\mathbf{e}_{i_k}^{(m)})^T \mathbf{A} \mathbf{e}_{j_l}^{(n)}$. Subsequently, by defining the row and column selection matrices:

$$\mathbf{M} = [\mathbf{e}_{i_1}^{(m)}, \cdots, \mathbf{e}_{i_{m_i}}^{(m)}] \text{ and } \mathbf{N} = [\mathbf{e}_{j_1}^{(n)}, \cdots, \mathbf{e}_{j_{n_j}}^{(n)}], \quad [5.1]$$

we can write $\mathbf{B}(m_i, n_j)$ as:

$$\mathbf{B}(m_i, n_j) = \mathbf{M}^T \mathbf{A} \mathbf{N} = \begin{bmatrix} a_{i_1, j_1} & \cdots & a_{i_1, j_{n_j}} \\ \vdots & & \vdots \\ a_{i_{m_i}, j_1} & \cdots & a_{i_{m_i}, j_{n_j}} \end{bmatrix}. \quad [5.2]$$

In the case of a square matrix $\mathbf{A}$ of order $n$, a principal submatrix of order $r$ is a submatrix $\mathbf{B}(r, r)$ whose elements are positioned at the intersections of the same set of $r$ rows and $r$ columns, that is, defined by the same set of indices $\alpha_r = \{i_k, k \in \langle r \rangle\} \subseteq \langle n \rangle$. A principal submatrix of order $r$ contains $r$ elements of the main diagonal of $\mathbf{A}$.

There are $C_n^r = \frac{n!}{r!(n-r)!}$ principal submatrices of order $r$.

EXAMPLE 5.1.– For

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

there are $C_3^2 = 3$ principal submatrices of order two which are:

$$\mathbf{A}_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \ , \ \mathbf{A}_2 = \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix} \ , \ \mathbf{A}_3 = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}.$$

## 5.3. Partitioned matrices

Let $\{\alpha_{m_1}, \cdots, \alpha_{m_R}\}$ and $\{\beta_{n_1}, \cdots, \beta_{n_S}\}$ be partitions of the sets $\{1, \cdots, m\}$ and $\{1, \cdots, n\}$, respectively, with $m_r \in \langle m \rangle$ and $n_s \in \langle n \rangle$, such that $\sum_{r=1}^{R} m_r = m$ and $\sum_{s=1}^{S} n_s = n$. It is said that matrices $\mathbf{A}_{rs}$ of dimensions $(m_r, n_s)$ form a partition of the matrix $\mathbf{A} \in \mathbb{K}^{m \times n}$ into $(R, S)$ blocks, or that $\mathbf{A}$ is partitioned into $(R, S)$ blocks, if $\mathbf{A}$ can be written as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1S} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{R1} & \mathbf{A}_{R2} & \cdots & \mathbf{A}_{RS} \end{bmatrix} = [\mathbf{A}_{rs}], \ r \in \langle R \rangle, \ s \in \langle S \rangle. \qquad [5.3]$$

Such a partitioning with blocks of different dimensions is said to be unbalanced.

The submatrix $\mathbf{A}_{rs}$ can be expressed as:

$$\mathbf{A}_{rs} = \begin{bmatrix} a_{m_1+\cdots+m_{r-1}+1, n_1+\cdots+n_{s-1}+1} & \cdots & a_{m_1+\cdots+m_{r-1}+1, n_1+\cdots+n_{s-1}+n_s} \\ \vdots & \ddots & \vdots \\ a_{m_1+\cdots+m_{r-1}+m_r, n_1+\cdots+n_{s-1}+1} & \cdots & a_{m_1+\cdots+m_{r-1}+m_r, n_1+\cdots+n_{s-1}+n_s} \end{bmatrix} \in \mathbb{K}^{m_r \times n_s}.$$

All submatrices of the same row-block $(r)$ contain the same number $(m_r)$ of rows. Similarly, all submatrices of the same column-block $(s)$ contain the same number $(n_s)$ of columns, that is:

$$\begin{bmatrix} \mathbf{A}_{r1} & \mathbf{A}_{r2} & \cdots & \mathbf{A}_{rS} \end{bmatrix} \in \mathbb{K}^{m_r \times n}, \quad \begin{bmatrix} \mathbf{A}_{1s} \\ \mathbf{A}_{2s} \\ \vdots \\ \mathbf{A}_{Rs} \end{bmatrix} \in \mathbb{K}^{m \times n_s}. \qquad [5.4]$$

It is then said that the submatrices $\mathbf{A}_{rs}$ are of compatible dimensions.

In the particular case where $n = 1$, the partitioned matrix [5.3] becomes a block-column vector:

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_R \end{bmatrix} \in \mathbb{K}^{m \times 1}, \ \mathbf{a}_r \in \mathbb{K}^{m_r \times 1}, \ r \in \langle R \rangle.$$

Similarly, when $m = 1$, the partitioned matrix [5.3] becomes a block-row vector:

$$\mathbf{a}^T = \begin{bmatrix} \mathbf{a}_1^T & \mathbf{a}_2^T & \cdots & \mathbf{a}_S^T \end{bmatrix} \in \mathbb{K}^{1 \times n}, \ \mathbf{a}_s \in \mathbb{K}^{n_s \times 1}, \ s \in \langle S \rangle.$$

NOTE 5.2.– If all the blocks $\mathbf{A}_{rs}$ have the same dimensions $P \times Q$, that is, when $m_r = P, \forall r \in \langle R \rangle$, and $n_s = Q, \forall s \in \langle S \rangle$, then the space of partitioned matrices into $(R, S)$ blocks, with entries in the space $\mathbb{K}^{P \times Q}$ (also written $\mathcal{M}_{P \times Q}(\mathbb{K})$), will be denoted $\mathcal{M}_{R \times S}\Big(\mathcal{M}_{P \times Q}(\mathbb{K})\Big)$. The partitioning is then said to be balanced.

## 5.4. Matrix products and partitioned matrices

### 5.4.1. *Matrix products*

Given two rectangular matrices $\mathbf{A} \in \mathbb{K}^{I \times J}$ and $\mathbf{B} \in \mathbb{K}^{J \times K}$, the product $\mathbf{C} = \mathbf{AB} \in \mathbb{K}^{I \times K}$ can be written in terms of matrices partitioned into column blocks or row blocks:

$$\mathbf{AB} = \left[\mathbf{AB}_{.1}, \mathbf{AB}_{.2}, \cdots, \mathbf{AB}_{.K}\right] = \begin{bmatrix} \mathbf{A}_{1.}\mathbf{B} \\ \mathbf{A}_{2.}\mathbf{B} \\ \vdots \\ \mathbf{A}_{I.}\mathbf{B} \end{bmatrix}.$$

Two matrix products play an important role in matrix calculation. These are the Kronecker and Khatri–Rao products.

### 5.4.2. *Vector Kronecker product*

Let $\mathbf{u} \in \mathbb{K}^I$ and $\mathbf{v} \in \mathbb{K}^J$. Their Kronecker product is defined as:

$$\mathbf{x} = \mathbf{u} \otimes \mathbf{v} = \begin{bmatrix} u_1\mathbf{v} \\ \vdots \\ u_I\mathbf{v} \end{bmatrix} \in \mathbb{K}^{IJ}$$

$$= [u_1v_1, u_1v_2, \cdots, u_1v_J, u_2v_1, \cdots, u_Iv_J]^T.$$

This is a vector partitioned into $I$ blocks of dimension $J$. The element $u_iv_j$ is positioned at position $j + (i-1)J$.

### 5.4.3. *Matrix Kronecker product*

Given $\mathbf{A} \in \mathbb{K}^{I \times J}$ and $\mathbf{B} \in \mathbb{K}^{M \times N}$, the Kronecker product to the right of $\mathbf{A}$ by $\mathbf{B}$ is the matrix $\mathbf{C} \in \mathbb{K}^{IM \times JN}$ defined as :

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2J}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & \cdots & a_{IJ}\mathbf{B} \end{bmatrix} = [a_{ij}\mathbf{B}]. \quad [5.5]$$

This is a matrix partitioned into $(I, J)$ blocks, the block $(i, j)$ being the matrix $a_{ij}\mathbf{B} \in \mathbb{K}^{M\times N}$. The element $a_{ij}b_{mn}$ is positioned at position $((i-1)M+m, (j-1)N+n)$ in $\mathbf{A}\otimes\mathbf{B}$.

EXAMPLE 5.3.– For $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$, we have:

$$\mathbf{A}\otimes\mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \vdots & a_{12}\mathbf{B} \\ \cdots & & \cdots \\ a_{21}\mathbf{B} & \vdots & a_{22}\mathbf{B} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & \vdots & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & \vdots & a_{12}b_{21} & a_{12}b_{22} \\ \cdots & \cdots & & \cdots & \cdots \\ a_{21}b_{11} & a_{21}b_{12} & \vdots & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & \vdots & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix} \quad [5.6]$$

The $j$th column-block of $\mathbf{A}\otimes\mathbf{B}$ is given by:

$$\mathbf{A}_{.j}\otimes\mathbf{B} = \begin{bmatrix} a_{1j}\mathbf{B} \\ \vdots \\ a_{Ij}\mathbf{B} \end{bmatrix} = \left[\, \mathbf{A}_{.j}\otimes\mathbf{B}_{.1} \;\; \mathbf{A}_{.j}\otimes\mathbf{B}_{.2} \cdots \mathbf{A}_{.j}\otimes\mathbf{B}_{.N} \,\right] \;,\; j \in \langle J\rangle.$$

Subsequently, the columns of $\mathbf{A}\otimes\mathbf{B}$ are composed of all the Kronecker products of a column of $\mathbf{A}$ with a column of $\mathbf{B}$, the columns being taken in lexicographical order. Similarly, $\mathbf{A}\otimes\mathbf{B}$ can be decomposed into $I$ row-blocks $\mathbf{A}_{i.}\otimes\mathbf{B}$, with $i \in \langle I\rangle$, the $IM$ rows being composed of all the Kronecker products of a row of $\mathbf{A}$ with a row of $\mathbf{B}$.

Therefore, $\mathbf{A}\otimes\mathbf{B}$ can be broken into blocks such that:

$$\begin{aligned} \mathbf{A}\otimes\mathbf{B} &= \left[\, \mathbf{A}_{.1}\otimes\mathbf{B} \cdots \mathbf{A}_{.J}\otimes\mathbf{B} \,\right] \\ &= \left[\, \mathbf{A}_{.1}\otimes\mathbf{B}_{.1} \cdots \mathbf{A}_{.1}\otimes\mathbf{B}_{.N} \cdots \mathbf{A}_{.J}\otimes\mathbf{B}_{.1} \cdots \mathbf{A}_{.J}\otimes\mathbf{B}_{.N} \,\right] \\ &= \begin{bmatrix} \mathbf{A}_{1.}\otimes\mathbf{B} \\ \vdots \\ \mathbf{A}_{I.}\otimes\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{1.}\otimes\mathbf{B}_{1.} \\ \vdots \\ \mathbf{A}_{1.}\otimes\mathbf{B}_{M.} \\ \vdots \\ \mathbf{A}_{I.}\otimes\mathbf{B}_{1.} \\ \vdots \\ \mathbf{A}_{I.}\otimes\mathbf{B}_{M.} \end{bmatrix}. \end{aligned}$$

The Kronecker product can be used to write the matrix $\mathbf{A}$ partitioned into $(R, S)$ blocks, defined in [5.3], as follows:

$$\mathbf{A} = \sum_{r=1}^{R} \sum_{s=1}^{S} \mathbf{E}_{rs}^{(R\times S)} \otimes \mathbf{A}_{rs}, \qquad [5.7]$$

where $\mathbf{E}_{rs}^{(R\times S)}$, for $r \in \langle R\rangle$ and $s \in \langle S\rangle$, are the matrices of the canonical basis of the space $\mathbb{K}^{R\times S}$, that is, with the $(r, s)$th element equal to 1 and all others equal to zero.

EXAMPLE 5.4.– For $R = 2$ and $S = 3$, we have:

$$\mathbf{A} = \sum_{r=1}^{2} \sum_{s=1}^{3} \mathbf{E}_{rs}^{(2\times 3)} \otimes \mathbf{A}_{rs} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \end{bmatrix}.$$

### 5.4.4. *Khatri–Rao product*

Given $\mathbf{A} \in \mathbb{K}^{I\times J}$ and $\mathbf{B} \in \mathbb{K}^{K\times J}$ having the same number of columns, the Khatri–Rao product of $\mathbf{A}$ with $\mathbf{B}$, denoted by $\mathbf{A} \diamond \mathbf{B} \in \mathbb{K}^{IK\times J}$, is defined as:

$$\mathbf{A} \diamond \mathbf{B} = \left[ \mathbf{A}_{.1} \otimes \mathbf{B}_{.1} \ , \ \mathbf{A}_{.2} \otimes \mathbf{B}_{.2} \ , \ \cdots \ , \ \mathbf{A}_{.J} \otimes \mathbf{B}_{.J} \right] \qquad [5.8]$$

This is a matrix that is partitioned into $J$ column-blocks, the $j$th block being equal to the Kronecker product of the $j$th column of $\mathbf{A}$ with the $j$th column of $\mathbf{B}$. It is said that $\mathbf{A} \diamond \mathbf{B}$ is a columnwise Kronecker product of $\mathbf{A}$ and $\mathbf{B}$.

PROPOSITION 5.5.– *The Khatri–Rao product can also be written as a matrix partitioned into $I$ row-blocks:*

$$\mathbf{A} \diamond \mathbf{B} = \begin{bmatrix} \mathbf{BD}_1(\mathbf{A}) \\ \mathbf{BD}_2(\mathbf{A}) \\ \vdots \\ \mathbf{BD}_I(\mathbf{A}) \end{bmatrix} \qquad [5.9]$$

*where $\mathbf{D}_i(\mathbf{A}) = diag(a_{i1}, a_{i2}, \cdots, a_{iJ})$ refers to the diagonal matrix whose diagonal elements are the elements of the $i$th row of $\mathbf{A}$.*

PROOF.– By definition of the Khatri–Rao product, the $i$th row-block is given by:

$$\begin{aligned} \left[ a_{i1}\mathbf{B}_{.1} \cdots a_{iJ}\mathbf{B}_{.J} \right] &= \left[ \mathbf{B}_{.1} \cdots \mathbf{B}_{.J} \right] \mathrm{diag}(a_{i1}, \cdots, a_{iJ}) \\ &= \mathbf{BD}_i(\mathbf{A}) \in \mathbb{C}^{K\times J}, \end{aligned} \qquad [5.10]$$

from which [5.9] is deduced by stacking the row-blocks [5.10] from $i = 1$ to $i = I$. □

## 5.5. Special cases of partitioned matrices

### 5.5.1. *Block-diagonal matrices*

A square matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ partitioned into $(R, R)$ blocks, of diagonal blocks $\mathbf{A}_{rr} \in \mathbb{K}^{n_r\times n_r}$, with $r \in \langle R\rangle$ and $\sum_{r=1}^{R} n_r = n$, whose off-diagonal blocks are zero, is called a block-diagonal matrix and can be written as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0}_{n_1\times n_2} & \cdots & \mathbf{0}_{n_1\times n_R} \\ \mathbf{0}_{n_2\times n_1} & \mathbf{A}_{22} & \cdots & \mathbf{0}_{n_2\times n_R} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_R\times n_1} & \mathbf{0}_{n_R\times n_2} & \cdots & \mathbf{A}_{RR} \end{bmatrix}. \qquad [5.11]$$

It is also written $\text{diag}(\mathbf{A}_{11}, \mathbf{A}_{22}, \cdots, \mathbf{A}_{RR})$ or simply $\text{diag}(\mathbf{A}_{rr})$ with the number $R$ of diagonal blocks implied.

### 5.5.2. *Signature matrices*

The signature matrix of a symmetric matrix $\mathbf{A}$, of full rank, is a diagonal matrix whose diagonal elements are equal to 1 or $-1$ (see section 4.15.7):

$$\mathbf{S} = \text{diag}(\underbrace{1, \cdots, 1}_{p \text{ terms}}, \underbrace{-1, \cdots, -1}_{q \text{ terms}}) \text{ with } p \geq 0, q \geq 0.$$

$p$ and $q$ correspond to the numbers of positive and negative eigenvalues of $\mathbf{A}$, respectively. A signature matrix is thus a block-diagonal matrix consisting of two diagonal blocks $\mathbf{I}_p$ and $-\mathbf{I}_q$:

$$\mathbf{S} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & -\mathbf{I}_q \end{bmatrix}$$

### 5.5.3. *Direct sum*

Given $\mathbf{A} \in \mathbb{K}^{I\times J}$ and $\mathbf{B} \in \mathbb{K}^{K\times L}$, the direct sum of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A} \oplus \mathbf{B}$, is the block-diagonal matrix $\begin{bmatrix} \mathbf{A} & \mathbf{0}_{I\times L} \\ \mathbf{0}_{K\times J} & \mathbf{B} \end{bmatrix} \in \mathbb{K}^{(I+K)\times(J+L)}$. In the case of $P$ matrices $\mathbf{A}^{(p)} \in \mathbb{K}^{I_p\times J_p}$, we have:

$$\bigoplus_{p=1}^{P} \mathbf{A}_p = \mathbf{A}_1 \oplus \mathbf{A}_2 \oplus \cdots \oplus \mathbf{A}_P \in \mathbb{K}^{\sum_{p=1}^{P} I_p \times \sum_{p=1}^{P} J_p}$$

$$= \text{diag}(\mathbf{A}_1, \cdots, \mathbf{A}_P) = \text{diag}(\mathbf{A}_p).$$

### 5.5.4. *Jordan forms*

A non-diagonalizable matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ can be transformed into a Jordan form:

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & & \mathbf{0} \\ & \mathbf{B}_2 & \\ & & \ddots \\ \mathbf{0} & & \mathbf{B}_p \end{bmatrix} \quad [5.12]$$

$$\mathbf{B}_i = \begin{bmatrix} \lambda_i & 1 & & \mathbf{0} \\ & \lambda_i & 1 & \\ & & \ddots & 1 \\ \mathbf{0} & & & \lambda_i \end{bmatrix} \in \mathbb{K}^{n_i\times n_i} \ , \ i \in \langle p \rangle$$

where $\{\lambda_1, \cdots, \lambda_p\}$ are the eigenvalues of $\mathbf{A}$, and $n_i$ is the multiplicity order of $\lambda_i$. Block-diagonal decomposition [5.12] into Jordan blocks $\mathbf{B}_i$, is called the Jordan form of $\mathbf{A}$. This decomposition is little used in practice because its numerical determination may be unstable.

### 5.5.5. *Block-triangular matrices*

When $\mathbf{A}_{rs} = \mathbf{0}_{m_r,n_s}$ for $s < r$ in [5.3], that is:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1S} \\ \mathbf{0} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{RS} \end{bmatrix},$$

$\mathbf{A}$ is said to be an upper block-triangular matrix. Similarly, when $\mathbf{A}_{rs} = \mathbf{0}_{m_r,n_s}$ for $s > r$, that is:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{R1} & \mathbf{A}_{R2} & \cdots & \mathbf{A}_{RS} \end{bmatrix},$$

then $\mathbf{A}$ is called a lower block-triangular matrix.

### 5.5.6. *Block Toeplitz and Hankel matrices*

An $I \times J$ block-Toeplitz matrix is an $IM \times JN$ matrix partitioned in the form:

$$\mathbf{A} = \begin{bmatrix} \mathbf{F}_0 & \mathbf{F}_{-1} & \mathbf{F}_{-2} & \cdots & \mathbf{F}_{1-J} \\ \mathbf{F}_1 & \mathbf{F}_0 & \mathbf{F}_{-1} & \cdots & \mathbf{F}_{2-J} \\ \mathbf{F}_2 & \mathbf{F}_1 & \mathbf{F}_0 & \cdots & \mathbf{F}_{3-J} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}_{I-1} & \mathbf{F}_{I-2} & \mathbf{F}_{I-3} & \cdots & \mathbf{F}_0 \end{bmatrix}, \quad [5.13]$$

where $\mathbf{F}_t \in \mathbb{C}^{M \times N}$, with $1 - J \leq t \leq I - 1$. When $M = N = 1$, we have a standard $I \times J$ Toeplitz matrix.

An $IJ \times IJ$ block-Hankel matrix is of the form:

$$\mathbf{A} = \begin{bmatrix} \mathbf{F}_0 & \mathbf{F}_1 & \mathbf{F}_2 & \cdots & \mathbf{F}_I \\ \mathbf{F}_1 & \mathbf{F}_2 & \mathbf{F}_3 & \cdots & \mathbf{F}_{I+1} \\ \mathbf{F}_2 & \mathbf{F}_3 & \mathbf{F}_4 & \cdots & \mathbf{F}_{I+2} \\ \vdots & \vdots & \vdots & \iddots & \vdots \\ \mathbf{F}_I & \mathbf{F}_{I+1} & \mathbf{F}_{I+2} & \cdots & \mathbf{F}_{2I} \end{bmatrix}, \quad [5.14]$$

where $\mathbf{F}_t$ is a $J \times J$ matrix for $t = 0, 1, \cdots, 2I$. As a Hankel matrix $\mathbf{A} = [a_{ij}] = [a_{i+j}]$, with $0 \leq i, j \leq I$, is determined by its first column and last row, a block-Hankel matrix is such that $\mathbf{A} = [\mathbf{A}_{ij}] = [\mathbf{F}_{i+j}]$ with $0 \leq i, j \leq I$. When each block $\mathbf{F}_t$ is a Hankel matrix, then $\mathbf{A}$ is a block-Hankel matrix with Hankel blocks.

## 5.6. Transposition and conjugate transposition

The transposition (or conjugate transposition) of a matrix $\mathbf{A}$ partitioned into $(R, S)$ blocks $\mathbf{A}_{rs}, r \in \langle R \rangle, s \in \langle S \rangle$, is obtained by transposing (or transconjugating) the blocks, followed by a blockwise transposition.

EXAMPLE 5.6.– For $R = S = 2$, we have:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \Rightarrow \mathbf{M}^T = \begin{bmatrix} \mathbf{A}^T & \mathbf{C}^T \\ \mathbf{B}^T & \mathbf{D}^T \end{bmatrix}, \mathbf{M}^H = \begin{bmatrix} \mathbf{A}^H & \mathbf{C}^H \\ \mathbf{B}^H & \mathbf{D}^H \end{bmatrix}.$$

The following proposition can be deduced.

PROPOSITION 5.7.– *For a matrix partitioned into* $(2, 2)$ *blocks with square blocks of same dimensions:*

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \quad [5.15]$$

*we have:*

$$\mathbf{M}\,\textit{symmetric} \;\Leftrightarrow\; \mathbf{A}^T = \mathbf{A},\ \mathbf{C} = \mathbf{B}^T,\ \mathbf{D}^T = \mathbf{D}$$

$$\mathbf{M}\,\textit{Hermitian} \;\Leftrightarrow\; \mathbf{A}^H = \mathbf{A},\ \mathbf{C} = \mathbf{B}^H,\ \mathbf{D}^H = \mathbf{D},$$

*that is, the diagonal blocks must be symmetric/Hermitian and the off-diagonal blocks transposed/conjugate transposed with respect to one another.*

## 5.7. Trace

The trace of a partitioned matrix $\mathbf{A} = [\mathbf{A}_{rs}]$, with $r, s \in \langle R \rangle$, of dimensions $(n, n)$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1R} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{R1} & \mathbf{A}_{R2} & \cdots & \mathbf{A}_{RR} \end{bmatrix},$$

with $\dim(\mathbf{A}_{rr}) = (n_r, n_r)$ and $\sum_{r=1}^{R} n_r = n$, is given by: $\mathrm{tr}(\mathbf{A}) = \sum_{r=1}^{R} \mathrm{tr}(\mathbf{A}_{rr})$.

## 5.8. Vectorization

Let us consider a balanced partitioning of $\mathbf{A}$ into $(R, S)$ blocks of dimensions $P \times Q$. The partitioned matrix $\mathbf{A}$ can be vectorized column-blockwise (or row-blockwise), that is, by vectorizing each column (or row) of blocks, and then stacking the resulting vectors. The corresponding vectorization operators are denoted as $\mathrm{vec}_c(.)$ and $\mathrm{vec}_r(.)$, respectively.

EXAMPLE 5.8.– For $R = S = 2$, we have:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \Rightarrow \mathrm{vec}_c(\mathbf{A}) = \begin{bmatrix} \mathrm{vec}(\mathbf{A}_{11}) \\ \mathrm{vec}(\mathbf{A}_{21}) \\ \mathrm{vec}(\mathbf{A}_{12}) \\ \mathrm{vec}(\mathbf{A}_{22}) \end{bmatrix},\ \mathrm{vec}_r(\mathbf{A}) = \begin{bmatrix} \mathrm{vec}(\mathbf{A}_{11}) \\ \mathrm{vec}(\mathbf{A}_{12}) \\ \mathrm{vec}(\mathbf{A}_{21}) \\ \mathrm{vec}(\mathbf{A}_{22}) \end{bmatrix}.$$

## 5.9. Blockwise addition

Let $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{I \times J}$ be two matrices partitioned into blocks having the same dimensions $\mathbf{A}_{rs}, \mathbf{B}_{rs} \in \mathbb{K}^{I_r \times J_s}$, with $\sum_{r=1}^{R} I_r = I, \sum_{s=1}^{S} J_s = J$. Their sum is a partitioned matrix $\mathbf{C} = \mathbf{A} + \mathbf{B} = [\mathbf{A}_{rs} + \mathbf{B}_{rs}]$, with $\mathbf{C}_{rs} = \mathbf{A}_{rs} + \mathbf{B}_{rs} \in \mathbb{K}^{I_r \times J_s}$.

EXAMPLE 5.9.– For $R = S = 2$, we have:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad , \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{bmatrix}.$$

## 5.10. Blockwise multiplication

Let $\mathbf{A} \in \mathbb{K}^{I \times J}$ and $\mathbf{B} \in \mathbb{K}^{J \times L}$ be two matrices partitioned into blocks $\mathbf{A}_{rs} \in \mathbb{K}^{I_r \times J_s}$ and $\mathbf{B}_{sn} \in \mathbb{K}^{J_s \times L_n}$, with $\sum_{r=1}^{R} I_r = I, \sum_{s=1}^{S} J_s = J$, and $\sum_{n=1}^{N} L_n = L$. The product $\mathbf{C} = \mathbf{AB}$ is a matrix that is partitioned into blocks $\mathbf{C}_{rn} = \sum_{s=1}^{S} \mathbf{A}_{rs}\mathbf{B}_{sn} \in \mathbb{K}^{I_r \times L_n}$.

EXAMPLE 5.10.– For $R = S = N = 2$, we have:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}.$$

EXAMPLE 5.11.– In the case of the product of two partitioned matrices built by adding a row and a column, respectively, we have:

$$\begin{bmatrix} \mathbf{X}^T \\ \mathbf{x}^T \end{bmatrix} \begin{bmatrix} \mathbf{Y} & \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{Y} & \mathbf{X}^T\mathbf{y} \\ \mathbf{x}^T\mathbf{Y} & \mathbf{x}^T\mathbf{y} \end{bmatrix}.$$

## 5.11. Hadamard product of partitioned matrices

It should be remembered first that the Hadamard product[1] of two matrices $\mathbf{A} \in \mathbb{K}^{I \times J}$ and $\mathbf{B} \in \mathbb{K}^{I \times J}$, of the same dimensions, gives a matrix $\mathbf{C} \in \mathbb{K}^{I \times J}$ defined as:

$$\mathbf{C} = \mathbf{A} \odot \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1J}b_{1J} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2J}b_{2J} \\ \vdots & \vdots & & \vdots \\ a_{I1}b_{I1} & a_{I2}b_{I2} & \cdots & a_{IJ}b_{IJ} \end{bmatrix} \qquad [5.16]$$

that is, $c_{ij} = a_{ij}b_{ij}$, and thus, $\mathbf{C} = [a_{ij}b_{ij}]$.

1 The Hadamard product is also known as the Schur product or the entrywise product.

Given $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{I \times J}$, partitioned into $(R, S)$ blocks $\mathbf{A}_{rs}, \mathbf{B}_{rs} \in \mathbb{K}^{I_r \times J_s}$, with $I = \sum_{r=1}^{R} I_r$ and $J = \sum_{s=1}^{S} J_s$ , then their Hadamard product $\mathbf{A} \odot \mathbf{B}$ is a partitioned matrix into $(R, S)$ blocks $\mathbf{C}_{rs} = \mathbf{A}_{rs} \odot \mathbf{B}_{rs} \in \mathbb{K}^{I_r \times J_s}$, with $r \in \langle R \rangle$, $s \in \langle S \rangle$.

Note that if $R = S = 1$, then the block Hadamard product becomes the classical Hadamard product [5.16].

As mentioned earlier, if all the blocks have the same dimensions $P \times Q$, that is, $I_r = P$, $\forall r \in \langle R \rangle$ , and $J_s = Q$, $\forall s \in \langle S \rangle$, then $\mathbf{A}$ and $\mathbf{B}$, and consequently $\mathbf{A} \odot \mathbf{B}$, belong to the space denoted by $\mathcal{M}_{R \times S}\Big(\mathcal{M}_{P \times Q}(\mathbb{K})\Big)$.

## 5.12. Kronecker product of partitioned matrices

Given a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ partitioned into $(R, S)$ blocks $\mathbf{A}_{rs} \in \mathbb{K}^{I_r \times J_s}$, with $\sum_{r=1}^{R} I_r = I$ and $\sum_{s=1}^{S} J_s = J$, and a matrix $\mathbf{B} \in \mathbb{K}^{M \times N}$, then their Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is a matrix partitioned into $(R, S)$ blocks $\mathbf{A}_{rs} \otimes \mathbf{B} \in \mathbb{K}^{I_r M \times J_s N}$.

EXAMPLE 5.12.– For $R = 2, S = 3$, we have:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \end{bmatrix} \Rightarrow \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} \otimes \mathbf{B} & \mathbf{A}_{12} \otimes \mathbf{B} & \mathbf{A}_{13} \otimes \mathbf{B} \\ \mathbf{A}_{21} \otimes \mathbf{B} & \mathbf{A}_{22} \otimes \mathbf{B} & \mathbf{A}_{23} \otimes \mathbf{B} \end{bmatrix}.$$

More generally, in the case of two matrices partitioned into blocks $\mathbf{A}_{rs} \in \mathbb{K}^{I_r \times J_s}$, with ($r \in \langle R \rangle$ , $s \in \langle S \rangle$), and $\mathbf{B}_{mn} \in \mathbb{K}^{K_m \times L_n}$, with ($m \in \langle M \rangle$ , $n \in \langle N \rangle$), the block Kronecker product, called the Tracy–Singh product (1972), is defined as:

$$\mathbf{A} \otimes_b \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} \otimes \mathbf{B}_{11} & \cdots & \mathbf{A}_{11} \otimes \mathbf{B}_{1N} & \cdots & \mathbf{A}_{1S} \otimes \mathbf{B}_{11} & \cdots & \mathbf{A}_{1S} \otimes \mathbf{B}_{1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_{11} \otimes \mathbf{B}_{M1} & \cdots & \mathbf{A}_{11} \otimes \mathbf{B}_{MN} & \cdots & \mathbf{A}_{1S} \otimes \mathbf{B}_{M1} & \cdots & \mathbf{A}_{1S} \otimes \mathbf{B}_{MN} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_{R1} \otimes \mathbf{B}_{11} & \cdots & \mathbf{A}_{R1} \otimes \mathbf{B}_{1N} & \cdots & \mathbf{A}_{RS} \otimes \mathbf{B}_{11} & \cdots & \mathbf{A}_{RS} \otimes \mathbf{B}_{1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_{R1} \otimes \mathbf{B}_{M1} & \cdots & \mathbf{A}_{R1} \otimes \mathbf{B}_{MN} & \cdots & \mathbf{A}_{RS} \otimes \mathbf{B}_{M1} & \cdots & \mathbf{A}_{RS} \otimes \mathbf{B}_{MN} \end{bmatrix}$$

of dimensions $(\sum_{r=1}^{R} \sum_{m=1}^{M} I_r K_m, \sum_{s=1}^{S} \sum_{n=1}^{N} J_s L_n)$. Note that if $R = S = M = N = 1$, this block Kronecker product becomes the classical Kronecker product [5.5], with $\mathbf{A} = \mathbf{A}_{11}$ and $\mathbf{B} = \mathbf{B}_{11}$.

Another Kronecker product of partitioned matrices, called the strong Kronecker product and denoted by $|\otimes|$, was introduced by de Launey and Seberry (1994) for generating orthogonal matrices from Hadamard matrices. This Kronecker product is also used to represent tensor train decompositions, in the case of large-scale tensors (Lee and Cichocki 2017). Given two matrices partitioned into blocks $\mathbf{A}_{rs} \in \mathbb{K}^{I \times J}$ and

$\mathbf{B}_{sn} \in \mathbb{K}^{K\times L}$, with $r \in \langle R\rangle$ , $s \in \langle S\rangle$, and $n \in \langle N\rangle$, the strong Kronecker product $\mathbf{A}|\otimes|\mathbf{B}$ is defined as the matrix partitioned into $(R, N)$ blocks $\mathbf{C}_{rn} \in \mathbb{K}^{IK\times JL}$, with $r \in \langle R\rangle$ and $n \in \langle N\rangle$, such as:

$$\mathbf{C}_{rn} = \sum_{s=1}^{S} \mathbf{A}_{rs} \otimes \mathbf{B}_{sn}.$$

This operation, which is completely determined by the parameters $(R, S, N)$, preserves the orthogonality.

In the next proposition, we present some properties of the block Hadamard and Kronecker products (Garcia-Bayona 2019).

PROPOSITION 5.13.– *The block Hadamard and Kronecker products satisfy the following properties:*

*– The matrix $\mathbf{1}_{RP\times SQ}$, whose all entries are equal to 1, is the identity element for $\odot$ in the space $\mathcal{M}_{R\times S}\Big(\mathcal{M}_{P\times Q}(\mathbb{K})\Big)$:*

$$\mathbf{A} \odot \mathbf{1}_{RP\times SQ} = \mathbf{1}_{RP\times SQ} \odot \mathbf{A} = \mathbf{A} \ , \ \forall \mathbf{A} \in \mathcal{M}_{R\times S}\Big(\mathcal{M}_{P\times Q}(\mathbb{K})\Big). \quad [5.17]$$

*Note that, for the Kronecker product, there is no identity element $\mathbf{E}$ such that $\mathbf{A}\otimes\mathbf{E} = \mathbf{E}\otimes\mathbf{A} = \mathbf{A}$ , $\forall\mathbf{A}$.*

*– Commutativity of $\odot$:*

$$\mathbf{A} \odot \mathbf{B} = \mathbf{B} \odot \mathbf{A}. \quad [5.18]$$

*– Associativity of $\odot$ and $\otimes$:*

$$\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} \quad [5.19]$$

$$\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C}. \quad [5.20]$$

*– Distributivity of $\odot$ and $\otimes$ over the addition:*

$$(\mathbf{A} + \mathbf{B}) \odot \mathbf{C} = (\mathbf{A} \odot \mathbf{C}) + (\mathbf{B} \odot \mathbf{C}) \quad [5.21]$$

$$\mathbf{A} \odot (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) + (\mathbf{A} \odot \mathbf{C}) \quad [5.22]$$

$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C}) \quad [5.23]$$

$$\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C}) \quad [5.24]$$

*– Distributivity of $\odot$ and $\otimes$ over the scalar multiplication. For any $\lambda \in \mathbb{K}$:*

$$\lambda(\mathbf{A} \odot \mathbf{B}) = (\lambda\mathbf{A}) \odot \mathbf{B} = \mathbf{A} \odot (\lambda\mathbf{B}) \quad [5.25]$$

$$\lambda(\mathbf{A} \otimes \mathbf{B}) = (\lambda\mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (\lambda\mathbf{B}) \quad [5.26]$$

As shown in section 4.2.3 for the set $\mathbb{K}^{I\times J}$, the set $\mathcal{M}_{R\times S}\Big(\mathcal{M}_{P\times Q}(\mathbb{K})\Big)$ of block matrices, equipped with addition and scalar multiplication, is a $\mathbb{K}$-v.s. From the properties presented in the above proposition, we can conclude that this set of block matrices equipped with the block Kronecker product as second internal law is an algebra, whereas it is a commutative algebra with the block Hadamard product operation as second internal law (see the definition of an algebra in section 2.5.16).

## 5.13. Elementary operations and elementary matrices

In this section, we present three types of elementary operations involving the rows or the columns of a matrix. These elementary operations, which can be represented by way of the so-called elementary matrices, are used to express a matrix in canonical form (i.e. unique) such as the echelon or normal form. These operations are described in the next proposition.

PROPOSITION 5.14.– *The elementary operations consist of:*

*– interchanging the $i$th and $j$th rows (columns):*

$$\mathbf{A}_{i.} \leftrightarrow \mathbf{A}_{j.}\ (\mathbf{A}_{.i} \leftrightarrow \mathbf{A}_{.j});$$

*– multiplying the elements of the $i$th row (column) by a scalar $k \neq 0$:*

$$k\mathbf{A}_{i.} \rightarrow \mathbf{A}_{i.}\ (k\mathbf{A}_{.i} \rightarrow \mathbf{A}_{.i});$$

*– adding to the elements of the $i$th row (column), the corresponding elements of the $j$th row (column) multiplied by $k$:*

$$\mathbf{A}_{i.} + k\mathbf{A}_{j.} \rightarrow \mathbf{A}_{i.}\ (\mathbf{A}_{.i} + k\mathbf{A}_{.j} \rightarrow \mathbf{A}_{.i}).$$

*The three corresponding transformations, respectively, denoted by $\mathbf{P}_{ij}$, $\mathbf{P}_i(k)$, and $\mathbf{P}_{ij}(k)$ can be represented using the so-called elementary matrices.*

*Designating by $(p_1, \cdots, p_n)$ a permutation of $(1, \cdots, n)$, these elementary matrices are such that:*

$$\begin{aligned}
\mathbf{P}_{ij} &= \left[\mathbf{e}_{p_1} \cdots \mathbf{e}_{p_n}\right] \text{ with } \begin{bmatrix} \mathbf{e}_{p_i} = \mathbf{e}_j \text{ and } \mathbf{e}_{p_j} = \mathbf{e}_i \\ \mathbf{e}_{p_k} = \mathbf{e}_k \text{ if } k \neq i \text{ and } j \end{bmatrix} \\
&= \mathbf{I} - (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \\
\mathbf{P}_i(k) &= \left[\mathbf{e}_1, \cdots, \mathbf{e}_{i-1}, k\mathbf{e}_i, \mathbf{e}_{i+1}, \cdots, \mathbf{e}_n\right] \\
\mathbf{P}_{ij}(k) &= \left[\mathbf{e}_1, \cdots, \mathbf{e}_{j-1}, \mathbf{e}'_j, \mathbf{e}_{j+1}, \cdots, \mathbf{e}_n\right] \quad \text{for rows,} \\
&= \mathbf{I} + k\mathbf{e}_i\mathbf{e}_j^T
\end{aligned}$$

$$\mathbf{P}_{ij}(k) = \Big[\mathbf{e}_1, \cdots, \mathbf{e}_{i-1}, \mathbf{e}'_i, \mathbf{e}_{i+1}, \cdots, \mathbf{e}_n\Big] \quad \textit{for columns},$$
$$= \mathbf{I} + k\mathbf{e}_j\mathbf{e}_i^T$$

*where* $\mathbf{e}'_j$ *(*$\mathbf{e}'_i$*) is a vector consisting of* 0*s except its* $i$*th (*$j$*th) component equal to* $k$ *and its* $j$*th (*$i$*th) component equal to* 1.

An elementary operation on rows (columns) is achieved via a pre-multiplication (post-multiplication) by the associated elementary matrix having one of the three above-mentioned forms.

EXAMPLE 5.15.– Consider $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.

For $\mathbf{P}_{12} = \mathbf{I}_3 - (\mathbf{e}_2 - \mathbf{e}_1)(\mathbf{e}_2 - \mathbf{e}_1)^T = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, we have:

$$\mathbf{P}_{12}\mathbf{A} = \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad , \quad \mathbf{A}\mathbf{P}_{12} = \begin{bmatrix} a_{12} & a_{11} & a_{13} \\ a_{22} & a_{21} & a_{23} \\ a_{32} & a_{31} & a_{33} \end{bmatrix}.$$

and

$$\mathbf{P}_{12}(k) = \mathbf{I}_3 + k\mathbf{e}_1\mathbf{e}_2^T = \begin{bmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \mathbf{P}_{12}(k)\mathbf{A}$$
$$= \begin{bmatrix} a_{11} + k\,a_{21} & a_{12} + k\,a_{22} & a_{13} + k\,a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$
$$\mathbf{P}_{12}(k) = \mathbf{I}_3 + k\mathbf{e}_2\mathbf{e}_1^T = \begin{bmatrix} 1 & 0 & 0 \\ k & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \mathbf{A}\mathbf{P}_{12}(k) = \begin{bmatrix} a_{11} + k\,a_{12} & a_{12} & a_{13} \\ a_{21} + k\,a_{22} & a_{22} & a_{23} \\ a_{31} + k\,a_{32} & a_{32} & a_{33} \end{bmatrix}.$$

Similarly, we can define elementary operations involving row-blocks or column-blocks of a partitioned matrix, to:

– interchange the $i$th and $j$th row-blocks (column-blocks);

– multiply the $i$th row-block (column-block) on the left-hand side (right-hand side) by a non-singular matrix;

– add the $i$th row-block (column-block) to the $j$th row-block (column-block) multiplied on the left-hand side (right-hand side) by a non-singular matrix.

For example, consider a matrix partitioned into (2,2) blocks, with square diagonal blocks:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \in \mathbb{C}^{(n+m)\times(n+m)}, \qquad [5.27]$$

$\mathbf{A}, \mathbf{E} \in \mathbb{C}^{n\times n}, \mathbf{D}, \mathbf{F} \in \mathbb{C}^{m\times m}, \mathbf{C}, \mathbf{G} \in \mathbb{C}^{m\times n}, \mathbf{B}, \mathbf{H} \in \mathbb{C}^{n\times m}$, with non-singular $\mathbf{E}$ and $\mathbf{F}$.

For the first type of elementary operation, we have:

$$\begin{bmatrix} \mathbf{0} & \mathbf{I}_m \\ \mathbf{I}_n & \mathbf{0} \end{bmatrix} \mathbf{M} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{A} & \mathbf{B} \end{bmatrix} \quad , \quad \mathbf{M} \begin{bmatrix} \mathbf{0} & \mathbf{I}_n \\ \mathbf{I}_m & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{B} & \mathbf{A} \\ \mathbf{D} & \mathbf{C} \end{bmatrix} \qquad [5.28]$$

For the second type of elementary operation:

$$\begin{bmatrix} \mathbf{E} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \mathbf{M} = \begin{bmatrix} \mathbf{EA} & \mathbf{EB} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \quad , \quad \mathbf{M} \begin{bmatrix} \mathbf{E} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{AE} & \mathbf{B} \\ \mathbf{CE} & \mathbf{D} \end{bmatrix} \qquad [5.29a]$$

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{FC} & \mathbf{FD} \end{bmatrix} \quad , \quad \mathbf{M} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{BF} \\ \mathbf{C} & \mathbf{DF} \end{bmatrix} \qquad [5.29b]$$

For the third type of elementary operation:

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{H} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \mathbf{M} = \begin{bmatrix} \mathbf{A} + \mathbf{HC} & \mathbf{B} + \mathbf{HD} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \qquad [5.30a]$$

$$\mathbf{M} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{G} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{A} + \mathbf{BG} & \mathbf{B} \\ \mathbf{C} + \mathbf{DG} & \mathbf{D} \end{bmatrix} \qquad [5.30b]$$

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{G} & \mathbf{I}_m \end{bmatrix} \mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} + \mathbf{GA} & \mathbf{D} + \mathbf{GB} \end{bmatrix} \qquad [5.30c]$$

$$\mathbf{M} \begin{bmatrix} \mathbf{I}_n & \mathbf{H} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} + \mathbf{AH} \\ \mathbf{C} & \mathbf{D} + \mathbf{CH} \end{bmatrix}. \qquad [5.30d]$$

## 5.14. Inversion of partitioned matrices

This section is devoted to the inversion of $2 \times 2$ block matrices. We must note that for a $2 \times 2$ block matrix $\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$, whose blocks $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and $\mathbf{D}$ have the dimensions $m \times n$, $m \times p$, $q \times n$ and $q \times p$, respectively, with $m+q = n+p$, its inverse $\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}$ is such that the blocks $\mathbf{E}, \mathbf{F}, \mathbf{G}$, and $\mathbf{H}$ must be of dimensions $n \times m$, $n \times q$, $p \times m$ and $p \times q$, respectively, in order to satisfy $\mathbf{MM}^{-1} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times q} \\ \mathbf{0}_{q\times m} & \mathbf{I}_q \end{bmatrix}$ and

$\mathbf{M}^{-1}\mathbf{M} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n\times p} \\ \mathbf{0}_{p\times n} & \mathbf{I}_p \end{bmatrix}$. So, we can conclude that the blocks of $\mathbf{M}^{-1}$ have the same dimensions as those of $\mathbf{M}^T$, and consequently, the partition of $\mathbf{M}^{-1}$ is transposed of that of $\mathbf{M}$.

We shall first consider special cases corresponding to non-singular structured $2 \times 2$ block matrices, with square diagonal blocks, then non-structured forms. In section 5.15, we shall consider general matrices partitioned into (2,2) blocks, that is, with singular or rectangular submatrices.

### 5.14.1. *Inversion of block-diagonal matrices*

Assuming that $\mathbf{A}$ and $\mathbf{D}$ are non-singular, we have:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \quad \Rightarrow \quad \mathbf{M}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix}. \qquad [5.31]$$

### 5.14.2. *Inversion of block-triangular matrices*

For block upper and lower triangular matrices, with non-singular square diagonal blocks, we have:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} \qquad [5.32]$$

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{D}^{-1} \end{bmatrix}. \qquad [5.33]$$

When $\mathbf{A} = \mathbf{I}_n$ and $\mathbf{D} = \mathbf{I}_m$, that is for unit block-triangular matrices, formulae [5.32] and [5.33] become:

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{B} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}_m & -\mathbf{B} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix} \qquad [5.34]$$

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_n \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ -\mathbf{C} & \mathbf{I}_n \end{bmatrix}. \qquad [5.35]$$

Similarly, with non-singular square off-diagonal blocks, we have:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{C}^{-1} \\ \mathbf{B}^{-1} & -\mathbf{B}^{-1}\mathbf{A}\mathbf{C}^{-1} \end{bmatrix} \qquad [5.36]$$

$$\begin{bmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} -\mathbf{C}^{-1}\mathbf{D}\mathbf{B}^{-1} & \mathbf{C}^{-1} \\ \mathbf{B}^{-1} & \mathbf{0} \end{bmatrix}, \qquad [5.37]$$

with the following particular cases:

$$\begin{bmatrix} \mathbf{A} & \mathbf{I}_m \\ \mathbf{I}_n & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_n \\ \mathbf{I}_m & -\mathbf{A} \end{bmatrix}, \begin{bmatrix} \mathbf{0} & \mathbf{I}_m \\ \mathbf{I}_n & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} -\mathbf{D} & \mathbf{I}_n \\ \mathbf{I}_m & \mathbf{0} \end{bmatrix}. \quad [5.38]$$

When the partitioned matrix has no special structure, its inverse and its determinant are determined from block-triangular factorization, this factorization being itself obtained from block-diagonalization.

### 5.14.3. *Block-triangularization and Schur complements*

Assuming that $\mathbf{D}$ is non-singular, and applying the elementary transformation [5.30a], the partitioned matrix $\mathbf{M}$, defined in [5.27], can be transformed into a lower block-triangular form. Indeed, by choosing $\mathbf{H} = -\mathbf{B}\mathbf{D}^{-1}$, we obtain:

$$\begin{bmatrix} \mathbf{I}_n & -\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \mathbf{M} = \begin{bmatrix} \mathbf{X_D} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \quad [5.39a]$$

$$\mathbf{X_D} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}, \quad [5.39b]$$

where $\mathbf{X_D}$, also denoted by $(\mathbf{M}/\mathbf{D})$, is called the Schur complement of $\mathbf{D}$ in $\mathbf{M}$.

Similarly, assuming that $\mathbf{A}$ is non-singular and choosing $\mathbf{H} = -\mathbf{A}^{-1}\mathbf{B}$, transformation [5.30d] yields:

$$\mathbf{M} \begin{bmatrix} \mathbf{I}_n & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{X_A} \end{bmatrix} \quad [5.40a]$$

$$\mathbf{X_A} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}, \quad [5.40b]$$

where $\mathbf{X_A}$, also denoted by $(\mathbf{M}/\mathbf{A})$, is the Schur complement of $\mathbf{A}$ in $\mathbf{M}$.

Similarly, elementary transformations [5.30c] and [5.30b] can be used to transform the partitioned matrix $\mathbf{M}$ into an upper block triangular form:

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_m \end{bmatrix} \mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{X_A} \end{bmatrix} \quad [5.41a]$$

$$\mathbf{M} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{X_D} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}, \quad [5.41b]$$

with $\mathbf{X_A}$ and $\mathbf{X_D}$ defined in [5.40b] and [5.39b].

### 5.14.4. *Block-diagonalization and block-factorization*

Assuming that $\mathbf{A}$ is invertible and combining block-triangularization operations [5.40a] and [5.41a], it is possible to put $\mathbf{M}$ in a block-diagonal form (Zhang 1999):

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{X_A} \end{bmatrix}. \quad [5.42]$$

Using the inversion formulae [5.34] and [5.35], the partitioned matrix $\mathbf{M}$ can be written in the following block-factorized form:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_\mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix}. \quad [5.43]$$

Similarly, assuming that $\mathbf{D}$ is invertible and combining [5.39a] and [5.41b], a second block-diagonal form is obtained:

$$\begin{bmatrix} \mathbf{I}_n & -\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{X}_\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}, \quad [5.44]$$

from which the following block-factorized form is deduced:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{X}_\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m \end{bmatrix}. \quad [5.45]$$

In the following, it is assumed that $\mathbf{X}_\mathbf{A}$ and $\mathbf{X}_\mathbf{D}$ are invertible.

### 5.14.5. *Block-inversion and partitioned inverse*

In this section, we present inversion formulae for the partitioned matrix [5.27] in terms of $\mathbf{A}^{-1}$ and $\mathbf{D}^{-1}$.

Using [5.34] and [5.35], the block-factorized form [5.43] gives the so-called Banachiewicz–Schur form, with $\mathbf{X}_\mathbf{A} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$:

$$\begin{aligned} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} &= \begin{bmatrix} \mathbf{I}_n & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_m \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{X}_\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{X}_\mathbf{A}^{-1} \\ -\mathbf{X}_\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{X}_\mathbf{A}^{-1} \end{bmatrix}. \end{aligned} \quad [5.46]$$

This inversion formula is valid if $\mathbf{A}$ and $\mathbf{X}_\mathbf{A}$ are invertible.

Similarly, the factorized form [5.45] leads to:

$$\begin{aligned} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} &= \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{X}_\mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & -\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{X}_\mathbf{D}^{-1} & -\mathbf{X}_\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{X}_\mathbf{D}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{X}_\mathbf{D}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}. \end{aligned} \quad [5.47]$$

This inversion formula is valid if $\mathbf{D}$ and $\mathbf{X}_\mathbf{D} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ are invertible.

If $\mathbf{A}$ and $\mathbf{D}$ are both non-singular, formulae [5.46] and [5.47] are equivalent and can be combined as shown in the next section.

### 5.14.6. *Other formulae for the partitioned* $2 \times 2$ *inverse*

From [5.46] and [5.47], the following other forms of $\mathbf{M}^{-1}$ can be deduced:

$$\begin{aligned}\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} &= \begin{bmatrix}\mathbf{A}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} + \begin{bmatrix}-\mathbf{A}^{-1}\mathbf{B}\\ \mathbf{I}_m\end{bmatrix} \mathbf{X}_{\mathbf{A}}^{-1} \begin{bmatrix}-\mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_m\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}^{-1}\end{bmatrix} + \begin{bmatrix}\mathbf{I}_n\\ -\mathbf{D}^{-1}\mathbf{C}\end{bmatrix} \mathbf{X}_{\mathbf{D}}^{-1} \begin{bmatrix}\mathbf{I}_n & -\mathbf{B}\mathbf{D}^{-1}\end{bmatrix}\end{aligned}$$

Combining formulae [5.46] and [5.47], we can also rewrite $\mathbf{M}^{-1}$ as:

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{X}_{\mathbf{D}}^{-1} & -\mathbf{X}_{\mathbf{D}}^{-1}\mathbf{B}\mathbf{D}^{-1}\\ -\mathbf{X}_{\mathbf{A}}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{X}_{\mathbf{A}}^{-1}\end{bmatrix} \tag{5.48a}$$

$$= \begin{bmatrix}\mathbf{X}_{\mathbf{D}}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{X}_{\mathbf{A}}^{-1}\\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{X}_{\mathbf{D}}^{-1} & \mathbf{X}_{\mathbf{A}}^{-1}\end{bmatrix}. \tag{5.48b}$$

which gives the following block-factorizations:

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{X}_{\mathbf{D}}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_{\mathbf{A}}^{-1}\end{bmatrix} \begin{bmatrix}\mathbf{I}_n & -\mathbf{B}\mathbf{D}^{-1}\\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_m\end{bmatrix} \tag{5.49a}$$

$$= \begin{bmatrix}\mathbf{I}_n & -\mathbf{A}^{-1}\mathbf{B}\\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m\end{bmatrix} \begin{bmatrix}\mathbf{X}_{\mathbf{D}}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_{\mathbf{A}}^{-1}\end{bmatrix}. \tag{5.49b}$$

By taking the first row-block of [5.46] and the second row-block of [5.47], we get:

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{X}_{\mathbf{A}}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{X}_{\mathbf{A}}^{-1}\\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{X}_{\mathbf{D}}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{X}_{\mathbf{D}}^{-1}\mathbf{B}\mathbf{D}^{-1}\end{bmatrix}. \tag{5.50}$$

PROOF.– (Alternative demonstration for the inversion formula [5.48b])

Defining the inverse of $\mathbf{M}$ as the partitioned matrix $\mathbf{M}^{-1} = \begin{bmatrix}\mathbf{E} & \mathbf{F}\\ \mathbf{G} & \mathbf{H}\end{bmatrix}$ and using the definition of the inverse, namely:

$$\mathbf{M}\mathbf{M}^{-1} = \begin{bmatrix}\mathbf{A}\mathbf{E} + \mathbf{B}\mathbf{G} & \mathbf{A}\mathbf{F} + \mathbf{B}\mathbf{H}\\ \mathbf{C}\mathbf{E} + \mathbf{D}\mathbf{G} & \mathbf{C}\mathbf{F} + \mathbf{D}\mathbf{H}\end{bmatrix} = \begin{bmatrix}\mathbf{I}_n & \mathbf{0}_{n\times m}\\ \mathbf{0}_{m\times n} & \mathbf{I}_m\end{bmatrix}. \tag{5.51}$$

one deduces the equations:

$$\mathbf{A}\mathbf{E} + \mathbf{B}\mathbf{G} = \mathbf{I}_n \tag{5.52a}$$

$$\mathbf{A}\mathbf{F} + \mathbf{B}\mathbf{H} = \mathbf{0}_{n\times m} \tag{5.52b}$$

$$\mathbf{C}\mathbf{E} + \mathbf{D}\mathbf{G} = \mathbf{0}_{m\times n} \tag{5.52c}$$

$$\mathbf{C}\mathbf{F} + \mathbf{D}\mathbf{H} = \mathbf{I}_m. \tag{5.52d}$$

From equations [5.52b] and [5.52c], we obtain:

$$\mathbf{F} = -\mathbf{A}^{-1}\mathbf{B}\mathbf{H} \qquad [5.53a]$$

$$\mathbf{G} = -\mathbf{D}^{-1}\mathbf{C}\mathbf{E} \qquad [5.53b]$$

Replacing $\mathbf{G}$ and $\mathbf{F}$ by their above expressions in [5.52a] and [5.52d], and using definitions [5.40b] and [5.39b] of $\mathbf{X_A}$ and $\mathbf{X_D}$, give:

$$\mathbf{E} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{X}_\mathbf{D}^{-1} \qquad [5.54a]$$

$$\mathbf{H} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} = \mathbf{X}_\mathbf{A}^{-1}, \qquad [5.54b]$$

then, by replacing $\mathbf{E}$ and $\mathbf{H}$ by their above expressions in [5.53a] and [5.53b], we get:

$$\mathbf{F} = -\mathbf{A}^{-1}\mathbf{B}\mathbf{X}_\mathbf{A}^{-1} \qquad [5.55a]$$

$$\mathbf{G} = -\mathbf{D}^{-1}\mathbf{C}\mathbf{X}_\mathbf{D}^{-1}. \qquad [5.55b]$$

Equations [5.54a]–[5.55b] define $\mathbf{M}^{-1}$ identically to formula [5.48b]. □

NOTE 5.16.– When the off-diagonal blocks $\mathbf{B}$ and $\mathbf{C}$ are non-singular square matrices and not the diagonal blocks, it is possible to derive inverse formulae similar to [5.46] and [5.47] using permutations of row- and column-blocks which transform the original partitioned matrix into a $2 \times 2$ block matrix with non-singular square diagonal blocks. Such inverse formulae are given by Lu and Shiou (2002).

### 5.14.7. *Solution of a system of linear equations*

Consider the following system of linear equations:

$$\begin{aligned}
\mathbf{A}\mathbf{x}_1 + \mathbf{B}\mathbf{x}_2 &= \mathbf{y}_1 \\
\mathbf{C}\mathbf{x}_1 + \mathbf{D}\mathbf{x}_2 &= \mathbf{y}_2 \\
&\Updownarrow \\
\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}.
\end{aligned}$$

Applying the inversion formula [5.48a] gives us the following solution:

$$\mathbf{x}_1 = \mathbf{X}_\mathbf{D}^{-1}(\mathbf{y}_1 - \mathbf{B}\mathbf{D}^{-1}\mathbf{y}_2)$$

$$\mathbf{x}_2 = \mathbf{X}_\mathbf{A}^{-1}(\mathbf{y}_2 - \mathbf{C}\mathbf{A}^{-1}\mathbf{y}_1),$$

where $\mathbf{X_A}$ and $\mathbf{X_D}$ are the Schur complements defined in [5.40b] and [5.39b].

### 5.14.8. *Inversion of a partitioned Gram matrix*

Given a matrix partitioned into two column blocks $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2]$, with $\mathbf{A}_1 \in \mathbb{R}^{n\times m}$ and $\mathbf{A}_2 \in \mathbb{R}^{n\times p}$, its Gram matrix $\mathbf{A}^T\mathbf{A}$ is partitioned as :

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \mathbf{A}_1^T\mathbf{A}_1 & \mathbf{A}_1^T\mathbf{A}_2 \\ \mathbf{A}_2^T\mathbf{A}_1 & \mathbf{A}_2^T\mathbf{A}_2 \end{bmatrix}.$$

Inversion of this type of matrix is employed for solving linear prediction problems or for estimating linear regression models. The application of inversion formula [5.48a], with $\mathbf{A} = \mathbf{A}_1^T\mathbf{A}_1$, $\mathbf{B} = \mathbf{A}_1^T\mathbf{A}_2$, $\mathbf{C} = \mathbf{A}_2^T\mathbf{A}_1$, $\mathbf{D} = \mathbf{A}_2^T\mathbf{A}_2$, gives:

$$[\mathbf{A}^T\mathbf{A}]^{-1} = \begin{bmatrix} (\mathbf{A}_1^T\mathbf{P}_2^\perp\mathbf{A}_1)^{-1} & -(\mathbf{A}_1^T\mathbf{P}_2^\perp\mathbf{A}_1)^{-1}\mathbf{A}_1^T\mathbf{A}_2(\mathbf{A}_2^T\mathbf{A}_2)^{-1} \\ -(\mathbf{A}_2^T\mathbf{P}_1^\perp\mathbf{A}_2)^{-1}\mathbf{A}_2^T\mathbf{A}_1(\mathbf{A}_1^T\mathbf{A}_1)^{-1} & (\mathbf{A}_2^T\mathbf{P}_1^\perp\mathbf{A}_2)^{-1} \end{bmatrix},$$

where $\mathbf{P}_1^\perp$ and $\mathbf{P}_2^\perp$ are the orthogonal complements of orthogonal projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ on column spaces $C(\mathbf{A}_1)$ and $C(\mathbf{A}_2)$, respectively, that is:

$$\mathbf{P}_1^\perp = \mathbf{I}_n - \mathbf{P}_1 = \mathbf{I}_n - \mathbf{A}_1(\mathbf{A}_1^T\mathbf{A}_1)^{-1}\mathbf{A}_1^T$$
$$\mathbf{P}_2^\perp = \mathbf{I}_n - \mathbf{P}_2 = \mathbf{I}_n - \mathbf{A}_2(\mathbf{A}_2^T\mathbf{A}_2)^{-1}\mathbf{A}_2^T.$$

When $\mathbf{A}_1$ and $\mathbf{A}_2$ are complex, the transposition operator is to be replaced by that of conjugate transposition in the above equations.

### 5.14.9. *Iterative inversion of a partitioned square matrix*

Consider the square matrix $\mathbf{M}_n$ of order $n$, partitioned into the following form:

$$\mathbf{M}_n = \begin{bmatrix} \mathbf{M}_{n-1} & \mathbf{c}_n \\ \mathbf{r}_n^T & \sigma_n \end{bmatrix} \qquad [5.56]$$

where $\mathbf{M}_{n-1}$ is a square matrix of order $n-1$, and $\mathbf{c}_n, \mathbf{r}_n \in \mathbb{K}^{n-1}$.

Assuming $\mathbf{M}_{n-1}$ and $\mathbf{M}_n$ are invertible, the application of the inversion formula [5.46] allows to perform the calculation of the inverse $\mathbf{M}_n^{-1}$ recursively with respect to order $n$, that is, in terms of $\mathbf{M}_{n-1}^{-1}$:

$$\mathbf{M}_n^{-1} = \begin{bmatrix} \mathbf{M}_{n-1}^{-1} + k_n\mathbf{M}_{n-1}^{-1}\mathbf{c}_n\mathbf{r}_n^T\mathbf{M}_{n-1}^{-1} & -k_n\mathbf{M}_{n-1}^{-1}\mathbf{c}_n \\ -k_n\mathbf{r}_n^T\mathbf{M}_{n-1}^{-1} & k_n \end{bmatrix} \qquad [5.57]$$

where $k_n = (\sigma_n - \mathbf{r}_n^T\mathbf{M}_{n-1}^{-1}\mathbf{c}_n)^{-1}$.

This recursive inversion formula will be used in section 5.18 to demonstrate the Levinson–Durbin algorithm.

## 5.14.10. *Matrix inversion lemma and applications*

### 5.14.10.1. *Matrix inversion lemma*

PROPOSITION 5.17.– *Let* $\mathbf{A} \in \mathbb{K}^{n\times n}, \mathbf{B} \in \mathbb{K}^{n\times m}, \mathbf{C} \in \mathbb{K}^{m\times n}$, *and* $\mathbf{D} \in \mathbb{K}^{m\times m}$. *By identifying the blocks (1,1) of the right-hand sides of [5.46] and [5.47], it can be deduced that:*

$$\begin{aligned}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} &= \mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}\mathbf{X}_{\mathbf{A}}^{-1}\mathbf{C}\mathbf{A}^{-1}\\ &= \mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}\end{aligned} \quad [5.58]$$

*This formula is known as the matrix inversion lemma. It is also called the Sherman–Morrison–Woodbury formula.*

It should be noted that $\mathbf{X}_{\mathbf{D}} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ is defined if $\mathbf{D}$ is invertible, and its inverse can be calculated using [5.58] if $\mathbf{A}$ and $\mathbf{X}_{\mathbf{A}} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ are invertible.

### 5.14.10.2. *Applications of the matrix inversion lemma*

In the next proposition, we present several applications of the matrix inversion lemma.

PROPOSITION 5.18.– *For different choices of* $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$, *the inversion lemma provides the following identities:*

– *For* $\mathbf{D} = -\mathbf{I}_m$, *the matrix inversion lemma [5.58] gives:*

$$[\mathbf{A}+\mathbf{B}\mathbf{C}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{I}_m + \mathbf{C}\mathbf{A}^{-1}\mathbf{B}]^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad [5.59]$$

– *From this identity, it can be deduced that for* $\mathbf{A} = \mathbf{I}_n$ *and* $\mathbf{C} = \mathbf{B}^H$

$$[\mathbf{I}_n + \mathbf{B}\mathbf{B}^H]^{-1} = \mathbf{I}_n - \mathbf{B}[\mathbf{I}_m + \mathbf{B}^H\mathbf{B}]^{-1}\mathbf{B}^H. \quad [5.60]$$

– *For* $m = n$, $\mathbf{D} = -\boldsymbol{\Delta}^{-1}$ *and* $\mathbf{B} = \mathbf{C} = \mathbf{I}_n$

$$[\mathbf{A}+\boldsymbol{\Delta}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}[\mathbf{A}^{-1}+\boldsymbol{\Delta}^{-1}]^{-1}\mathbf{A}^{-1}.$$

– *For* $m = 1$, $\mathbf{D} = -1/\alpha$, $\mathbf{B} = \mathbf{u} \in \mathbb{K}^n$, *and* $\mathbf{C}^T = \mathbf{v} \in \mathbb{K}^n$, *assuming that* $\alpha^{-1} + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u} \neq 0$, *we get:*

$$[\mathbf{A}+\alpha\mathbf{u}\mathbf{v}^T]^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{\alpha^{-1}+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}. \quad [5.61]$$

In Table 5.1, we summarize key results related to the inversion of partitioned matrices and the matrix inversion lemma.

**Inversion of $2 \times 2$ block matrices, with invertible diagonal blocks**

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{X}_{\mathbf{A}}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{X}_{\mathbf{A}}^{-1} \\ -\mathbf{X}_{\mathbf{A}}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{X}_{\mathbf{A}}^{-1} \end{bmatrix}$$

$$\mathbf{X}_{\mathbf{A}} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$$

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{X}_{\mathbf{D}}^{-1} & -\mathbf{X}_{\mathbf{D}}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{X}_{\mathbf{D}}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{X}_{\mathbf{D}}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}$$

$$\mathbf{X}_{\mathbf{D}} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$$

**Recursive inversion with respect to order**

$$\begin{bmatrix} \mathbf{M}_{n-1} & \mathbf{c}_n \\ \mathbf{r}_n^T & \sigma_n \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{M}_{n-1}^{-1} + k_n \mathbf{M}_{n-1}^{-1}\mathbf{c}_n\mathbf{r}_n^T\mathbf{M}_{n-1}^{-1} & -k_n\mathbf{M}_{n-1}^{-1}\mathbf{c}_n \\ -k_n\mathbf{r}_n^T\mathbf{M}_{n-1}^{-1} & k_n \end{bmatrix}$$

$$k_n = (\sigma_n - \mathbf{r}_n^T\mathbf{M}_{n-1}^{-1}\mathbf{c}_n)^{-1}$$

**Matrix inversion lemma**

$$(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{X}_{\mathbf{A}}^{-1}\mathbf{C}\mathbf{A}^{-1}$$

$$\mathbf{X}_{\mathbf{A}} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$$

**Special case**

$$[\mathbf{A} + \alpha\mathbf{u}\mathbf{v}^T]^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{\alpha^{-1} + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}$$

**Table 5.1.** *Inversion formulae*

## 5.15. Generalized inverses of $2 \times 2$ block matrices

In this section, we consider the extension of the Banachiewicz–Schur form [5.46] to the case of singular or rectangular matrices partitioned into $2 \times 2$ blocks, more specifically with singular or rectangular submatrices, written as:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \in \mathbb{K}^{(m+q)\times(n+p)}, \qquad [5.62]$$

with $\mathbf{A} \in \mathbb{K}^{m\times n}, \mathbf{B} \in \mathbb{K}^{m\times p}, \mathbf{C} \in \mathbb{K}^{q\times n}$, and $\mathbf{D} \in \mathbb{K}^{q\times p}$. Unlike the case of a partitioned matrix $\mathbf{M}$ with square diagonal blocks, previously addressed, we now consider rectangular diagonal blocks, needing to define a generalized inverse for $\mathbf{M}$.

The notion of generalized inverse was introduced by Moore (1935)[2], in a book published after his death. It was Penrose (1955) who demonstrated the uniqueness of the Moore generalized inverse, which explains the name given to the Moore–Penrose pseudo-inverse (see section 4.11.9). This pseudo-inverse that generalizes the inverse of a regular square matrix to the case of rectangular matrices plays a very important role for solving systems of linear equations using the method of least squares.

In the following, we define different types of generalized inverse $\mathbf{A}^\sharp \in \mathbb{K}^{n\times m}$ of a matrix $\mathbf{A} \in \mathbb{K}^{m\times n}$ according to the equations that are satisfied among:

$$(1) \quad \mathbf{A}\mathbf{A}^\sharp\mathbf{A} = \mathbf{A} \qquad [5.63a]$$

$$(2) \quad \mathbf{A}^\sharp\mathbf{A}\mathbf{A}^\sharp = \mathbf{A}^\sharp \qquad [5.63b]$$

$$(3) \quad (\mathbf{A}\mathbf{A}^\sharp)^H = \mathbf{A}\mathbf{A}^\sharp \qquad [5.63c]$$

$$(4) \quad (\mathbf{A}^\sharp\mathbf{A})^H = \mathbf{A}^\sharp\mathbf{A}. \qquad [5.63d]$$

Note that (3) and (4) means that $\mathbf{A}\mathbf{A}^\sharp$ and $\mathbf{A}^\sharp\mathbf{A}$ are Hermitian, respectively, or symmetric in the real case.

Any inverse only satisfying conditions $\{c_1\}$, or $\{c_1, c_2\}$, or $\{c_1, c_2, c_3\}$, with $c_1, c_2, c_3 \in \{(1), (2), (3), (4)\}$, is denoted $\mathbf{A}^{\{c_1\}}$, $\mathbf{A}^{\{c_1,c_2\}}$, and $\mathbf{A}^{\{c_1,c_2,c_3\}}$, respectively (Burns *et al.* 1974). The properties of this type of inverse were studied by Ben-Israel and Greville (2001). In the literature, the inverses $\mathbf{A}^{\{1\}}$, $\mathbf{A}^{\{2\}}$, $\mathbf{A}^{\{1,2\}}$, and $\mathbf{A}^{\{1,2,3\}}$ are often called inner inverse, outer inverse, reflexive generalized inverse (or semi-inverse), and weak generalized inverse (or least-squares reflexive generalized inverse), respectively.

For any matrix $\mathbf{A}$, there exists a unique matrix $\mathbf{A}^{\{1,2,3,4\}}$, namely, satisfying the four equations [5.63a]–[5.63d]. This matrix corresponds to the Moore–Penrose pseudo-inverse of $\mathbf{A}$ and is often denoted by $\mathbf{A}^\dagger$ (Penrose 1955).

When $\mathbf{M}$ and $\mathbf{A}$ in [5.62] are singular, the Banachiewicz–Schur formula [5.46] can be extended by replacing the inverses of $\mathbf{A}$ and $\mathbf{X}_\mathbf{A}$ by generalized inverses $\mathbf{A}^\sharp$

2 Eliakim Hastings Moore (1862–1932), American mathematician whose research topics have focused on group theory, algebraic geometry, and integral equations. He is particularly known for the pseudo-inverse matrix that bears his name.

and $\mathbf{X}_{\mathbf{A}}^{\sharp}$, which gives:

$$\mathbf{M}(\mathbf{A}^{\sharp}, \mathbf{X}_{\mathbf{A}}^{\sharp}) = \begin{bmatrix} \mathbf{I}_n & -\mathbf{A}^{\sharp}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{A}^{\sharp} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{\mathbf{A}}^{\sharp} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{C}\mathbf{A}^{\sharp} & \mathbf{I}_m \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{A}^{\sharp} + \mathbf{A}^{\sharp}\mathbf{B}\mathbf{X}_{\mathbf{A}}^{\sharp}\mathbf{C}\mathbf{A}^{\sharp} & -\mathbf{A}^{\sharp}\mathbf{B}\mathbf{X}_{\mathbf{A}}^{\sharp} \\ -\mathbf{X}_{\mathbf{A}}^{\sharp}\mathbf{C}\mathbf{A}^{\sharp} & \mathbf{X}_{\mathbf{A}}^{\sharp} \end{bmatrix}, \quad [5.64]$$

where $\mathbf{X}_{\mathbf{A}}$ is now the generalized Schur complement of $\mathbf{A}$ in $\mathbf{M}$, defined as:

$$\mathbf{X}_{\mathbf{A}} = \mathbf{D} - \mathbf{C}\mathbf{A}^{\sharp}\mathbf{B} \in \mathbb{K}^{q \times p}. \quad [5.65]$$

It is important to note that using a generalized inverse $\mathbf{A}^{\sharp}$ of type $\{c_1\}$, or $\{c_1, c_2\}$, or $\{c_1, c_2, c_3\}$ does not necessarily imply the same kind of generalized inverse for $\mathbf{M}(\mathbf{A}^{\sharp}, \mathbf{X}_{\mathbf{A}}^{\sharp})$, that is $\mathbf{M}^{\sharp}$. However, it is possible to establish rank conditions for this property to be satisfied. Thus, in Marsaglia and Styan (1974) and Baksalary and Styan (2002), necessary and sufficient conditions are provided such that the use of the Moore–Penrose inverse $\mathbf{A}^{\dagger}$ and $\mathbf{X}_{\mathbf{A}}^{\dagger}$ in [5.64] induces the Moore–Penrose inverse $\mathbf{M}^{\dagger}$ of $\mathbf{M}$:

$$\mathbf{M}^{\dagger} = \mathbf{M}(\mathbf{A}^{\dagger}, \mathbf{X}_{\mathbf{A}}^{\dagger}) \;\Leftrightarrow\; \mathbf{F}_{\mathbf{A}}\mathbf{B} = \mathbf{0}\,,\; \mathbf{C}\mathbf{E}_{\mathbf{A}} = \mathbf{0}\,,\; \mathbf{B}\mathbf{E}_{\mathbf{X}_{\mathbf{A}}} = \mathbf{0}\,,\; \mathbf{F}_{\mathbf{X}_{\mathbf{A}}}\mathbf{C} = \mathbf{0},$$

where

$$\mathbf{X}_{\mathbf{A}} = \mathbf{D} - \mathbf{C}\mathbf{A}^{\dagger}\mathbf{B} \quad [5.66a]$$

$$\mathbf{E}_{\mathbf{A}} = \mathbf{I}_n - \mathbf{A}^{\dagger}\mathbf{A} \quad , \quad \mathbf{F}_{\mathbf{A}} = \mathbf{I}_m - \mathbf{A}\mathbf{A}^{\dagger} \quad [5.66b]$$

$$\mathbf{E}_{\mathbf{X}_{\mathbf{A}}} = \mathbf{I}_p - \mathbf{X}_{\mathbf{A}}^{\dagger}\mathbf{X}_{\mathbf{A}} \;\;,\;\; \mathbf{F}_{\mathbf{X}_{\mathbf{A}}} = \mathbf{I}_q - \mathbf{X}_{\mathbf{A}}\mathbf{X}_{\mathbf{A}}^{\dagger}. \quad [5.66c]$$

Further results highlighting the type of generalized inverse of $\mathbf{M}$ which results from $\mathbf{M}(\mathbf{A}^{\sharp}, \mathbf{X}_{\mathbf{A}}^{\sharp})$ can be found in Baksalary and Styan (2002), Tian and Takane (2005, 2009), and Liu *et al.* (2016).

## 5.16. Determinants of partitioned matrices

### 5.16.1. *Determinant of block-diagonal matrices*

Assuming that $\mathbf{A}$ and $\mathbf{D}$ are square matrices, we have:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \quad \Rightarrow \quad \det(\mathbf{M}) = \det(\mathbf{A})\det(\mathbf{D}).$$

### 5.16.2. *Determinant of block-triangular matrices*

For upper and lower block-triangular matrices, with square diagonal blocks, we have:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \quad \Rightarrow \quad \det(\mathbf{M}) = \det(\mathbf{A})\det(\mathbf{D}). \tag{5.67a}$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \quad \Rightarrow \quad \det(\mathbf{M}) = \det(\mathbf{A})\det(\mathbf{D}). \tag{5.67b}$$

For unit block-triangular matrices ($\mathbf{A} = \mathbf{I}_n, \mathbf{D} = \mathbf{I}_m$), we have $\det(\mathbf{M}) = 1$.

More generally, we have:

$$\det \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1R} \\ & \mathbf{A}_{22} & & \vdots \\ & \mathbf{0} & \ddots & \vdots \\ & & & \mathbf{A}_{RR} \end{bmatrix} = \prod_{r=1}^{R} \det(\mathbf{A}_{rr}), \tag{5.68}$$

which means that the determinant of an upper or a lower block-triangular matrix, composed of $R$ square diagonal blocks $\mathbf{A}_{rr} \in \mathbb{K}^{n_r \times n_r}$, with $\sum_{r=1}^{R} n_r = n$, is equal to the product of the determinants of the $R$ diagonal blocks.

EXAMPLE 5.19.–

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \vdots & a_{13} \\ \cdots & & \cdots \\ \mathbf{0}_{1\times 2} & \vdots & a_{33} \end{bmatrix}$$

$$\Downarrow$$

$$\det(\mathbf{A}) = a_{33}\det(\mathbf{A}_{11}) = a_{33}\left(a_{11}a_{22} - a_{12}a_{21}\right)$$

which corresponds to the Laplace expansion [4.13] with respect to the third row of $\mathbf{A}$.

### 5.16.3. *Determinant of partitioned matrices with square diagonal blocks*

From relations [5.42] and [5.44], the following expressions can be deduced:

$$\det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \det(\mathbf{A})\det(\mathbf{X}_\mathbf{A}) = \det(\mathbf{A})\det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}) \tag{5.69a}$$

$$= \det(\mathbf{D})\det(\mathbf{X}_\mathbf{D}) = \det(\mathbf{D})\det(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}). \tag{5.69b}$$

In Powell (2011), a formula can be found for the computation of the determinant of a matrix $\mathbf{A} = [\mathbf{A}_{rs}] \in \mathbb{K}^{nR \times nR}$ partitioned into $(R, R)$ blocks $\mathbf{A}_{rs} \in \mathbb{K}^{n \times n}$, with $r, s \in \langle R \rangle$. For $R = 3$, we have:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} \end{bmatrix}$$

$$\Downarrow$$

$$\begin{aligned} \det(\mathbf{A}) = \det\Big( & [\mathbf{A}_{11} - \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{A}_{31}] \\ & - [\mathbf{A}_{12} - \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{A}_{32}][\mathbf{A}_{22} - \mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{A}_{32}]^{-1}[\mathbf{A}_{21} - \mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{A}_{31}]\Big) \\ & \times \det(\mathbf{A}_{22} - \mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{A}_{32})\det(\mathbf{A}_{33}). \end{aligned} \quad [5.70]$$

By choosing $\mathbf{A}_{13} = \mathbf{A}_{31} = \mathbf{A}_{23} = \mathbf{A}_{32} = \mathbf{0}_n$ and $\mathbf{A}_{33} = \mathbf{I}_n$ in the above formula, we revisit formula [5.69b] of the determinant of a $2 \times 2$ block matrix, with $\mathbf{A} = \mathbf{A}_{11}, \mathbf{B} = \mathbf{A}_{12}, \mathbf{C} = \mathbf{A}_{21}, \mathbf{D} = \mathbf{A}_{22}$. On the other hand, for $R = 3$ and $n = 1$, we find again the Sarrus formula for the computation of the determinant of a $3 \times 3$ matrix, demonstrated in Example 4.28.

### 5.16.4. *Determinants of specific partitioned matrices*

PROPOSITION 5.20.– *When* $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ *are square matrices of order* $n$*, we have:*

– *If* $\mathbf{A}$ *and* $\mathbf{C}$ *commute:*

$$det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = det(\mathbf{AD} - \mathbf{CB}). \quad [5.71]$$

– *If* $\mathbf{B}$ *and* $\mathbf{D}$ *commute:*

$$det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = det(\mathbf{DA} - \mathbf{BC}). \quad [5.72]$$

– *If* $\mathbf{A}$ *and* $\mathbf{B}$ *commute:*

$$det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = det(\mathbf{DA} - \mathbf{CB}). \quad [5.73]$$

– *If* $\mathbf{C}$ *and* $\mathbf{D}$ *commute:*

$$det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = det(\mathbf{AD} - \mathbf{BC}). \quad [5.74]$$

PROOF.– According to formula [5.69a] and taking into account the property $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ and the assumption of commutativity $\mathbf{AC} = \mathbf{CA}$, it follows that

$$\det\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = \det(\mathbf{A})\det(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}) = \det(\mathbf{A}\mathbf{D}-\mathbf{A}\mathbf{C}\mathbf{A}^{-1}\mathbf{B})$$
$$= \det(\mathbf{A}\mathbf{D}-\mathbf{C}\mathbf{A}\mathbf{A}^{-1}\mathbf{B}) = \det(\mathbf{A}\mathbf{D}-\mathbf{C}\mathbf{B}), \qquad [5.75]$$

which demonstrates [5.71]. The other formulae [5.72]–[5.74] can be demonstrated in the same way. □

PROPOSITION 5.21.– *The application of the previous formulae yields:*

$$det\begin{bmatrix}\mathbf{I}_n & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = det(\mathbf{D}-\mathbf{C}\mathbf{B}) \quad ; \quad det\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{I}_n\end{bmatrix} = det(\mathbf{A}-\mathbf{B}\mathbf{C}).$$
$$det\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{A} & \mathbf{D}\end{bmatrix} = det(\mathbf{A})det(\mathbf{D}-\mathbf{B}) \; ; \; det\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{B}\end{bmatrix} = det(\mathbf{B})det(\mathbf{A}-\mathbf{C}).$$
$$det\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{0}\end{bmatrix} = det\begin{bmatrix}\mathbf{0} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = (-1)^n det(\mathbf{B})det(\mathbf{C}).$$

Defining:

$$\mathbf{B} = \begin{bmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{bmatrix} \text{ with } \mathbf{B}_{11} \in \mathbb{K}^{n\times n}, \mathbf{B}_{22} \in \mathbb{K}^{m\times m} \; \mathbf{A} \in \mathbb{K}^{n\times n}, \mathbf{C} \in \mathbb{K}^{m\times m},$$

and using the property $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$, as well as determinant formulae for block diagonal matrices and block-triangular matrices, the other following determinants can be deduced:

$$\det\begin{bmatrix}\mathbf{A}\mathbf{B}_{11} & \mathbf{A}\mathbf{B}_{12}\\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{bmatrix} = \det\left(\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix}\mathbf{B}\right) = \det(\mathbf{A})\det(\mathbf{B}),$$

and

$$\det\begin{bmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\\ \mathbf{B}_{21}+\mathbf{C}\mathbf{B}_{11} & \mathbf{B}_{22}+\mathbf{C}\mathbf{B}_{12}\end{bmatrix} = \det\left(\begin{bmatrix}\mathbf{I}_n & \mathbf{0}\\ \mathbf{C} & \mathbf{I}_m\end{bmatrix}\mathbf{B}\right) = \det(\mathbf{B}).$$

### 5.16.5. *Eigenvalues of* CB *and* BC

The following proposition establishes the link between the eigenvalues of a product of two matrices and those of the product of the same matrices taken in reverse order.

PROPOSITION 5.22.– *For* $\mathbf{B} \in \mathbb{K}^{n\times m}$ *and* $\mathbf{C} \in \mathbb{K}^{m\times n}$*, we have:*

$$\lambda^n det(\lambda\mathbf{I}_m - \mathbf{CB}) = \lambda^m det(\lambda\mathbf{I}_n - \mathbf{BC}). \quad [5.76]$$

*This identity allows one to conclude that the eigenvalues of the products* $\mathbf{CB}$ *and* $\mathbf{BC}$ *are identical, with* $m - n$ *additional zero eigenvalues for* $\mathbf{CB}$ *if* $m > n$*, or* $n - m$ *additional zero eigenvalues for* $\mathbf{BC}$ *if* $n > m$*. For* $\lambda = 1$*, we obtain:*

$$det(\mathbf{I}_m - \mathbf{CB}) = det(\mathbf{I}_n - \mathbf{BC}). \quad [5.77]$$

PROOF.– The application of formulae [5.69a] and [5.69b], with $\mathbf{A} = \lambda\mathbf{I}_n$ and $\mathbf{D} = \mathbf{I}_m$, gives:

$$\begin{aligned} \det \begin{bmatrix} \lambda\mathbf{I}_n & \mathbf{B} \\ \mathbf{C} & \mathbf{I}_m \end{bmatrix} &= \det(\lambda\mathbf{I}_n)\det(\mathbf{I}_m - \lambda^{-1}\mathbf{CB}) = \lambda^{n-m}\det(\lambda\mathbf{I}_m - \mathbf{CB}) \\ &= \det(\lambda\mathbf{I}_n - \mathbf{BC}), \end{aligned} \quad [5.78]$$

from which the identity [5.76] is deduced. □

## 5.17. Rank of partitioned matrices

Let the partitioned matrix be:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}.$$

Obviously, we have:

$$\begin{aligned} \mathrm{r}(\mathbf{M}) &\leq \mathrm{r}\Big( \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \Big) + \mathrm{r}\Big( \begin{bmatrix} \mathbf{C} & \mathbf{D} \end{bmatrix} \Big) \\ &\leq \mathrm{r}\Big( \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} \Big) + \mathrm{r}\Big( \begin{bmatrix} \mathbf{B} \\ \mathbf{D} \end{bmatrix} \Big). \end{aligned}$$

In general, for a partitioned matrix $\mathbf{A} = [\mathbf{A}_{ij}]$, we have $\mathrm{r}(\mathbf{A}_{ij}) \leq \mathrm{r}(\mathbf{A})$.

Based on block-factorization formulae [5.42] and [5.44], the following relations can be deduced, for $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $\mathbf{D} \in \mathbb{C}^{m\times m}$:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

$$\Downarrow \quad [5.79a]$$

$$\mathrm{r}(\mathbf{M}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{X}_\mathbf{A}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}) \quad [5.79b]$$

$$= \mathrm{r}(\mathbf{D}) + \mathrm{r}(\mathbf{X}_\mathbf{D}) = \mathrm{r}(\mathbf{D}) + \mathrm{r}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}). \quad [5.79c]$$

In the case of rectangular submatrices $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{B} \in \mathbb{C}^{m\times k}$, $\mathbf{C} \in \mathbb{C}^{l\times n}$, and $\mathbf{D} \in \mathbb{C}^{l\times k}$, the rank of some partitioned matrices verifies the following equalities (Marsaglia and Styan 1974):

$$\mathrm{r}\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix} = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B} - \mathbf{A}\mathbf{A}^{\dagger}\mathbf{B}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{F}_{\mathbf{A}}\mathbf{B}) \quad [5.80a]$$

$$= \mathrm{r}(\mathbf{B}) + \mathrm{r}(\mathbf{A} - \mathbf{B}\mathbf{B}^{\dagger}\mathbf{A}) = \mathrm{r}(\mathbf{B}) + \mathrm{r}(\mathbf{F}_{\mathbf{B}}\mathbf{A}) \quad [5.80b]$$

$$\mathrm{r}\begin{bmatrix}\mathbf{A}\\ \mathbf{C}\end{bmatrix} = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{C} - \mathbf{C}\mathbf{A}^{\dagger}\mathbf{A}) = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{C}\mathbf{E}_{\mathbf{A}}) \quad [5.80c]$$

$$= \mathrm{r}(\mathbf{C}) + \mathrm{r}(\mathbf{A} - \mathbf{A}\mathbf{C}^{\dagger}\mathbf{C}) = \mathrm{r}(\mathbf{C}) + \mathrm{r}(\mathbf{A}\mathbf{E}_{\mathbf{C}}). \quad [5.80d]$$

Note that formulae [5.80b] and [5.80d] can be obtained from [5.80a] and [5.80c], respectively, by interchanging $\mathbf{A}$ and $\mathbf{B}$, on the one hand, and $\mathbf{A}$ and $\mathbf{C}$, on the other hand. We have the following simplified rank equalities:

$$\mathrm{r}\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix} = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) \Leftrightarrow \mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B}) = \{0\} \quad [5.81]$$

$$\mathrm{r}\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix} = \mathrm{r}(\mathbf{A}) \Leftrightarrow \mathcal{C}(\mathbf{B}) \subseteq \mathcal{C}(\mathbf{A}) \quad [5.82]$$

$$\mathrm{r}\begin{bmatrix}\mathbf{A}\\ \mathbf{C}\end{bmatrix} = \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{C}) \Leftrightarrow \mathcal{C}(\mathbf{A}^{H}) \cap \mathcal{C}(\mathbf{C}^{H}) = \{0\}, \quad [5.83]$$

where $\mathcal{C}(\mathbf{A})$ stands for the column space of $\mathbf{A}$.

$$\mathrm{r}\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{0}\end{bmatrix} = \mathrm{r}(\mathbf{B}) + \mathrm{r}(\mathbf{C}) + \mathrm{r}(\mathbf{F}_{\mathbf{B}}\mathbf{A}\mathbf{E}_{\mathbf{C}}), \quad [5.84a]$$

$$\mathrm{r}\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = \mathrm{r}(\mathbf{A}) + \mathrm{r}\begin{bmatrix}\mathbf{0} & \mathbf{F}_{\mathbf{A}}\mathbf{B}\\ \mathbf{C}\mathbf{E}_{\mathbf{A}} & \mathbf{D} - \mathbf{C}\mathbf{A}^{\dagger}\mathbf{B}\end{bmatrix}, \quad [5.84b]$$

where $(.)^{\dagger}$ denotes the pseudo-inverse, $\mathbf{E}_{\mathbf{A}}$ and $\mathbf{F}_{\mathbf{A}}$ are two projectors induced by $\mathbf{A}$, defined in [5.66b]. Different rank equalities for partitioned matrices and their Moore–Penrose inverses are provided in Tian (2004).

NOTE 5.23.– By choosing $\mathbf{B} = \mathbf{0}$ and thus $\mathbf{F}_{\mathbf{B}} = \mathbf{I}$ in [5.84a], we find again [5.80d]. Similarly, with $\mathbf{C} = \mathbf{0}$ and thus $\mathbf{E}_{\mathbf{C}} = \mathbf{I}$ in [5.84a], one obtains [5.80b].

Similarly, the choices ($\mathbf{C} = \mathbf{0}$ , $\mathbf{D} = \mathbf{0}$), on the one hand, and ($\mathbf{B} = \mathbf{0}$ , $\mathbf{D} = \mathbf{0}$), on the other hand, in [5.84b] lead to equalities [5.80a] and [5.80c], respectively.

## 5.18. Levinson–Durbin algorithm

Autoregressive (AR) processes or models are widely used in signal processing for the representation and classification of random signals. The autocorrelation function of such a process satisfies a system of linear equations in AR coefficients, called

Yule[3]–Walker equations. These equations form a Toeplitz[4] system whose inversion can be achieved in a numerically efficient way, by means of the Levinson[5] algorithm, which is an algorithm recursive with respect to the order of the model.

Besides the fact that the Levinson algorithm, also known as the Levinson–Durbin algorithm, plays a fundamental role in signal processing (Durbin 1960; Levinson 1947), the purpose of its presentation is to illustrate an application of the recursive formula [5.57], based on the Banachiewicz–Schur form [5.46], for solving the Yule–Walker equations. We shall also highlight the link existing between the estimation of AR parameters and that of the coefficients of one-step forward and backward linear predictors. These linear predictors will be interpreted in terms of orthogonal projectors over subspaces of the Hilbert space of second-order stationary random signals.

Before demonstrating the Levinson algorithm, we first establish the Yule–Walker equations.

### 5.18.1. *AR process and Yule–Walker equations*

A real process $x(k)$ is an AR process of order $N$, denoted AR($N$), if it satisfies the following input–output equation:

$$x(k) = -\sum_{n=1}^{N} a_{N,n} x(k-n) + e(k) = -\mathbf{a}_N^T \mathbf{x}_N(k) + e(k) \qquad [5.85]$$

3 George Udny Yule (1871–1951), Scottish statistician who concluded a year of study on electric waves, in Bonn, with the German physicist and mathematician Heinrich Hertz (1857–1894), before devoting his works to statistics, in London, from 1895. Inspired by Karl Pearson's (1857–1936) works, in 1911 he published a book entitled *Introduction to the Theory of Statistics*, which has been very successful, and whose 14th and final edition was written jointly with Maurice Kendall (1907–1983) and published in 1950. His contributions have focused more specifically on correlation and regression. He introduced the concepts of correlogram and autoregressive process, and he performed fundamental works on the theory of time series.

4 Otto Toeplitz (1881–1940), German mathematician who was a student of David Hilbert. He worked on integral equations and quadratic forms, and he developed a general theory of infinite-dimensional spaces. He gave his name to a class of matrices that play an important role in signal processing, as shown in this section.

5 Norman Levinson (1912–1975), American mathematician who worked with Norbert Wiener (1894–1964), early in his career, at the MIT. His contributions concern the theory of numbers, differential equations, partial derivative equations, integral equations, harmonic analysis, time series, filtering, and prediction.

where the regressor vector $\mathbf{x}_N(k)$ of dimension $N$ and the vector $\mathbf{a}_N$ of parameters of the AR($N$) model, called AR coefficients or regressor coefficients, are defined as:

$$\mathbf{a}_N^T = \left[a_{N,1}, \cdots, a_{N,N}\right] \quad [5.86a]$$

$$\mathbf{x}_N^T(k) = \left[x(k-1), \cdots, x(k-N)\right], \quad [5.86b]$$

and $e(k)$ is a white noise, of zero mean and variance $\sigma_e^2$, namely:

$$E[e(k+\tau)\, e(k)] = \sigma_e^2\, \delta_{\tau 0},$$

where $E[.]$ represents the mathematical expectation, and $\delta_{\tau 0}$ is the Kronecker delta, with $\delta_{\tau 0} = 1$ if $\tau = 0$, and $\delta_{\tau 0} = 0$ if $\tau \neq 0$. It is said that $x(k)$ is generated by filtering white noise through an all-pole filter (or AR filter), with the operatorial transfer function:

$$H(q^{-1}) = \frac{1}{1 + \sum_{n=1}^{N} a_{N,n} q^{-n}}$$

where $q^{-1}$ is the unit delay operator such that $q^{-1}\, x(k) = x(k-1)$ and $q^{-n}\, x(k) = x(k-n)$.

Equation [5.85] means that the signal $x(k)$ depends linearly on its previous $N$ samples and on the white noise $e(k)$. Observing that $x(k)$ depends on noise samples $e(t)$ for $t \leq k$, the white noise assumption implies $E[e(k+\tau)\, x(k)] = \sigma_e^2 \delta_{\tau 0},\ \forall \tau \geq 0$.

Assuming that the process $x(k)$ is second-order stationary[6], its autocorrelation function satisfies the following recurrent equation for $\tau > 0$:

$$\varphi_x(\tau) = E\left[x(k+\tau)x(k)\right] = -\sum_{n=1}^{N} a_{N,n} E\left[x(k+\tau-n)x(k)\right] + E\left[e(k+\tau)x(k)\right]$$

$$= -\sum_{n=1}^{N} a_{N,n} \varphi_x(\tau - n)\ ,\ \forall \tau > 0. \quad [5.87]$$

6 A random signal is strictly stationary (or stationary in the strict sense) if its statistics are independent of the origine of time or, equivalently, if they are invariant to any translation of the time origin. In signal processing, this assumption is often employed, together with the assumption of ergodicity, because it allows statistics (in the sense of the mathematical expectation) to be estimated based on time averages, in other words averages calculated using signals measured over a finite time interval. A random signal $x(k)$ is second-order stationary (also known as weak stationarity or stationarity in the broad sense), if its mean is constant, and its autocorrelation function $\varphi_x(k,t) = E[x(k)\, x(t)]$ depends only on the time interval $\tau = k - t$. Assuming second-order stationarity and ergodicity, the autocorrelation function can be estimated as:

$$\widehat{\varphi}_x(\tau) = \frac{1}{T} \sum_{k=1}^{T} x(k)x(k-\tau).$$

In addition, for $\tau = 0$, we have:

$$\varphi_x(0) = E\Big[x^2(k)\Big] = E\Big[\Big(-\sum_{n=1}^{N} a_{N,n}x(k-n) + e(k)\Big)x(k)\Big]$$

$$= -\sum_{n=1}^{N} a_{N,n}\varphi_x(n) + \sigma_e^2. \qquad [5.88]$$

Considering the autocorrelation function for $\tau \in \langle N \rangle$, and taking into account its symmetry property $\big(\varphi_x(\tau) = \varphi_x(-\tau)\big)$, we obtain the following system of equations:

$$\begin{bmatrix} \varphi_x(1) \\ \varphi_x(2) \\ \vdots \\ \varphi_x(N) \end{bmatrix} = - \begin{bmatrix} \varphi_x(0) & \varphi_x(1) & \cdots & \varphi_x(N-1) \\ \varphi_x(1) & \varphi_x(0) & \cdots & \varphi_x(N-2) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_x(N-1) & \varphi_x(N-2) & \cdots & \varphi_x(0) \end{bmatrix} \begin{bmatrix} a_{N,1} \\ a_{N,2} \\ \vdots \\ a_{N,N} \end{bmatrix}$$

or still in a compact form:

$$\mathbf{c}_N = \big[\varphi_x(1), \cdots, \varphi_x(N)\big]^T = -\mathbf{\Phi}_N \mathbf{a}_N, \qquad [5.89]$$

where $\mathbf{c}_N$ is the autocorrelation vector and $\mathbf{\Phi}_N$ is the autocorrelation matrix of order $N$ of the signal $x(k)$. Equation [5.88] can also be written as:

$$\varphi_x(0) = -\mathbf{c}_N^T \mathbf{a}_N + \sigma_e^2. \qquad [5.90]$$

Equations [5.89] and [5.90], which are linear with respect to the coefficients $\{a_{N,n}\}$ of the model AR$(N)$, are called the Yule–Walker equations, or normal equations. A method for estimating these AR parameters consists in estimating the autocorrelation function $\varphi_x(\tau)$ using time averages (hypothesis of ergodicity), and then in inverting the autocorrelation matrix $\mathbf{\Phi}_N$. Given that this matrix has a symmetric Toeplitz structure inducing its centrosymmetry property (see definition [5.97]), it is possible to invert it recursively with respect to order $N$, which corresponds to the Levinson–Durbin algorithm presented in the following section.

### 5.18.2. *Levinson–Durbin algorithm*

For the model AR$(N+1)$, the Yule–Walker equations [5.89] become:

$$\mathbf{c}_{N+1} = -\mathbf{\Phi}_{N+1}\mathbf{a}_{N+1} \qquad [5.91]$$

with the following partitionings:

$$\mathbf{\Phi}_{N+1} = \begin{bmatrix} \mathbf{\Phi}_N & \vdots & \widetilde{\mathbf{c}}_N \\ \cdots\cdots & & \cdots\cdots \\ \widetilde{\mathbf{c}}_N^T & \vdots & \varphi_x(0) \end{bmatrix}, \quad \mathbf{c}_{N+1} = \begin{bmatrix} \mathbf{c}_N \\ \cdots\cdots \\ \varphi_x(N+1) \end{bmatrix}, \qquad [5.92]$$

where $\widetilde{\mathbf{c}}_N$ is the vector $\mathbf{c}_N$ defined in [5.89], with its components in reverse order:

$$\widetilde{\mathbf{c}}_N^T = \left[\varphi_x(N), \cdots, \varphi_x(1)\right]. \qquad [5.93]$$

Let us define the matrix $\mathbf{J}_N$, with 1's on the antidiagonal and 0's elsewhere:

$$\mathbf{J}_N = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \iddots & \iddots & 0 \\ 0 & \iddots & \iddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}.$$

It satisfies the following properties:

$$\mathbf{J}_N^T = \mathbf{J}_N^{-1} = \mathbf{J}_N \quad ; \quad \mathbf{J}_N^2 = \mathbf{I}_N. \qquad [5.94]$$

The pre-multiplication of a vector by $\mathbf{J}_N$ inverting the order of its components, it follows that:

$$\widetilde{\mathbf{c}}_N = \mathbf{J}_N \mathbf{c}_N \;\Rightarrow\; \widetilde{\mathbf{c}}_N^T = \mathbf{c}_N^T \mathbf{J}_N \qquad [5.95a]$$

$$\widetilde{\mathbf{a}}_N = \mathbf{J}_N \mathbf{a}_N \qquad [5.95b]$$

where $\widetilde{\mathbf{a}}_N$ is the vector $\mathbf{a}_N$ with its components in reverse order.

By using properties [5.94] of $\mathbf{J}_N$ and definitions [5.95a] and [5.95b], it is easy to deduce the following identities:

$$\mathbf{c}_N^T \mathbf{a}_N = \mathbf{c}_N^T \mathbf{J}_N \mathbf{J}_N \mathbf{a}_N = \widetilde{\mathbf{c}}_N^T \widetilde{\mathbf{a}}_N \qquad [5.96a]$$

$$\mathbf{c}_N^T \widetilde{\mathbf{a}}_N = \mathbf{c}_N^T \mathbf{J}_N \mathbf{a}_N = \widetilde{\mathbf{c}}_N^T \mathbf{a}_N, \qquad [5.96b]$$

In addition, by making use of the centro-symmetry property of the autocorrelation matrix:

$$\mathbf{J}_N \mathbf{\Phi}_N \mathbf{J}_N = \mathbf{\Phi}_N \;\Leftrightarrow\; \mathbf{J}_N \mathbf{\Phi}_N = \mathbf{\Phi}_N \mathbf{J}_N \qquad [5.97]$$

and identities [5.95a], and [5.95b], the pre-multiplication by $\mathbf{J}_N$ of both members of Yule–Walker equations [5.89] yields:

$$\begin{aligned} \widetilde{\mathbf{c}}_N &= \mathbf{J}_N \mathbf{c}_N = -\mathbf{J}_N \mathbf{\Phi}_N \mathbf{a}_N = -\mathbf{\Phi}_N \mathbf{J}_N \mathbf{a}_N \\ &= -\mathbf{\Phi}_N \widetilde{\mathbf{a}}_N, \end{aligned} \qquad [5.98]$$

which means that Yule–Walker equations remain valid if we reverse the order of the components of vectors $\mathbf{a}_N$ and $\mathbf{c}_N$.

The recursive inversion formula [5.57] applied to the partitioned matrix $\mathbf{\Phi}_{N+1}$ defined in [5.92], with correspondences:

$$(\mathbf{M}_n, \mathbf{c}_n, \mathbf{r}_n, \sigma_n, k_n) \Leftrightarrow (\mathbf{\Phi}_N, \widetilde{\mathbf{c}}_N, \widetilde{\mathbf{c}}_N, \varphi_x(0), \sigma_N^{-2}),$$

gives:

$$\mathbf{\Phi}_{N+1}^{-1} = \begin{bmatrix} \mathbf{\Phi}_N^{-1} + \sigma_N^{-2}\mathbf{\Phi}_N^{-1}\widetilde{\mathbf{c}}_N\widetilde{\mathbf{c}}_N^T\mathbf{\Phi}_N^{-1} & \vdots & -\sigma_N^{-2}\mathbf{\Phi}_N^{-1}\widetilde{\mathbf{c}}_N \\ \cdots\cdots\cdots\cdots\cdots & & \cdots\cdots\cdots \\ -\sigma_N^{-2}\widetilde{\mathbf{c}}_N^T\mathbf{\Phi}_N^{-1} & \vdots & \sigma_N^{-2} \end{bmatrix}, \quad [5.99a]$$

$$\begin{aligned} \sigma_N^2 &= \varphi_x(0) - \widetilde{\mathbf{c}}_N^T\mathbf{\Phi}_N^{-1}\widetilde{\mathbf{c}}_N \\ &= \varphi_x(0) + \widetilde{\mathbf{c}}_N^T\widetilde{\mathbf{a}}_N \quad \text{(according to [5.98])} \\ &= \varphi_x(0) + \mathbf{c}_N^T\mathbf{a}_N \quad \text{(according to [5.96a])}. \end{aligned} \quad [5.99b]$$

Using partitioned forms [5.92] of $\mathbf{c}_{N+1}$, and [5.99a] of $\mathbf{\Phi}_{N+1}^{-1}$, and Yule–Walker equations [5.89] and [5.98], and bearing in mind that $\widetilde{\mathbf{c}}_N^T\mathbf{a}_N$ is a scalar, the $(N+1)$th-order solution of Yule–Walker equation [5.91] is written as:

$$\begin{aligned} \mathbf{a}_{N+1} &= -\mathbf{\Phi}_{N+1}^{-1}\mathbf{c}_{N+1} \\ &= \begin{bmatrix} \mathbf{a}_N + \sigma_N^{-2}[\varphi_x(N+1) + \widetilde{\mathbf{c}}_N^T\mathbf{a}_N]\mathbf{\Phi}_N^{-1}\widetilde{\mathbf{c}}_N \\ \cdots\cdots\cdots\cdots\cdots \\ -\sigma_N^{-2}[\varphi_x(N+1) + \widetilde{\mathbf{c}}_N^T\mathbf{a}_N] \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a}_N + k_{N+1}\mathbf{\Phi}_N^{-1}\widetilde{\mathbf{c}}_N \\ \cdots\cdots\cdots\cdots \\ -k_{N+1} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_N - k_{N+1}\widetilde{\mathbf{a}}_N \\ \cdots\cdots\cdots\cdots \\ -k_{N+1} \end{bmatrix}, \end{aligned} \quad [5.100]$$

where:

$$k_{N+1} = \rho_N^2/\sigma_N^2 \quad [5.101a]$$

$$\rho_N^2 = \varphi_x(N+1) + \widetilde{\mathbf{c}}_N^T\mathbf{a}_N. \quad [5.101b]$$

Taking into account partitionings [5.92] and [5.100], identity [5.96b], and definitions [5.101b] and [5.101a], identity [5.99b] at order $(N+1)$ can be written as:

$$\sigma_{N+1}^2 = \varphi_x(0) + \mathbf{c}_{N+1}^T\mathbf{a}_{N+1}. \quad [5.102a]$$

$$\begin{aligned} &= \varphi_x(0) + \begin{bmatrix} \mathbf{c}_N^T \vdots \varphi_x(N+1) \end{bmatrix} \begin{bmatrix} \mathbf{a}_N - k_{N+1}\widetilde{\mathbf{a}}_N \\ \cdots\cdots\cdots\cdots \\ -k_{N+1} \end{bmatrix} \\ &= \varphi_x(0) + \mathbf{c}_N^T\mathbf{a}_N - k_{N+1}[\varphi_x(N+1) + \mathbf{c}_N^T\widetilde{\mathbf{a}}_N]. \\ &= \sigma_N^2 - k_{N+1}\rho_N^2 = (1 - k_{N+1}^2)\sigma_N^2. \end{aligned} \quad [5.102b]$$

In summary, the Levinson–Durbin algorithm consists of equations [5.100], [5.101b], [5.101a] and [5.102b]. By decomposing the vector of AR coefficients using its components $a_{N,n}$, with $n \in \langle N \rangle$. the algorithm can be detailed in the following way:

1) Estimate the autocorrelation function $\varphi_x(\tau)$ using time averages.

2) Initialization: $N = 0, a_{0,0} = 1, \sigma_0^2 = \varphi_x(0)$.

3) Computation loop:

$$\rho_N^2 = \varphi_x(N+1) + \widetilde{\mathbf{c}}_N^T \mathbf{a}_N \qquad [5.103a]$$

$$k_{N+1} = \rho_N^2 / \sigma_N^2 \qquad [5.103b]$$

$$a_{N+1,n} = \begin{cases} a_{N,n} - k_{N+1} a_{N,N+1-n} \ , & n \in \langle N \rangle \\ -k_{N+1} \quad , & n = N+1 \end{cases} \qquad [5.103c]$$

$$\sigma_{N+1}^2 = (1 - k_{N+1}^2)\sigma_N^2. \qquad [5.103d]$$

4) Return to step 3 (with $N \to N+1$) until the stopping criterion be satisfied.

NOTE 5.24.– We can make the following observations:

– The coefficients $k_N$ are called partial correlation coefficients (PARCOR).

– The Levinson–Durbin algorithm allows to replace the inversion of the $(N+1)$th-order matrix $\Phi_{N+1}$ by the inversion of $N+1$ scalars for computing the PARCOR coefficients, which provides a fast computation of AR parameters.

– Iterating equation [5.103d] gives: $\sigma_{N+1}^2 = \prod_{n=1}^{N+1}(1 - k_n^2)\sigma_0^2$.

– Since $\sigma_N^2$ must be positive for all orders, the PARCOR coefficients are less than unit in absolute value: $|k_n| \leq 1, \forall n \in \langle N+1 \rangle$.

– In theory, an AR($N$) process can be whitened by filtering it through the inverse of the AR model.

– In practice, the stopping criterion of the algorithm can be built from the decrease of $\sigma_N^2$. For example, we can use:

$$\frac{\sigma_N^2 - \sigma_{N+1}^2}{\sigma_N^2} < \eta^2,$$

or equivalently $k_{N+1} < \eta$, where $\eta$ is the convergence threshold of the algorithm.

– For a model AR($N$), we have $\rho_N = 0$. Indeed, from the expression [5.87] of the autocorrelation function of an AR($N$) process, for $\tau = N+1$, we have:

$$\varphi_x(N+1) = -\widetilde{\mathbf{c}}_N^T \mathbf{a}_N \ \Rightarrow \ \rho_N = 0.$$

– As it will be seen in the following section, for one-step linear prediction, $\sigma_N^2$ represents the variance of the prediction error. From equation [5.102b], it can then be deduced that the increase of the order of the predictor implies a decrease of this variance, thereby an improvement in prediction quality.

EXAMPLE 5.25.– Let the autocorrelation function be defined by its first four points: $\varphi_x(0) = 12, \varphi_x(1) = 8, \varphi_x(2) = 2$, and $\varphi_x(3) = -2$. Applying the Levinson–Durbin algorithm gives:

– For N=0:

$$\begin{aligned}\rho_0^2 &= \varphi_x(1) = 8\\ k_1 &= \rho_0^2/\sigma_0^2 = 2/3\\ a_{1,1} &= -k_1 = -2/3\\ \sigma_1^2 &= (1-k_1^2)\sigma_0^2 = 20/3.\end{aligned}$$

– For N=1:

$$\begin{aligned}\rho_1^2 &= \varphi_x(2) + a_{1,1}\varphi_x(1) = -10/3\\ k_2 &= \rho_1^2/\sigma_1^2 = -1/2\\ a_{2,1} &= (1-k_2)a_{1,1} = -1\\ a_{2,2} &= -k_2 = 1/2\\ \sigma_2^2 &= (1-k_2^2)\sigma_1^2 = 5.\end{aligned}$$

– For N=2:

$$\rho_2^2 = \varphi_x(3) + a_{2,1}\varphi_x(2) + a_{2,2}\varphi_x(1) = 0.$$

It can be concluded that the considered autocorrelation function is that of an AR(2) process, whose vector of coefficients is $\mathbf{a}_2 = [-1,\ 1/2]^T$, modeled as:

$$x(k) = -\sum_{n=1}^{2} a_{2,n}x(k-n) + e(k) = x(k-1) - \frac{1}{2}x(k-2) + e(k)$$

the white noise $e(k)$ having variance $\sigma_e^2 = \sigma_2^2 = 5$. The AR coefficients and variance $\sigma_e^2$ can also be found by solving the following Yule–Walker equations:

$$\begin{bmatrix}\varphi_x(1)\\ \varphi_x(2)\end{bmatrix} = -\begin{bmatrix}\varphi_x(0) & \varphi_x(1)\\ \varphi_x(1) & \varphi_x(0)\end{bmatrix}\begin{bmatrix}a_{2,1}\\ a_{2,2}\end{bmatrix}$$

$$\sigma_e^2 = \varphi_x(0) + a_{2,1}\varphi_x(1) + a_{2,2}\varphi_x(2).$$

### 5.18.3. *Linear prediction*

We are going to show the relation that exists between the estimation of the AR parameters and that of the coefficients of one-step forward and backward linear predictors, optimal in the sense of the MMSE (for minimum mean square error). These predictors will be interpreted as orthogonal projectors over subspaces of the Hilbert space of second-order stationary random signals[7], defined on the same probability space, with the following inner product between two random variables $x$ and $y$:

$$\langle x, y \rangle = E[xy]. \qquad [5.104]$$

For centered random signals $x(k)$ and $y(k)$, their inner product corresponds to their intercorrelation function $\langle x(k+\tau), y(k) \rangle = E[x(k+\tau)y(k)] = \varphi_{xy}(\tau)$, and their orthogonality is equivalent to their non-correlation: $\varphi_{xy}(\tau) = 0$, $\forall \tau \in \mathbb{Z}$. In addition, the square of the norm of $x(k)$ is equal to its variance:

$$\|x(k)\|^2 = \langle x(k), x(k) \rangle = E[x^2(k)] = \varphi_x(0) = \sigma_x^2.$$

#### 5.18.3.1. *Forward linear predictor*

Let $x_N^{(d)}(k) = x_N(k/k-N, k-1)$ be the one-step forward linear predictor[8], of order $N$, that is, the linear predictor of $x(k)$ based on the $N$ previous samples $\{x(k-N), \cdots, x(k-1)\}$. This predictor is written in the form of a linear regression:

$$x_N^{(d)}(k) = \sum_{n=1}^{N} p_n x(k-n) = \mathbf{p}_N^T \mathbf{x}_N(k), \qquad [5.105]$$

where the regression vector $\mathbf{x}_N(k)$ is defined as in [5.86b].

The predictor optimal in the MMSE sense minimizes the variance of the prediction error with respect to the vector $\mathbf{p}_N = \left[p_1, \cdots, p_N\right]^T$ of the predictor coefficients, or more precisely the following optimality criterion:

$$\begin{aligned} J_{\text{MSE}} &= E\left[(x(k) - x_N^{(d)}(k))^2\right] = E\left[(x(k) - \mathbf{p}_N^T \mathbf{x}_N(k))^2\right] \\ &= \mathbf{p}_N^T \mathbf{\Sigma}_{\mathbf{x}_N} \mathbf{p}_N - 2\mathbf{p}_N^T \mathbf{\Sigma}_{x\mathbf{x}_N} + \varphi_x(0), \end{aligned} \qquad [5.106]$$

---

7 The intercorrelation function of two jointly weakly second-order stationary random signals $x(k)$ and $y(k)$ is such that $\varphi_{x,y}(k,t) = E[x(k)y(t)]$ only depends on the time interval $\tau = k-t$, that is $\varphi_{x,y}(k,t) = \varphi_{x,y}(k-t) = \varphi_{x,y}(\tau) = E[x(k+\tau)y(k)]$. Similarly, the covariance function is given by:

$$\sigma_{x,y}(\tau) = E[(x(k+\tau) - \mu_x)(y(k) - \mu_y)] = \varphi_{x,y}(\tau) - \mu_x \mu_y$$

where $\mu_x$ and $\mu_y$ are the mean of $x(k)$ and $y(k)$, respectively. For centered signals, that is, of zero mean, we have $\varphi_{x,y}(\tau) = \sigma_{x,y}(\tau)$.

8 Superscript $(d)$ denotes direct.

where:

$$\boldsymbol{\Sigma}_{\mathbf{x}_N} = E\big[\mathbf{x}_N(k)\mathbf{x}_N^T(k)\big] = \begin{bmatrix} \varphi_x(0) & \varphi_x(1) & \cdots & \varphi_x(N-1) \\ \varphi_x(1) & \varphi_x(0) & \cdots & \varphi_x(N-2) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_x(N-1) & \varphi_x(N-2) & \cdots & \varphi_x(0) \end{bmatrix} = \boldsymbol{\Phi}_N$$

$$\begin{aligned} \boldsymbol{\Sigma}_{x\mathbf{x}_N} &= E\big[x(k)\mathbf{x}_N(k)\big] && [5.107] \\ &= \big[E[x(k)x(k-1)], \cdots, E[x(k)x(k-N)]\big]^T \\ &= \big[\varphi_x(1), \cdots, \varphi_x(N)\big]^T = \mathbf{c}_N. && [5.108] \end{aligned}$$

$\mathbf{c}_N$ and $\boldsymbol{\Phi}_N$ are defined as in Yule–Walker equation [5.89].

The coefficient vector of the optimal forward predictor is obtained by canceling the gradient of $J_{\text{MSE}}$ with respect to the vector $\mathbf{p}_N$, which gives:

$$\frac{\partial J_{\text{MSE}}}{\partial \mathbf{p}_N} = 2\boldsymbol{\Sigma}_{\mathbf{x}_N}\widehat{\mathbf{p}}_N - 2\boldsymbol{\Sigma}_{x\mathbf{x}_N} = \mathbf{0}, \qquad [5.109]$$

or equivalently:

$$\mathbf{c}_N = \boldsymbol{\Phi}_N\widehat{\mathbf{p}}_N. \qquad [5.110]$$

We thus find the Yule–Walker equation [5.89] for estimating the coefficients of an AR($N$) model, with a change of sign due to the notation [5.105] of the predictor with respect to equation [5.85] of the AR model.

It should be noted that this solution for $\widehat{\mathbf{p}}_N$ effectively corresponds to a minimum of the criterion because the Hessian matrix $\frac{\partial^2 J_{\text{MSE}}}{\partial \mathbf{p}_N^2} = 2\boldsymbol{\Sigma}_{\mathbf{x}_N}$ is equal to the $N$th-order autocorrelation matrix of $x(k)$, up to factor 2, which is a positive semi-definite matrix, because: $\mathbf{u}^T\boldsymbol{\Sigma}_{\mathbf{x}_N}\mathbf{u} = \mathbf{u}^T E[\mathbf{x}_N(k)\mathbf{x}_N^T(k)]\mathbf{u} = E[(\mathbf{x}_N^T(k)\mathbf{u})^2] \geq 0$, $\forall \mathbf{u}$.

The variance of the order $N$ forward prediction error $\widetilde{x}_N^{(d)}(k) = x(k) - \widehat{\mathbf{x}}_N^{(d)}(k)$, associated with the optimal predictor $\widehat{\mathbf{x}}_N^{(d)}(k) = \widehat{\mathbf{p}}_N^T\mathbf{x}_N = \mathbf{c}_N^T\boldsymbol{\Phi}_N^{-1}\mathbf{x}_N$, is given by the minimum value of the criterion $J_{\text{MSE}}$ obtained by replacing $\mathbf{p}_N$ by $\widehat{\mathbf{p}}_N$ in [5.106]:

$$\begin{aligned} \sigma^2_{\widetilde{x}_N^{(d)}} &= E[(x(k) - \widehat{\mathbf{x}}_N^{(d)}(k))^2] = \widehat{\mathbf{p}}_N^T\boldsymbol{\Sigma}_{\mathbf{x}_N}\widehat{\mathbf{p}}_N - 2\widehat{\mathbf{p}}_N^T\boldsymbol{\Sigma}_{x\mathbf{x}_N} + \varphi_x(0) \\ &= \varphi_x(0) - \mathbf{c}_N^T\boldsymbol{\Phi}_N^{-1}\mathbf{c}_N = \varphi_x(0) - \mathbf{c}_N^T\widehat{\mathbf{p}}_N. && [5.111] \end{aligned}$$

By comparing [5.111] with [5.90], it can be concluded that the variance of the one-step prediction error, of order $N$, is the same as the variance of the input white noise of the AR($N$) model.

NOTE 5.26.– (Interpretation of the predictor in terms of orthogonal projector): The $N$th-order forward linear predictor of $x(k)$, defined by [5.105], can be interpreted as the orthogonal projector of $x(k)$ over the subspace spanned by signals $\{x(k-N),\cdots,x(k-1)\}$ of the Hilbert space of random signals, equipped with the inner product [5.104]. By following the same approach as in section 3.6.2, the vector of the predictor coefficients is then determined from the relationship of orthogonality between the prediction error and the signals $x(k-\tau), \tau \in \langle N\rangle$, of the projection subspace:

$$\langle \widetilde{x}_N^{(d)}(k), x(k-\tau)\rangle = E[\widetilde{x}_N^{(d)}(k)\, x(k-\tau)] = 0 \ , \ \ \forall \tau \in \langle N\rangle \qquad [5.112]$$

which leads to:

$$E\big[\big(x(k) - \sum_{n=1}^{N} \widehat{p}_n x(k-n)\big)x(k-\tau)\big] = \ \varphi_x(\tau) - \sum_{n=1}^{N} \widehat{p}_n \varphi_x(\tau - n) = 0, \qquad [5.113]$$

namely, the Yule–Walker equation [5.110]. In addition, the mean square projection error can be obtained by using the orthogonality property [5.112]:

$$E\big[(\widetilde{x}_N^{(d)}(k))^2\big] = E\big[\widetilde{x}_N^{(d)}(k)\,\big(x(k) - \sum_{n=1}^{N} \widehat{p}_n x(k-n)\big)\big] = E\big[\widetilde{x}_N^{(d)}(k)\, x(k)\big]$$

$$= E\big[\big(x(k) - \sum_{n=1}^{N} \widehat{p}_n x(k-n)\big)\, x(k)\big] = \varphi_x(0) - \sum_{n=1}^{N} \widehat{p}_n \varphi_x(\tau),$$

which corresponds to the variance of the prediction error [5.111].

### 5.18.3.2. *Backward linear predictor*

We now consider the one-step backward linear predictor, of order $N$, or more specifically the linear predictor of $x(k-N)$ based on the $N$ samples $\{x(k-N+1),\cdots,x(k)\}$, posterior to $x(k-N)$, in the form:

$$x_N^{(b)}(k-N) = \sum_{n=1}^{N} b_n x(k-N+n) = \mathbf{b}_N^T \widetilde{\mathbf{x}}_N(k+1), \qquad [5.114]$$

where superscript $(b)$ denotes backward, $\mathbf{b}_N = \big[b_1,\cdots,b_N\big]^T$ is the coefficient vector of the backward predictor, and $\widetilde{\mathbf{x}}_N(k+1) = \big[x(k-N+1)\cdots,x(k)\big]^T$, namely, the vector $\mathbf{x}_N(k+1)$ whose components are taken in reverse order.

Following the same approach as for the forward predictor, the optimal backward predictor can be deduced by minimizing the criterion $J_{\text{MSE}}$ defined as:

$$J_{\text{MSE}} = E\Big[\Big(x(k-N) - x_N^{(b)}(k-N)\Big)^2\Big] = E\Big[\Big(x(k-N) - \mathbf{b}_N^T \widetilde{\mathbf{x}}_N(k+1)\Big)^2\Big]$$

$$= \mathbf{b}_N^T \boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}_N} \mathbf{b}_N - 2\mathbf{b}_N^T \boldsymbol{\Sigma}_{x\widetilde{\mathbf{x}}_N} + \varphi_x(0), \qquad [5.115]$$

where:

$$\begin{aligned}\boldsymbol{\Sigma}_{\widetilde{\mathbf{x}}_N} &= E\left[\widetilde{\mathbf{x}}_N(k+1)\widetilde{\mathbf{x}}_N^T(k+1)\right] = \boldsymbol{\Phi}_N \\ \boldsymbol{\Sigma}_{x\widetilde{\mathbf{x}}_N} &= E\left[x(k-N)\widetilde{\mathbf{x}}_N(k+1)\right] \\ &= \left[E[x(k-N)x(k-N+1)], \cdots, E[x(k-N)x(k)]\right]^T \\ &= \left[\varphi_x(1), \cdots, \varphi_x(N)\right]^T = \mathbf{c}_N.\end{aligned}$$

The minimization of the criterion [5.115] gives the following solution:

$$\mathbf{c}_N = \boldsymbol{\Phi}_N \widehat{\mathbf{b}}_N, \qquad [5.116]$$

and the variance of the order $N$ backward prediction error $\widetilde{x}_N^{(r)}(k-N) = x(k-N) - \widehat{x}_N^{(r)}(k-N)$, associated with the optimal predictor $\widehat{\mathbf{x}}_N^{(r)}(k-N) = \widehat{\mathbf{b}}_N^T \widetilde{\mathbf{x}}_N(k+1) = \mathbf{c}_N^T \boldsymbol{\Phi}_N^{-1} \widetilde{x}_N(k+1)$, is given by:

$$\begin{aligned}\sigma^2_{\widetilde{x}_N^{(r)}} &= E[(x(k-N) - \widehat{x}_N^{(r)}(k-N))^2] \\ &= \varphi_x(0) - \mathbf{c}_N^T \boldsymbol{\Phi}_N^{-1} \mathbf{c}_N = \varphi_x(0) - \mathbf{c}_N^T \widehat{\mathbf{b}}_N. \end{aligned} \qquad [5.117]$$

From equations [5.89], [5.110], and [5.116], we deduce that the coefficients of the order $N$ optimal backward predictor are equal to those of the order $N$ optimal forward predictor and to the AR($N$) parameters with the opposite sign: $\widehat{\mathbf{b}}_N = \widehat{\mathbf{p}}_N = -\mathbf{a}_N$. In addition, from [5.111] and [5.117], it is concluded that order $N$ forward and backward prediction errors have the same variance than the input white noise of the AR($N$) model.

NOTE 5.27.– As for the forward predictor, it can be shown that the backward predictor is the orthogonal projector of $x(k-N)$ over the subspace spanned by the signals $\{x(k-N+1), \cdots, x(k-1)\}$.

In Table 5.2, we summarize the links highlighted between AR model and forward and backward linear predictors.

| Regressor and autocorrelation vectors | |
|---|---|
| $\mathbf{x}_N^T(k) = [x(k-1), \cdots, x(k-N)]$ | |
| $\widetilde{\mathbf{x}}_N(k+1) = \mathbf{J}_N \mathbf{x}_N(k+1)$ | |
| $\mathbf{c}_N^T = [\varphi_x(1), \cdots, \varphi_x(N)]$ | |
| **AR model** | |
| $x(k) = -\mathbf{a}_N^T \mathbf{x}_N(k) + e(k)$ | |
| $\sigma_e^2 = E[e^2(k)]$ | |
| $\mathbf{c}_N = -\mathbf{\Phi}_N \mathbf{a}_N$ | |
| **Forward predictor** | **Backward predictor** |
| $x_N^{(d)}(k) = \mathbf{p}_N^T \mathbf{x}_N(k)$ | $x_N^{(r)}(k-N) = \mathbf{b}_N^T \widetilde{\mathbf{x}}_N(k+1)$ |
| $\mathbf{c}_N = \mathbf{\Phi}_N \widehat{\mathbf{p}}_N$ | $\mathbf{c}_N = \mathbf{\Phi}_N \widehat{\mathbf{b}}_N$ |
| $\widehat{x}_N^{(d)}(k) = \widehat{\mathbf{p}}_N^T \mathbf{x}_N(k)$ | $\widehat{x}_N^{(r)}(k-N) = \widehat{\mathbf{b}}_N^T \widetilde{\mathbf{x}}_N(k+1)$ |
| $\widetilde{x}_N^{(d)}(k) = x(k) - \widehat{x}_N^{(d)}(k)$ | $\widetilde{x}_N^{(r)}(k-N) = x(k-N) - \widehat{x}_N^{(r)}(k-N)$ |
| $\sigma^2_{\widetilde{x}_N^{(d)}} = E\left[\left(\widetilde{x}_N^{(d)}(k)\right)^2\right]$ | $\sigma^2_{\widetilde{x}_N^{(r)}} = E\left[\left(\widetilde{x}_N^{(r)}(k-N)\right)^2\right]$ |
| $\sigma_e^2 = \sigma^2_{\widetilde{x}_N^{(d)}} = \sigma^2_{\widetilde{x}_N^{(r)}}$ $= \varphi_x(0) - \mathbf{c}_N^T \mathbf{\Phi}_N^{-1} \mathbf{c}_N$ | |

**Table 5.2.** *AR model and one-step linear predictors: Yule–Walker equations*

# 6

# Tensor Spaces and Tensors

## 6.1. Chapter summary

The concepts of hypermatrix and outer product, as well as the operations of contraction and $n$-mode hypermatrix–matrix product, are first defined. The links between multilinear forms and multilinear maps, on the one hand, and homogeneous polynomials and hypermatrices, on the other hand, are established. The special cases of bilinear maps and bilinear forms are considered. The notion of symmetric multilinear form allows us to introduce the concept of symmetric hypermatrix. The case of symmetric hypermatrices of third and fourth orders is detailed.

$N$th-order tensors are then introduced based on the tensor product of $N$ v.s., called tensor space. The hypermatrix of coordinates of a tensor with respect to a given basis is highlighted, and the canonical writing of tensors is presented. The universal property of the tensor product is described and illustrated by means of vector Khatri–Rao and dot products. The effect of a change of basis in the vector spaces of a tensor product, on the coordinate hypermatrix of a tensor, is analyzed. The notions of tensor rank and tensor decomposition are introduced through the presentation of the canonical polyadic decomposition (CPD) which plays a fundamental role in numerous applications. The notions of eigenvalue and singular value of an hypermatrix are then defined as extensions of those for matrices, presented in section 4.16. Finally, different examples of isomorphisms of tensor spaces are described.

In this chapter, we present the main notations and definitions relating to hypermatrices and tensors that will be used in the following volumes. Basic tensor operations as well as the main tensor models will be presented in Volume 2.

## 6.2. Hypermatrices

Before introducing tensors in a formal way as elements of a tensor space, we are going to describe hypermatrix v.s.

*From Algebraic Structures to Tensors,*
First Edition. Edited by Gérard Favier.

### 6.2.1. *Hypermatrix vector spaces*

A hypermatrix $\mathcal{A} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ of order $N$ and of dimensions $I_1 \times \cdots \times I_N$, denoted $[a_{i_1,\cdots,i_N}]$ or $[a_{i_1 \cdots i_N}]$, is an array composed of elements depending on $N$ indices $i_n \in \langle I_n \rangle$, with $n \in \langle N \rangle$. Indices are also called modes, dimensions or axes. The hypermatrix $\mathcal{A} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ can be defined as a map $f : \overset{N}{\underset{n=1}{\times}} \langle I_n \rangle \to \mathbb{K}$ such that:

$$\overset{N}{\underset{n=1}{\times}} \langle I_n \rangle \ni (i_1, \cdots, i_N) \mapsto f(i_1, \cdots, i_N) = a_{i_1,\cdots,i_N} \in \mathbb{K},$$

where $I_n$ is the upper bound of index $i_n$, corresponding to the $n$-mode dimension.

Hypermatrices generalize vectors and matrices to orders higher than two, and the special cases of hypermatrices of orders $N = 1$ and $N = 2$ correspond to vectors and matrices, respectively.

If $I_1 = \cdots = I_N = I$, the hypermatrix $\mathcal{A} \in \mathbb{K}^{I \times I \times \cdots \times I}$ of order $N$ is said to be hypercubic.

Moreover, if $a_{i_1,\cdots,i_N} \neq 0$ only for $i_1 = i_2 = \cdots = i_N$, then $\mathcal{A}$ is called a diagonal hypermatrix. The set of elements $a_{i,i,\cdots,i}$, with $i \in \langle I \rangle$, form the diagonal of $\mathcal{A}$, and the sum of these elements is the trace of $\mathcal{A}$, denoted by $\mathrm{tr}(\mathcal{A})$:

$$\mathrm{tr}(\mathcal{A}) = \sum_{i=1}^{I} a_{i,i,\cdots,i}.$$

The identity hypermatrix of order $N$ and of dimensions $I \times \cdots \times I$ is denoted $\mathcal{I}_{N,I} = [\delta_{i_1,\cdots,i_N}]$, with $i_n \in \langle I \rangle$ for $n \in \langle N \rangle$, or simply $\mathcal{I}$. This is a hypercubic hypermatrix whose elements are defined using the generalized Kronecker delta:

$$\delta_{i_1,\cdots,i_N} = \begin{cases} 1 & \text{if} \quad i_1 = \cdots = i_N \\ 0 & \text{otherwise} \end{cases}.$$

Let $\mathcal{A} = [a_{i_1,\cdots,i_N}], \mathcal{B} = [b_{i_1,\cdots,i_N}]$, and $\lambda \in \mathbb{K}$. By defining addition and scalar multiplication operations for hypermatrices of $\mathbb{K}^{I_1 \times \cdots \times I_N}$, such as:

$$\begin{aligned} \mathcal{A} + \mathcal{B} &= [a_{i_1,\cdots,i_N} + b_{i_1,\cdots,i_N}] \\ \lambda\, \mathcal{A} &= [\lambda\, a_{i_1,\cdots,i_N}], \end{aligned}$$

the set $\mathbb{K}^{I_1 \times \cdots \times I_N}$ of hypermatrices, of order $N$, is a v.s. of dimension $\prod_{n=1}^{N} I_n$. This dimension expresses the fact that a hypermatrix of $\mathbb{K}^{I_1 \times \cdots \times I_N}$ can be vectorized as a vector of $\mathbb{K}^{\prod_{n=1}^{N} I_n}$. The vectorization operation will be detailed in Volume 2.

### 6.2.2. *Hypermatrix inner product and Frobenius norm*

Given two complex hypermatrices $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$, their Hermitian inner product is defined as:

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} a_{i_1,\cdots,i_N} b^*_{i_1,\cdots,i_N} = a_{i_1,\cdots,i_N} b^*_{i_1,\cdots,i_N} \quad [6.1]$$

where the last equality results from the use of Einstein's summation convention which consists in summing over repeated indices.

For example, $\sum_{i=1}^{I} x_i \mathbf{e}_i^{(I)}$ is replaced by $x_i \mathbf{e}_i^{(I)}$. Similarly, $\sum_{j=1}^{J} a_{ij} b_{jk}$ will be replaced by $a_{ij} b_{jk}$. This convention, which makes it possible to simplify the writing of equations involving quantities defined using multiple indices, will be utilized later in this chapter. In Volume 2, it will be used in a generalized form.

The Frobenius[1] norm, also called Hilbert–Schmidt norm, of $\mathcal{A}$ is defined as:

$$\|\mathcal{A}\|_F^2 = \langle \mathcal{A}, \mathcal{A} \rangle = a_{i_1,\cdots,i_N} a^*_{i_1,\cdots,i_N} = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} |a_{i_1,\cdots,i_N}|^2. \quad [6.2]$$

In the matrix case, for $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{I \times J}$, equations [6.1] and [6.2] become:

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} b^*_{ij} = a_{ij} b^*_{ij} \quad [6.3a]$$

$$\|\mathbf{A}\|_F^2 = \langle \mathbf{A}, \mathbf{A} \rangle = \sum_{i=1}^{I} \sum_{j=1}^{J} |a_{ij}|^2 = a_{ij} a^*_{ij}. \quad [6.3b]$$

NOTE 6.1.– In the case of two real hypermatrices $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, their Euclidean inner product and the Frobenius norm are obtained by removing the conjugation in definitions [6.1] and [6.2].

### 6.2.3. *Contraction operation and $n$-mode hypermatrix–matrix product*

Two important operations that will be used in this chapter are defined here. On the one hand, the contraction of two hypermatrices sharing common modes.

1 Ferdinand Georg Frobenius (1849–1917), German mathematician, member of the Academy of Sciences and Letters of Berlin, who was a student of Weierstrass and Kronecker. He made significant contributions to group theory, in particular for the representation of symmetric and alternating groups, in linear algebra, in analysis, and number theory. Several concepts, theorems, and methods are named after him, for example, the norm, the decomposition, the endomorphism, or the Frobenius algebra.

On the other hand, the $n$-mode product of a hypermatrix with a matrix having a common mode, which can be viewed as a special case of contraction.

#### 6.2.3.1. *Contraction operation*

Let $\mathcal{A} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ and $\mathcal{B} \in \mathbb{K}^{J_1 \times \cdots \times J_P}$ be two hypermatrices of orders $N$ and $P$, respectively, with $M$ common modes. Contraction corresponds to summations over common modes. The result of this contraction is a hypermatrix $\mathcal{C}$ of order $N+P-2M$.

EXAMPLE 6.2.– For $\mathcal{A} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$ and $\mathcal{B} \in \mathbb{K}^{I_2 \times I_3 \times J_1 \times J_2}$, the contraction over common indices $i_2$ and $i_3$ gives the contracted hypermatrix $\mathcal{C} \in \mathbb{K}^{I_1 \times J_1 \times J_2}$ such that $c_{i_1,j_1,j_2} = \sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} a_{i_1,i_2,i_3} b_{i_2,i_3,j_1,j_2}$. We write $C = \mathcal{A} \underset{1,2}{\overset{2,3}{\times}} \mathcal{B}$, where numbers $\{1,2\}$ are relative to the position of the common indices of tensor $\mathcal{B}$, while numbers $\{2,3\}$ define the position of the common indices of $\mathcal{A}$.

FACT 6.3.– For $\mathbb{K} = \mathbb{R}$ and $M = N = P$, the contraction of $\mathcal{A}$ and $\mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ corresponds to their inner product:

$$\mathcal{A} \underset{1,2,\cdots,N}{\overset{1,2,\cdots,N}{\times}} \mathcal{B} = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} a_{i_1,\cdots,i_N} b_{i_1,\cdots,i_N} = \langle \mathcal{A}, \mathcal{B} \rangle.$$

#### 6.2.3.2. *$n$-mode hypermatrix–matrix product*

The $n$-mode product of the hypermatrix $\mathcal{X} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ with the matrix $\mathbf{A} \in \mathbb{K}^{J_n \times I_n}$, denoted by $\mathcal{X} \times_n \mathbf{A}$, is equivalent to a contraction of the hypermatrix $\mathcal{X}$ with matrix $\mathbf{A}$ over their common mode $i_n$. This contraction gives the $N$-order hypermatrix $\mathcal{Y}$ of dimensions $I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N$, such that (Carroll *et al.* 1980):

$$y_{i_1,\ldots,i_{n-1},j_n,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_N} a_{j_n,i_n} x_{i_1,\ldots,i_{n-1},i_n,i_{n+1},\ldots,i_N}. \qquad [6.4]$$

Thus, for $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$, $\mathbf{A} \in \mathbb{K}^{P \times I}$, $\mathbf{B} \in \mathbb{K}^{Q \times J}$, $\mathbf{C} \in \mathbb{K}^{R \times K}$, we have:

$$(\mathcal{X} \times_1 \mathbf{A})_{p,j,k} = \sum_{i=1}^{I} a_{pi} x_{i,j,k}$$

$$(\mathcal{X} \times_2 \mathbf{B})_{i,q,k} = \sum_{j=1}^{J} b_{qj} x_{i,j,k}$$

$$(\mathcal{X} \times_3 \mathbf{C})_{i,j,r} = \sum_{k=1}^{K} c_{rk} x_{i,j,k}$$

and therefore, we deduce that:

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \qquad [6.5]$$

$$\Updownarrow$$

$$y_{p,q,r} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} a_{pi} b_{qj} c_{rk}\, x_{i,j,k}. \qquad [6.6]$$

In Volume 2, we shall see that this equation corresponds to a Tucker decomposition of the tensor $\mathcal{Y}$ (Tucker 1966). This is also referred to as the Tucker model for $\mathcal{Y}$. Then, $\mathcal{X}$ is called the core tensor and $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ are the factor matrices of the decomposition.

More generally, by considering a permutation $\pi(.)$ of the first $N$ natural numbers $n \in \langle N \rangle$ such as $p_n = \pi(n)$, a series of $p_n$-mode products of $\mathcal{X} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ with $\mathbf{A}^{(p_n)} \in \mathbb{K}^{J_{p_n} \times I_{p_n}}$, $n \in \langle N \rangle$, will be concisely denoted as:

$$\mathcal{X} \times_{p_1} \mathbf{A}^{(p_1)} \cdots \times_{p_N} \mathbf{A}^{(p_N)} = \mathcal{X} \times_{p=p_1}^{p_N} \mathbf{A}^{(p)} \in \mathbb{K}^{J_{p_1} \times \cdots \times J_{p_N}}. \qquad [6.7]$$

In the case where the hypermatrix–matrix product is performed for all modes, another concise notation of [6.7] was proposed by de Silva and Lim (2008):

$$\mathcal{X} \times_{p_1} \mathbf{A}^{(p_1)} \cdots \times_{p_N} \mathbf{A}^{(p_N)} = (\mathbf{A}^{(p_1)}, \cdots, \mathbf{A}^{(p_N)}) \cdot \mathcal{X} \qquad [6.8]$$

where the order of the matrices $\mathbf{A}^{(p_1)}, \cdots, \mathbf{A}^{(p_N)}$ is identical to the order of the $p_n$-mode products, that is, $\times_{p_1}, \cdots, \times_{p_N}$.

Similarly, it is possible to define $n$-mode product of $\mathcal{X} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ with vector $\mathbf{u}^{(n)} \in \mathbb{K}^{I_n}$, for $n \in \langle N \rangle$. This gives a scalar:

$$y = \mathcal{X} \times_{n=1}^{N} \mathbf{u}^{(n)} = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} u_i^{(1)} \cdots u_i^{(N)} x_{i_1, \cdots, i_N}. \qquad [6.9]$$

### 6.2.3.3. *Properties of hypermatrix-matrix product*

The $n$-mode product plays a very important role in tensor calculus, so we describe its main properties hereafter.

PROPOSITION 6.4.– *The $n$-mode product satisfies the following properties:*

*– For any permutation $\pi(.)$ of the first $N$ natural numbers $n \in \langle N \rangle$ such as $p_n = \pi(n)$, we have:*

$$\mathcal{X} \times_{p=p_1}^{p_N} \mathbf{A}^{(p)} = \mathcal{X} \times_{n=1}^{N} \mathbf{A}^{(n)},$$

*which means that the order of the $n$-mode products is irrelevant when modes $n$ are all distinct.*

*– For two products of $\mathcal{X} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ along $n$-mode, with $\mathbf{A} \in \mathbb{K}^{J_n \times I_n}$ and $\mathbf{B} \in \mathbb{K}^{K_n \times J_n}$, we have (de Lathauwer 1997):*

$$\mathcal{Y} = \mathcal{X} \times_n \mathbf{A} \times_n \mathbf{B} = \mathcal{X} \times_n (\mathbf{BA}) \in \mathbb{K}^{I_1 \times \cdots \times I_{n-1} \times K_n \times I_{n+1} \times \cdots \times I_N}. \quad [6.10]$$

PROOF.– Defining $\mathcal{Z} = \mathcal{X} \times_n \mathbf{A}$, it is deduced that $\mathcal{Y} = \mathcal{Z} \times_n \mathbf{B}$. Then, using definition [6.4] of the $n$-mode product, we can write:

$$z_{i_1,\ldots,i_{n-1},j_n,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} a_{j_n,i_n} x_{i_1,\ldots,i_{n-1},i_n,i_{n+1},\ldots,i_N} \quad [6.11]$$

$$y_{i_1,\ldots,i_{n-1},k_n,i_{n+1},\ldots,i_N} = \sum_{j_n=1}^{J_n} b_{k_n,j_n} y_{i_1,\ldots,i_{n-1},j_n,i_{n+1},\ldots,i_N}. \quad [6.12]$$

Using expression [6.11] in [6.12] and reversing the order of summations, we obtain:

$$y_{i_1,\ldots,i_{n-1},k_n,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} \Big[ \sum_{j_n=1}^{J_n} b_{k_n,j_n} a_{j_n,i_n} \Big] x_{i_1,\ldots,i_{n-1},i_n,i_{n+1},\ldots,i_N}.$$

The summation in brackets gives the current element $c_{k_n,i_n}$ of matrix $\mathbf{C} = \mathbf{BA}$, which demonstrates property [6.10]. □

PROPOSITION 6.5.– *The $n$-mode product satisfies the other two following properties:*

*– For $\mathcal{X} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$, $\mathbf{A}^{(n)} \in \mathbb{K}^{J_n \times I_n}$ and $\mathbf{B}^{(n)} \in \mathbb{K}^{K_n \times J_n}$, $n \in \langle N \rangle$, property [6.10] can be generalized as follows:*

$$\mathcal{Y} = \mathcal{X} \times_{n=1}^{N} \mathbf{A}^{(n)} \times_{n=1}^{N} \mathbf{B}^{(n)} = \mathcal{X} \times_{n=1}^{N} (\mathbf{B}^{(n)} \mathbf{A}^{(n)}) \in \mathbb{K}^{K_1 \times \cdots \times K_N}. \quad [6.13]$$

*– If the factors $\mathbf{A}^{(n)}$ are full column rank, we have*

$$\mathcal{Y} = \mathcal{X} \times_{n=1}^{N} \mathbf{A}^{(n)} \;\Leftrightarrow\; \mathcal{X} = \mathcal{Y} \times_{n=1}^{N} \mathbf{A}^{(n)\dagger},$$

*where $\mathbf{A}^{(n)\dagger}$ denotes the Moore–Penrose pseudo-inverse of $\mathbf{A}^{(n)}$.*

NOTE 6.6.– Considering the rows and the columns of a matrix as its 1-mode and 2-mode, respectively, the $n$-mode product can be used for matrix multiplication in such a way that: $\mathbf{A} \times_1 \mathbf{B} = \mathbf{BA}$ and $\mathbf{A} \times_2 \mathbf{C} = \mathbf{AC}^T$, which gives:

$$\mathbf{A} \times_1 \mathbf{B} \times_2 \mathbf{C} = \mathbf{BAC}^T \in \mathbb{K}^{K \times L},$$

with $\mathbf{A} \in \mathbb{K}^{I \times J}$, $\mathbf{B} \in \mathbb{K}^{K \times I}$, $\mathbf{C} \in \mathbb{K}^{L \times J}$.

## 6.3. Outer products

The outer product[2] of two vectors $\mathbf{u} \in \mathbb{K}^I$ and $\mathbf{v} \in \mathbb{K}^J$, denoted $\mathbf{u} \circ \mathbf{v}$ and also called dyadic product of $\mathbf{u}$ with $\mathbf{v}$, gives a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ of rank one[3], such that $a_{ij} = (\mathbf{u} \circ \mathbf{v})_{ij} = u_i v_j$, and therefore, $\mathbf{u} \circ \mathbf{v} = \mathbf{u}\mathbf{v}^T = [u_i v_j]$, with $i \in \langle I \rangle, j \in \langle J \rangle$.

The outer product is to be compared with the inner product of $\mathbf{u}, \mathbf{v} \in \mathbb{K}^I$ which provides a scalar: $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v} = \sum_{i=1}^{I} u_i v_i$.

EXAMPLE 6.7.– For $I = 2, J = 3$, we have:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \circ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \end{bmatrix}.$$

The outer product of three vectors $\mathbf{u} \in \mathbb{K}^I, \mathbf{v} \in \mathbb{K}^J$, and $\mathbf{w} \in \mathbb{K}^K$ gives a rank one third-order hypermatrix $\mathcal{A} \in \mathbb{K}^{I \times J \times K}$, such that $a_{ijk} = (\mathbf{u} \circ \mathbf{v} \circ \mathbf{w})_{ijk} = u_i v_j w_k$, and therefore, $\mathbf{u} \circ \mathbf{v} \circ \mathbf{w} = [u_i v_j w_k]$, with $i \in \langle I \rangle, j \in \langle J \rangle, k \in \langle K \rangle$.

For the outer product of $N$ vectors $\mathbf{u}^{(n)} \in \mathbb{K}^{I_n}$, denoted by $\overset{N}{\underset{n=1}{\circ}} \mathbf{u}^{(n)}$, a rank one hypermatrix $\mathcal{U} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$, of order $N$, is obtained such that:

$$\mathcal{U} = [u_{i_1, \cdots, i_N}] = [\prod_{n=1}^{N} u_{i_n}^{(n)}] \ , \ i_n \in \langle I_n \rangle \ , \ n \in \langle N \rangle.$$

When the vectors $\mathbf{u}^{(n)}$ are identical, that is, $\mathbf{u}^{(n)} = \mathbf{u}$, for $n \in \langle N \rangle$, their outer product $\overset{N}{\underset{n=1}{\circ}} \mathbf{u}^{(n)}$ is written as $\mathbf{u}^{\circ N}$.

EXAMPLE 6.8.– Consider the outer product defined as:

$$\mathcal{B} = \begin{bmatrix} i \\ 1 \end{bmatrix} \circ \begin{bmatrix} i \\ 1 \end{bmatrix} \circ \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} i \\ 1 \end{bmatrix}^{\circ 3} \in \mathbb{C}^{2 \times 2 \times 2}, \qquad [6.14]$$

with $i^2 = -1$. Ranging the elements $b_{ijk}$ of the hypermatrix $\mathcal{B}$, with $i, j, k \in \langle 2 \rangle$, in a matrix defined as:

$$\mathbf{B} = \left[\begin{array}{cc|cc} b_{111} & b_{112} & b_{121} & b_{122} \\ b_{211} & b_{212} & b_{221} & b_{222} \end{array}\right] \in \mathbb{C}^{2 \times 4}, \qquad [6.15]$$

2 It should be noted that we use the same symbol $\circ$ to designate the outer product, the map composition and an external law, due to the fact that these are the most commonly used notations, the mathematical context allowing the elimination of any ambiguity.

3 The notions of rank-one matrix, hypermatrix and tensor are defined in section 6.7.

we have:

$$\mathbf{B} = \left[\begin{array}{cc|cc} -i & -1 & -1 & i \\ -1 & i & i & 1 \end{array}\right] \in \mathbb{C}^{2\times 4}. \qquad [6.16]$$

Similarly, for the outer product:

$$\mathcal{C} = \begin{bmatrix} -i \\ 1 \end{bmatrix}^{\circ 3} \in \mathbb{C}^{2\times 2\times 2}, \qquad [6.17]$$

the matrix $\mathbf{C}$ corresponding to the arrangement (6.15) is given by:

$$\mathbf{C} = \left[\begin{array}{cc|cc} i & -1 & -1 & -i \\ -1 & -i & -i & 1 \end{array}\right] \in \mathbb{C}^{2\times 4}. \qquad [6.18]$$

Matrices $\mathbf{B}$ and $\mathbf{C}$ above are called matrix unfoldings of the hypermatrices $\mathcal{B}$ and $\mathcal{C}$, respectively. Such unfoldings will be considered in more detail in section 6.9.

The identity hypermatrix $\mathcal{I}_{N,I} = [\delta_{i_1,\cdots,i_N}]$, defined in section 6.2.1, can be written using outer products of canonical basis vectors as:

$$\mathcal{I}_{N,I} = \sum_{i=1}^{I} \underbrace{\mathbf{e}_i^{(I)} \circ \cdots \circ \mathbf{e}_i^{(I)}}_{N \text{ terms}},$$

where $\mathbf{e}_i^{(I)} \circ \cdots \circ \mathbf{e}_i^{(I)}$ is the $N$th-order hypermatrix composed of 0's except one 1 on the diagonal, at position $i_1 = \cdots = i_N = i$, with $i \in \langle I \rangle$.

More generally, given $\mathbf{u} \in \mathbb{K}^K, \mathbf{A} \in \mathbb{K}^{I\times J}, \mathbf{B} \in \mathbb{K}^{K\times N}, \mathcal{A} \in \mathbb{K}^{I_1\times\cdots\times I_M}$, and $\mathcal{B} \in \mathbb{K}^{J_1\times\cdots\times J_N}$, we define the following outer products:

$$\begin{aligned} &\mathbf{A} \circ \mathbf{u} = \mathcal{C} \in \mathbb{K}^{I\times J\times K} && , \quad c_{ijk} = a_{ij}u_k \\ &\mathbf{A} \circ \mathbf{B} = \mathcal{C} \in \mathbb{K}^{I\times J\times K\times N} && , \quad c_{ijkn} = a_{ij}b_{kn} \\ &\mathcal{A} \circ \mathcal{B} = \mathcal{C} \in \mathbb{K}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_N} && , \quad c_{i_1\cdots i_M j_1\cdots j_N} = a_{i_1\cdots i_M} b_{j_1\cdots j_N}. \end{aligned}$$

The vector spaces $\mathbb{K}^{I\times J}$ of matrices and $\mathbb{K}^{I_1\times\cdots\times I_N}$ of hypermatrices of order $N$ have the following respective canonical bases:

$$\mathbf{E}_{ij}^{(I\times J)} = \mathbf{e}_i^{(I)} \circ \mathbf{e}_j^{(J)} \qquad [6.19a]$$

$$\mathcal{E}_{i_1\cdots i_N}^{(I_1\times\cdots\times I_N)} = \mathbf{e}_{i_1}^{(I_1)} \circ \cdots \circ \mathbf{e}_{i_N}^{(I_N)} = \bigcirc_{n=1}^{N} \mathbf{e}_{i_n}^{(I_n)}, \qquad [6.19b]$$

with $i \in \langle I \rangle$, $j \in \langle J \rangle$, and $i_n \in \langle I_n \rangle$ for $n \in \langle N \rangle$. Matrix $\mathbf{E}_{ij}^{(I \times J)}$ and hypermatrix $\mathcal{E}_{i_1 \cdots i_N}^{(I_1 \times \cdots \times I_N)}$ contain one 1 at positions $(i, j)$ and $(i_1, \cdots, i_N)$, respectively, and 0's elsewhere.

## 6.4. Multilinear forms, homogeneous polynomials and hypermatrices

### 6.4.1. *Hypermatrix associated to a multilinear form*

Consider a multilinear form $\psi : \overset{N}{\underset{n=1}{\times}} E_n \rightarrow \mathbb{K}$, as defined in section 2.5.11.2, with $\dim(E_n) = I_n$, such that[4]:

$$\overset{N}{\underset{n=1}{\times}} E_n \ni (\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) \mapsto \psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) \in \mathbb{K},$$

with the basis $\mathcal{B}^{(I_n)} = \{\mathbf{b}_1^{(I_n)}, \cdots, \mathbf{b}_{I_n}^{(I_n)}\}$ in $E_n$, $n \in \langle N \rangle$. Let us define the scalar coefficients equal to the images of these basis vectors by $\psi$:

$$b_{i_1, \cdots, i_N} = \psi(\mathbf{b}_{i_1}^{(I_1)}, \cdots, \mathbf{b}_{i_N}^{(I_N)}) \in \mathbb{K} \ , \ i_n \in \langle I_n \rangle, \ n \in \langle N \rangle,$$

and express $\psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)})$ in terms of the coordinates of the vectors $\mathbf{x}^{(n)}$, in the bases $\mathcal{B}^{(I_n)}$, that is, $\mathbf{x}^{(n)} = \sum_{i_n=1}^{I_n} c_{i_n}^{(n)} \mathbf{b}_{i_n}^{(I_n)}$. Using the multilinearity property of $\psi$ and Einstein's convention, we get:

$$\begin{aligned}
\psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) &= \psi(\sum_{i_1=1}^{I_1} c_{i_1}^{(1)} \mathbf{b}_{i_1}^{(I_1)}, \cdots, \sum_{i_N=1}^{I_N} c_{i_N}^{(N)} \mathbf{b}_{i_N}^{(I_N)}) \\
&= \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} c_{i_1}^{(1)} \cdots c_{i_N}^{(N)} \psi(\mathbf{b}_{i_1}^{(I_1)}, \cdots, \mathbf{b}_{i_N}^{(I_N)}) \\
&= c_{i_1}^{(1)} \cdots c_{i_N}^{(N)} \psi(\mathbf{b}_{i_1}^{(I_1)}, \cdots, \mathbf{b}_{i_N}^{(I_N)}) \qquad [6.20] \\
&= b_{i_1, \cdots, i_N} c_{i_1}^{(1)} \cdots c_{i_N}^{(N)}. \qquad [6.21]
\end{aligned}$$

In this equation, $\mathcal{B} = [b_{i_1, \cdots, i_N}] \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ is called the hypermatrix associated to the multilinear form $\psi$, and its elements $b_{i_1, \cdots, i_N}$ represent the components of $\psi$ with respect to the $N$-linear terms $c_{i_1}^{(1)} \cdots c_{i_N}^{(N)}$. This equation can be interpreted in terms of homogeneous polynomial of degree $N$ in the components $c_{i_n}^{(n)}$, $i_n \in \langle I_n \rangle$, of

4 In this chapter, the vectors of the v.s. $E_n$ will be also denoted by way of bold lower case letters in order to better distinguish them from their scalar components in a basis of $E_n$.

vectors $\mathbf{x}^{(n)}$. It is linear in every component of these vectors. The terms $b_{i_1,\cdots,i_N}$ are the coefficients of the polynomial.

The $N$-linear terms are the elements of the rank-one hypermatrix $\mathcal{C} = \overset{N}{\underset{n=1}{\circ}} \mathbf{c}^{(n)} = [c_{i_1,\cdots,i_N}] = [c_{i_1}^{(1)} \cdots c_{i_N}^{(N)}]$, where $\mathbf{c}^{(n)} = [c_1^{(n)}, \cdots, c_{I_n}^{(n)}]^T$ is the coordinate vector of $\mathbf{x}^{(n)}$ in the basis $\mathcal{B}^{(I_n)}$. Therefore, [6.21] can also be viewed as the contraction of order $N$ of the hypermatrices $\mathcal{B}$ and $\mathcal{C}$; in other words, their inner product:

$$\psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) = \langle \mathcal{B}, \mathcal{C} \rangle. \qquad [6.22]$$

FACT 6.9.– When $E_n = \mathbb{K}^{I_n}$, for $n \in \langle N \rangle$, with the canonical bases $\mathbf{b}_{i_n}^{(I_n)} = \mathbf{e}_{i_n}^{(I_n)}, i_n \in \langle I_n \rangle$, the coordinates $c_{i_n}^{(n)}$ are the components of the vector $\mathbf{x}^{(n)}$ in the canonical basis of $\mathbb{K}^{I_n}$, that is, $\mathbf{x}^{(n)} = \sum_{i_n=1}^{I_n} x_{i_n}^{(n)} \mathbf{e}_{i_n}^{(I_n)}$, and equation [6.20] then becomes:

$$\psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) = x_{i_1}^{(1)} \cdots x_{i_N}^{(N)} \psi(\mathbf{e}_{i_1}^{(I_1)}, \cdots, \mathbf{e}_{i_N}^{(I_N)}). \qquad [6.23]$$

The term $\prod_{n=1}^{N} x_{i_n}^{(n)}$ can be interpreted as an element of the rank-one hypermatrix $\mathcal{X} = \overset{N}{\underset{n=1}{\circ}} \mathbf{x}^{(n)}$. Equation [6.23] expresses the multilinear form $\psi$ in terms of $N$-tuples of the canonical bases transformed by $\psi$. It is to be compared with [6.46] that expresses the tensor $\mathcal{X} = \overset{N}{\underset{n=1}{\otimes}} \mathbf{x}^{(n)}$ in the canonical basis.

### 6.4.2. *Symmetric multilinear forms and symmetric hypermatrices*

As seen in section 2.5.11.3, where $E_n = E, \forall n \in \langle N \rangle$, with $\dim(E) = J$, the multilinear form $\psi \in \mathcal{ML}_N(E, \mathbb{K})$ is symmetric if and only if for any permutation $\pi \in \mathcal{S}_\mathcal{N}$, we have:

$$\psi(\mathbf{x}^{(\pi(1))}, \cdots, \mathbf{x}^{(\pi(N))}) = \psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}). \qquad [6.24]$$

From [6.21], it is inferred that:

$$\psi(\mathbf{x}^{(\pi(1))}, \cdots, \mathbf{x}^{(\pi(N))}) = b_{i_{\pi(1)},\cdots,i_{\pi(N)}} c_{i_{\pi(1)}}^{(\pi(1))} \cdots c_{i_{\pi(N)}}^{(\pi(N))}, \qquad [6.25]$$

and consequently, equation [6.24] implies that:

$$b_{i_{\pi(1)},\cdots,i_{\pi(N)}} = b_{i_1,\cdots,i_N}, \qquad [6.26]$$

that is, the symmetry of the hypermatrix $\mathcal{B}$. This means that the elements of $\mathcal{B}$ are invariant with respect to $N!$ permutations $\pi(.)$ of indices $i_n \in \langle J \rangle, n \in \langle N \rangle$.

For example, for a bilinear form $\psi : E^2 \rightarrow \mathbb{K}$, with $\dim(E) = J$, such that:

$$E^2 \ni (\mathbf{x}, \mathbf{y}) \mapsto \psi(\mathbf{x}, \mathbf{y}) \in \mathbb{K}, \quad [6.27]$$

the expansion of vectors $\mathbf{x} = \sum_{i=1}^{J} x_i \mathbf{u}_i$ and $\mathbf{y} = \sum_{j=1}^{J} y_j \mathbf{u}_j$ in the basis $\{\mathbf{u}\} = \{\mathbf{u}_1, \cdots, \mathbf{u}_J\}$ of $E$ leads to:

$$\psi(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{J} \sum_{j=1}^{J} x_i y_j \psi(\mathbf{u}_i, \mathbf{u}_j). \quad [6.28]$$

or still:

$$\psi(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{J} \sum_{j=1}^{J} b_{ij} x_i y_j = \mathbf{x}_{\mathbf{u}}^T \mathbf{B}_{\mathbf{u}} \mathbf{y}_{\mathbf{u}}, \quad [6.29]$$

where $\mathbf{x}_{\mathbf{u}}$ and $\mathbf{y}_{\mathbf{u}}$ are the coordinate vectors of $\mathbf{x}$ and $\mathbf{y}$ in the basis $\{\mathbf{u}\}$, and the matrix $\mathbf{B}_{\mathbf{u}} \in \mathbb{K}^{J \times J}$ associated with the bilinear form is defined as:

$$b_{ij} = \psi(\mathbf{u}_i, \mathbf{u}_j). \quad [6.30]$$

Equation [6.29] is a special case of equation [6.21] corresponding to $N = 2$. We thus find again equation [4.56], up to a transposition due to the definition of the matrix $\mathbf{B}_{\mathbf{u}}$.

From equation [6.29], we can conclude that the symmetry assumption for $\psi$ implies that $\mathbf{B}_{\mathbf{u}}$ is itself symmetric:

$$b_{ij} = b_{ji} \ , \ \forall i, j \in \langle J \rangle \quad [6.31]$$

$$\mathbf{B}_{\mathbf{u}}^T = \mathbf{B}_{\mathbf{u}}. \quad [6.32]$$

NOTE 6.10.– It should be noted that for $E = \mathbb{K}^J$, the vectors $\mathbf{x}$ and $\mathbf{y}$ belong to the same v.s. as their coordinate vectors $\mathbf{x}_{\mathbf{u}}$ and $\mathbf{y}_{\mathbf{u}}$. By choosing the canonical basis $\{\mathbf{e}_j^{(J)}\}$ for $\{\mathbf{u}\}$, the coordinate vectors are often expressed as the vectors $\mathbf{x}$ and $\mathbf{y}$ themselves.

For a trilinear form $\psi : \mathbb{R}^J \times \mathbb{R}^J \times \mathbb{R}^J \rightarrow \mathbb{R}$, equation [6.21] becomes:

$$\mathbb{R}^J \times \mathbb{R}^J \times \mathbb{R}^J \ni (\mathbf{x}, \mathbf{y}, \mathbf{t}) \mapsto \psi(\mathbf{x}, \mathbf{y}, \mathbf{t}) = b_{ijk} \, x_i y_j t_k,$$

with:

$$\mathbf{x} = \sum_{i=1}^{J} x_i \mathbf{u}_i \ , \ \mathbf{y} = \sum_{j=1}^{J} y_j \mathbf{u}_j \ , \ \mathbf{t} = \sum_{k=1}^{J} t_k \mathbf{u}_k$$

$$b_{ijk} = \psi(\mathbf{u}_i, \mathbf{u}_j, \mathbf{u}_k). \qquad [6.33]$$

The symmetry of the trilinear form $\psi$ implies that of the hypermatrix $\mathcal{B}$, which gives the following equalities:

$$\psi(\mathbf{x}, \mathbf{y}, \mathbf{t}) = \psi(\mathbf{x}, \mathbf{t}, \mathbf{y}) = \psi(\mathbf{y}, \mathbf{x}, \mathbf{t}) = \psi(\mathbf{y}, \mathbf{t}, \mathbf{x}) = \psi(\mathbf{t}, \mathbf{x}, \mathbf{y}) = \psi(\mathbf{t}, \mathbf{y}, \mathbf{x})$$

$$b_{ijk} = b_{ikj} = b_{jik} = b_{jki} = b_{kij} = b_{kji}, \quad \forall\, i, j, k \in \langle J \rangle. \qquad [6.34]$$

Similarly, for a quadrilinear form $\psi$ such that:

$$\mathbb{R}^J \times \mathbb{R}^J \times \mathbb{R}^J \times \mathbb{R}^J \ni (\mathbf{x}, \mathbf{y}, \mathbf{t}, \mathbf{w}) \mapsto \psi(\mathbf{x}, \mathbf{y}, \mathbf{t}, \mathbf{w}) = b_{ijkl}\, x_i y_j t_k w_l,$$

the symmetry of $\psi$ means that:

$$\psi(\mathbf{x}, \mathbf{y}, \mathbf{t}, \mathbf{w}) = \psi(\mathbf{x}, \mathbf{y}, \mathbf{w}, \mathbf{t}) = \psi(\mathbf{x}, \mathbf{t}, \mathbf{y}, \mathbf{w}) = \psi(\mathbf{x}, \mathbf{t}, \mathbf{w}, \mathbf{y}) = \cdots = \psi(\mathbf{w}, \mathbf{t}, \mathbf{y}, \mathbf{x}).$$

and the symmetry of the hypermatrix $\mathcal{B}$ then implies that the coefficient $b_{ijkl}$ is not altered by the 24 permutations of the indices $i, j, k$ and $l$, that is:

$$b_{ijkl} = b_{ijlk} = b_{ikjl} = b_{iklj} = \cdots = b_{lkji}, \quad \forall\, i, j, k, l \in \langle J \rangle.$$

FACT 6.11.– The following observations can be made:

– The set of symmetric hypermatrices of order $N$ form a subspace of the v.s. $\mathbb{K}^{I_1 \times \cdots \times I_N}$ of the hypermatrices.

– A rank-one $N$th-order hypermatrix $\mathcal{B} \in \mathbb{K}^{J \times J \times \cdots \times J}$ is symmetric if it can be written as the outer product of $N$ identical vectors, that is:

$$\mathcal{B} = \underbrace{\mathbf{u} \circ \cdots \circ \mathbf{u}}_{N \text{ terms}} = \mathbf{u}^{\circ N}.$$

– Partial symmetries can be defined with respect to a subset of indices. For example, the cubic hypermatrix $\mathcal{B} = [b_{ijk}]$ is said to be partially symmetric with respect to its first two indices if $b_{ijk} = b_{jik}, \forall\, i, j, k \in \langle J \rangle$. An hypermatrix can always be transformed into a symmetric hypermatrix with respect to a subset of indices. For example, $\mathcal{B} = [b_{ijk}]$ can be transformed into a symmetric hypermatrix $\mathcal{C}$ with respect to its first two indices, by defining $c_{ijk} = \frac{1}{2}(b_{ijk} + b_{jik})$.

## 6.5. Multilinear maps and homogeneous polynomials

Let us consider a multilinear map $\psi : \overset{N}{\underset{n=1}{\times}} E_n \rightarrow F$, with $\dim(E_n) = I_n$ for $n \in \langle N \rangle$ and $\dim(F) = P$. Let $\mathcal{B}^{(I_n)} = \{\mathbf{b}_1^{(I_n)}, \cdots, \mathbf{b}_{I_n}^{(I_n)}\}$ be a basis in $E_n$, $n \in \langle N \rangle$, and $\mathcal{B}^{(P)} = \{\mathbf{b}_p^{(P)}, p \in \langle P \rangle\}$ a basis in $F$. Expanding $\psi(\mathbf{b}_{i_1}^{(I_1)}, \cdots, \mathbf{b}_{i_N}^{(I_N)})$ in the basis $\mathcal{B}^{(P)}$ of $F$ gives:

$$\psi(\mathbf{b}_{i_1}^{(I_1)}, \cdots, \mathbf{b}_{i_N}^{(I_N)}) = \sum_{p=1}^{P} a_{i_1,\cdots,i_N,p}\, \mathbf{b}_p^{(P)} = a_{i_1,\cdots,i_N,p}\, \mathbf{b}_p^{(P)}, \qquad [6.35]$$

and [6.20] becomes:

$$\psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) = c_{i_1}^{(1)} \cdots c_{i_N}^{(N)}\, a_{i_1,\cdots,i_N,p}\, \mathbf{b}_p^{(P)} = d_p\, \mathbf{b}_p^{(P)}. \qquad [6.36]$$

This expansion is now expressed using $P \prod_{n=1}^{N} I_n$ components. We thus recover dimension [2.9] of the v.s. $\mathcal{ML}(E_1, \cdots, E_N; F)$. The $P$ coordinates $\{d_p, p \in \langle P \rangle\}$ of $\psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)})$ in the basis of $F$ are $P$ homogeneous polynomials of degree $N$ in the components of vectors $\mathbf{x}^{(n)}$, and linear in the components of each vector. The coefficients of these polynomials are the coordinates $a_{i_1,\cdots,i_N,p}$ of $\psi(\mathbf{b}_{i_1}^{(I_1)}, \cdots, \mathbf{b}_{i_N}^{(I_N)})$ in the basis of $F$. Defining the hypermatrix $\mathcal{A} = [a_{i_1,\cdots,i_N,p}]$ of order $N+1$ and dimensions $I_1 \times \cdots \times I_N \times P$, the coordinate $d_p$ can be written as the contraction:

$$d_p = c_{i_1}^{(1)} \cdots c_{i_N}^{(N)}\, a_{i_1,\cdots,i_N,p} = \mathcal{A} \overset{1,2,\cdots,N}{\underset{1,2,\cdots,N}{\times}} \mathcal{C},$$

where $\mathcal{C} = [c_{i_1,\cdots,i_N}] = [c_{i_1}^{(1)} \cdots c_{i_N}^{(N)}]$ is defined as for a multilinear form.

## 6.6. Tensor spaces and tensors

### 6.6.1. *Definitions*

Just as a vector is defined as an element of a v.s., also called linear space, a tensor of order $N$ can be defined as an element of a tensor space, that is, a tensor product of $N$ v.s. The linearity of each v.s. induces the multilinearity ($N$-linearity) property of the tensor space. There are several ways to define the concept of tensor product. Here, we choose an approach based on the use of a multilinear map and the choice of a basis.

The tensor product of $N$ $\mathbb{K}$-vector spaces $E_n$ of dimension $I_n$, with $n \in \langle N \rangle$, denoted $\overset{N}{\underset{n=1}{\otimes}} E_n$, can be defined as the image of a multilinear map $\psi$:

$$\overset{N}{\underset{n=1}{\times}} E_n \overset{\psi}{\rightarrow} \overset{N}{\underset{n=1}{\otimes}} E_n \qquad [6.37]$$

satisfying the following condition: considering a basis $\mathcal{B}^{(I_n)} = \{\mathbf{b}_1^{(I_n)}, \cdots, \mathbf{b}_{I_n}^{(I_n)}\}$ for $E_n$, with $n \in \langle N \rangle$, then the set[5]:

$$\mathcal{B} = \left\{ \psi(\mathbf{b}_{i_1}^{(I_1)}, \cdots, \mathbf{b}_{i_N}^{(I_N)}) = \mathbf{b}_{i_1}^{(I_1)} \otimes \cdots \otimes \mathbf{b}_{i_N}^{(I_N)},\ i_n \in \langle I_n \rangle,\ n \in \langle N \rangle \right\} \qquad [6.38]$$

constitutes a basis for the tensor product space $\overset{N}{\underset{n=1}{\otimes}} E_n$.

EXAMPLE 6.12.– Given two $\mathbb{K}$-vector spaces $E$ and $F$, of respective dimensions $I$ and $J$, with the respective bases $\{\mathbf{b}_i^{(I)}, i \in \langle I \rangle\}$ and $\{\mathbf{b}_j^{(J)}, j \in \langle J \rangle\}$, then the $IJ$ vectors $\mathbf{b}_i^{(I)} \otimes \mathbf{b}_j^{(J)}$ constitute a basis of $E \otimes F$.

The tensor product $\overset{N}{\underset{n=1}{\otimes}} \mathbf{u}^{(n)}$ of $N$ vectors $\mathbf{u}^{(n)} \in E_n$ represents an $N$-order tensor, of rank one, called elementary tensor, or pure tensor, or still decomposable tensor. The tensor product space $\overset{N}{\underset{n=1}{\otimes}} E_n$, also referred to as tensor space, consists of the set of linear combinations (see definition in section 2.5.12.2) of $N$-order elementary tensors:

$$\overset{N}{\underset{n=1}{\otimes}} E_n = \text{lc}\{ \overset{N}{\underset{n=1}{\otimes}} \mathbf{u}^{(n)}, \mathbf{u}^{(n)} \in E_n, n \in \langle N \rangle\}. \qquad [6.39]$$

The tensor space $\overset{N}{\underset{n=1}{\otimes}} E_n$ is a v.s. of dimension:

$$\dim(\overset{N}{\underset{n=1}{\otimes}} E_n) = \prod_{n=1}^{N} \dim(E_n) = \prod_{n=1}^{N} I_n. \qquad [6.40]$$

5 Note that two different uses are made of the symbol $\otimes$: (i) for the tensor product $\overset{N}{\underset{n=1}{\otimes}} E_n$ of vector spaces that defines a tensor space and (ii) for designating an element of the tensor space. The symbol $\otimes$ has also been used to denote the Kronecker product in Chapter 5.

## 6.6.2. *Multilinearity and associativity*

### 6.6.2.1. *Multilinearity*

The tensor product satisfies the multilinearity property, that is, for all $\mathbf{x}^{(n)}, \mathbf{y}^{(n)} \in E_n, n \in \langle N \rangle$, and all $\alpha \in \mathbb{K}$, we have:

$$\mathbf{x}^{(1)} \otimes \cdots \otimes (\alpha \mathbf{x}^{(n)} + \mathbf{y}^{(n)}) \otimes \cdots \otimes \mathbf{x}^{(N)} = \alpha\big(\mathbf{x}^{(1)} \otimes \cdots \otimes x^{(n)} \otimes \cdots \otimes \mathbf{x}^{(N)}\big) + \big(\mathbf{x}^{(1)} \otimes \cdots \otimes \mathbf{y}^{(n)} \otimes \cdots \otimes \mathbf{x}^{(N)}\big).$$

EXAMPLE 6.13.– Consider the case $N = 2$ corresponding to bilinear maps. Let $E$ and $F$ be two $\mathbb{K}$-v.s. of respective dimensions $I$ and $J$. Their tensor product $E \otimes F$ consists of elements of the form $\mathbf{u} \otimes \mathbf{v}$, with $\mathbf{u} \in E$ and $\mathbf{v} \in F$, and we have:

$$\forall \mathbf{u} \in E, \forall (\mathbf{v}_1, \mathbf{v}_2) \in F^2 \ , \ \mathbf{u} \otimes (\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{u} \otimes \mathbf{v}_1 + \mathbf{u} \otimes \mathbf{v}_2$$

$$\forall (\mathbf{u}_1, \mathbf{u}_2) \in E^2, \forall \mathbf{v} \in F \ , \ (\mathbf{u}_1 + \mathbf{u}_2) \otimes \mathbf{v} = \mathbf{u}_1 \otimes \mathbf{v} + \mathbf{u}_2 \otimes \mathbf{v}$$

$$\forall (\mathbf{u}, \mathbf{v}) \in E \times F, \forall \lambda \in \mathbb{K} \ , \ \lambda(\mathbf{u} \otimes \mathbf{v}) = \lambda \mathbf{u} \otimes \mathbf{v} = \mathbf{u} \otimes \lambda \mathbf{v}.$$

### 6.6.2.2. *Associativity*

The tensor product is associative. Thus, for three $\mathbb{K}$-v.s. $E, F, G$, we have:

$$(E \otimes F) \otimes G = E \otimes (F \otimes G) = E \otimes F \otimes G.$$

This property directly results from the definition of tensor product and from the property related to the basis of a Cartesian product of v.s. (see section 2.5.13.4).

## 6.6.3. *Tensors and coordinate hypermatrices*

A tensor $\mathcal{X} \in \overset{N}{\underset{n=1}{\otimes}} E_n$ is written in the bases $\mathcal{B}^{(I_n)} = \{\mathbf{b}_1^{(I_n)}, \cdots, \mathbf{b}_{I_n}^{(I_n)}\}$ of $E_n$, with $n \in \langle N \rangle$, as:

$$\mathcal{X} = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} a_{i_1,\cdots,i_N} \mathbf{b}_{i_1}^{(I_1)} \otimes \cdots \otimes \mathbf{b}_{i_N}^{(I_N)}$$

$$= a_{i_1,\cdots,i_N} \overset{N}{\underset{n=1}{\otimes}} \mathbf{b}_{i_n}^{(I_n)} \text{ (with Einstein's convention).} \qquad [6.41]$$

The coefficients $a_{i_1,\cdots,i_N}$ are the coordinates of $\mathcal{X}$ in the basis $\mathcal{B}$ defined in [6.38]. These coordinates can be viewed as the elements of a hypermatrix[6] $\mathcal{A} = [a_{i_1,\cdots,i_N}]$ of

6 Tensors and coordinate hypermatrices relatively to different bases are denoted by calligraphic letters, although they be mathematical objects belonging to different spaces. Similarly, an element (vector or matrix) of a v.s. and its coordinate vector or matrix in a given basis are both denoted by bold lower and upper case letters, respectively.

the space $\mathbb{K}^{I_1 \times \cdots \times I_N}$, that is, a multidimensional array of numbers accessible by way of $N$ indices, each index $i_n$ being associated with a mode of the tensor.

It should be noted that for every set of bases $\{\mathcal{B}^{(I_1)}, \cdots, \mathcal{B}^{(I_N)}\}$ corresponds a different coordinate hypermatrix. In section 6.6.7, we shall see how the coordinate hypermatrix is transformed following a change of basis in every vector space $E_n$ (see equation [6.54]).

It is important to point out the distinction between tensors and hypermatrices. The first ones express mathematical objects in a given basis of the tensor space under consideration, while the second ones correspond to arrays of numbers resulting in practice of measurements with units set during acquisition. This is the case, for instance, in signal processing applications for which the data contained in tensors generally correspond to measurements achieved with units fixed *a priori* or, more specifically, to bases implicitly chosen for the vector spaces associated with each recording mode. In this case, the data tensors are assimilated to coordinate hypermatrices, the bases of the vector spaces underlying the data having been chosen *a priori*.

In summary, formally speaking, a tensor $\mathcal{X}$ can be associated with a set $\{\mathcal{B}^{(I_n)}, n \in \langle N \rangle, \mathcal{A}\}$ comprising the bases of the v.s. $E_n$ and the coordinate hypermatrix in these bases.

### 6.6.4. *Canonical writing of tensors*

If $E_n = \mathbb{K}^{I_n}$, by choosing the canonical basis $\{\mathbf{e}_1^{(I_n)}, \cdots, \mathbf{e}_{I_n}^{(I_n)}\}$, then the tensor $\bigotimes_{n=1}^{N} \mathbf{e}_{i_n}^{(I_n)}$ is an element of the canonical basis of the tensor space $\bigotimes_{n=1}^{N} \mathbb{K}^{I_n}$. It should be noted that in the tensor product $\bigotimes_{n=1}^{N} \mathbf{e}_{i_n}^{(I_n)}$, the symbol $\otimes$ can be replaced by the outer product, as defined by [6.19b][7].

Any tensor $\mathcal{X}$ of the tensor space $\bigotimes_{n=1}^{N} \mathbb{K}^{I_n}$ is then written as a linear combination of these basis tensors, namely:

$$\mathcal{X} = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} x_{i_1,\cdots,i_N} \bigcirc_{n=1}^{N} \mathbf{e}_{i_n}^{(I_n)} \qquad [6.42a]$$

$$= x_{i_1,\cdots,i_N} \mathcal{E}_{i_1 \cdots i_N}^{(I_1 \times \cdots \times I_N)}, \qquad [6.42b]$$

7 If $E_n = \mathbb{K}^{I_n}$, the symbols $\otimes$ and $\circ$ will be indifferently employed to refer to the tensor product and the outer product of vectors of $\mathbb{K}^{I_n}$.

where the coefficients $x_{i_1,\cdots,i_N}$ are the coordinates of $\mathcal{X}$ in the canonical basis, the last equality resulting from the use of Einstein's convention.

It should be noted that the coordinates define a hypermatrix $\mathcal{X}$ which will be generally denoted as the tensor itself. Table 6.1 summarizes the expansions of tensors of orders 1, 2, 3, and $N$, in their respective canonical basis, with Einstein's convention.

| Tensors | Spaces | Canonical expansions | Dimensions |
|---|---|---|---|
| Vectors | $\mathbf{x} \in \mathbb{K}^I$ | $\mathbf{x} = x_i \mathbf{e}_i^{(I)}$ | $I$ |
| Matrices | $\mathbf{X} \in \mathbb{K}^{I \times J}$ | $\mathbf{X} = x_{ij} \mathbf{e}_i^{(I)} \circ \mathbf{e}_j^{(J)}$ | $I \times J$ |
| Third-order tensors | $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$ | $\mathcal{X} = x_{ijk} \mathbf{e}_i^{(I)} \circ \mathbf{e}_j^{(J)} \circ \mathbf{e}_k^{(K)}$ | $I \times J \times K$ |
| $N$-order tensors | $\mathcal{X} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ | $\mathcal{X} = x_{i_1,\cdots,i_N} \overset{N}{\underset{n=1}{\circ}} \mathbf{e}_{i_n}^{(I_n)}$ | $I_1 \times \cdots \times I_N$ |

**Table 6.1.** *Canonical expansions of tensors*

For example, in the particular case $N = 2$, the tensor space $\overset{2}{\underset{n=1}{\otimes}} \mathbb{K}^{I_n} = \mathbb{K}^{I_1} \otimes \mathbb{K}^{I_2} = \mathbb{K}^{I_1 \times I_2}$ is the vector space of matrices of dimensions $I_1 \times I_2$, and in the canonical basis [6.19a], $\mathbf{X} \in \mathbb{K}^{I_1 \times I_2}$ is written as:

$$\mathbf{X} = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} x_{i_1 i_2} \mathbf{e}_{i_1}^{(I_1)} \circ \mathbf{e}_{i_2}^{(I_2)} = x_{i_1 i_2} \mathbf{E}_{i_1 i_2}^{(I_1 \times I_2)}.$$

EXAMPLE 6.14.– For $I_1 = I_2 = 2$, we have:

$$\begin{aligned}\mathbf{X} &= x_{11}\mathbf{E}_{11}^{(2\times 2)} + x_{12}\mathbf{E}_{12}^{(2\times 2)} + x_{21}\mathbf{E}_{21}^{(2\times 2)} + x_{22}\mathbf{E}_{22}^{(2\times 2)} \\ &= x_{11}\begin{bmatrix}1&0\\0&0\end{bmatrix} + x_{12}\begin{bmatrix}0&1\\0&0\end{bmatrix} + x_{21}\begin{bmatrix}0&0\\1&0\end{bmatrix} + x_{22}\begin{bmatrix}0&0\\0&1\end{bmatrix},\end{aligned}$$

with the following coordinate matrix in the canonical basis: $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$.

As already mentioned, the same notation $(\mathbf{X})$ is employed to designate the element of the matrix space $\mathbb{K}^{2\times 2}$ and its coordinate matrix in the canonical basis.

EXAMPLE 6.15.– Let a tensor $\mathcal{X} \in \mathbb{K}^{I\times J\times K}$ with $I = J = K = 2$, such that:

$$\begin{aligned}\mathcal{X} &= x_{111}\mathbf{e}_1^{(I)} \otimes \mathbf{e}_1^{(J)} \otimes \mathbf{e}_1^{(K)} + x_{212}\mathbf{e}_2^{(I)} \otimes \mathbf{e}_1^{(J)} \otimes \mathbf{e}_2^{(K)} \\ &+ x_{221}\mathbf{e}_2^{(I)} \otimes \mathbf{e}_2^{(J)} \otimes \mathbf{e}_1^{(K)} + x_{222}\mathbf{e}_2^{(I)} \otimes \mathbf{e}_2^{(J)} \otimes \mathbf{e}_2^{(K)} \\ &= x_{111}\mathcal{E}_{111}^{(2\times 2\times 2)} + x_{212}\mathcal{E}_{212}^{(2\times 2\times 2)} + x_{221}\mathcal{E}_{221}^{(2\times 2\times 2)} + x_{222}\mathcal{E}_{222}^{(2\times 2\times 2)}. \quad [6.43]\end{aligned}$$

The coordinate hypermatrix in the canonical basis is given by:

$$\mathcal{X} = \begin{bmatrix}1\\0\end{bmatrix} \circ \begin{bmatrix}1\\0\end{bmatrix} \circ \begin{bmatrix}x_{111}\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} \circ \begin{bmatrix}1\\0\end{bmatrix} \circ \begin{bmatrix}0\\x_{212}\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} \circ \begin{bmatrix}0\\1\end{bmatrix} \circ \begin{bmatrix}x_{221}\\x_{222}\end{bmatrix}.$$

It should be noted that, according to [6.43], four coordinates of the tensor $\mathcal{X}$ are zero.

### 6.6.5. *Expansion of the tensor product of $N$ vectors*

Consider a rank-one tensor $\mathcal{X}$ resulting from the tensor product of $N$ vectors $\mathbf{x}^{(n)} \in E_n$, expanded in the bases $\mathcal{B}^{(I_n)} = \{\mathbf{b}_1^{(I_n)}, \cdots, \mathbf{b}_{I_n}^{(I_n)}\}$, for $n \in \langle N \rangle$, that is, $\mathbf{x}^{(n)} = \sum_{i_n=1}^{I_n} c_{i_n}^{(n)} \mathbf{b}_{i_n}^{(I_n)}$. Using the multilinearity property of the tensor product and Einstein's convention, we get:

$$\begin{aligned}\bigotimes_{n=1}^{N} \mathbf{x}^{(n)} &= \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} c_{i_1}^{(1)} \cdots c_{i_N}^{(N)} \bigotimes_{n=1}^{N} \mathbf{b}_{i_n}^{(I_n)} \\ &= c_{i_1}^{(1)} \cdots c_{i_N}^{(N)} \bigotimes_{n=1}^{N} \mathbf{b}_{i_n}^{(I_n)}. \end{aligned} \quad [6.44]$$

The coordinate hypermatrix $\mathcal{A} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ is then defined as $[a_{i_1,\cdots,i_N}] = [c_{i_1}^{(1)} \cdots c_{i_N}^{(N)}]$, with $i_n \in \langle I_n \rangle, n \in \langle N \rangle$.

By comparing [6.44] with [6.41], we can conclude that if any $N$th-order tensor is of the form [6.41], the decomposition of the coefficients $a_{i_1,\cdots,i_N}$ of the coordinate hypermatrix $\mathcal{A}$ in the form of products $c_{i_1}^{(1)} \cdots c_{i_N}^{(N)}$ is characteristic of a tensor of rank one. This hypermatrix is equal to the outer product of the $N$ coordinate vectors $\mathbf{c}^{(n)} = [c_1^{(n)}, \cdots, c_{I_n}^{(n)}]^T$, that is, $\mathcal{A} = [c_{i_1}^{(1)} \cdots c_{i_N}^{(N)}] = \overset{N}{\underset{n=1}{\circ}} \mathbf{c}^{(n)}$, is a rank-one hypermatrix.

FACT 6.16.– In the case where $E_n = \mathbb{K}^{I_n}$ with $\mathbf{x}^{(n)}$ expressed in the canonical basis of $E_n$, namely, $\mathbf{x}^{(n)} = \sum_{i_n=1}^{I_n} x_{i_n}^{(n)} \mathbf{e}_{i_n}^{(I_n)}$, equation [6.44] becomes:

$$\bigotimes_{n=1}^{N} \mathbf{x}^{(n)} = x_{i_1}^{(1)} \cdots x_{i_N}^{(N)} \overset{N}{\underset{n=1}{\circ}} \mathbf{e}_{i_n}^{(I_n)} \quad [6.45]$$

$$= x_{i_1}^{(1)} \cdots x_{i_N}^{(N)} \mathcal{E}_{i_1 \cdots i_N}^{(I_1 \times \cdots \times I_N)}, \quad [6.46]$$

which represents the expansion of the rank-one tensor $\bigotimes_{n=1}^{N} \mathbf{x}^{(n)} = \overset{N}{\underset{n=1}{\circ}} \mathbf{x}^{(n)}$ in the canonical basis $\{\mathcal{E}_{i_1 \cdots i_N}^{(I_1 \times \cdots \times I_N)}\}$.

EXAMPLE 6.17.– Consider the case $N = 2$, with $E_1 = E_2 = \mathbb{K}^3$. Let $\mathbf{x} = \sum_{i=1}^{3} x_i \mathbf{e}_i^{(3)}$ and $\mathbf{y} = \sum_{j=1}^{3} y_j \mathbf{e}_j^{(3)}$. The tensor product $\mathbf{x} \otimes \mathbf{y}$ defines a matrix of rank one:

$$\mathbf{x} \otimes \mathbf{y} = \sum_{i,j=1}^{3} x_i y_j \, \mathbf{e}_i^{(3)} \otimes \mathbf{e}_j^{(3)} = x_i y_j \, \mathbf{E}_{ij}^{(3\times 3)}.$$

and the coordinate matrix in the canonical basis is given by:

$$\begin{bmatrix} x_1y_1 & x_1y_2 & x_1y_3 \\ x_2y_1 & x_2y_2 & x_2y_3 \\ x_3y_1 & x_3y_2 & x_3y_3 \end{bmatrix} \in \mathbb{K}^{3\times 3}.$$

## 6.6.6. *Properties of the tensor product*

### 6.6.6.1. *Symmetry property*

When the $N$ v.s. are identical ($E_n = E$), for $n \in \langle N \rangle$, with $\dim(E) = I$, the elements of the tensor product $\underbrace{E \otimes E \otimes \cdots \otimes E}_{N \text{terms}}$, which is written as $E^{\otimes N}$, are hypercubic tensors of dimensions $I \times \cdots \times I$. As in section 6.4.2 for a symmetric multilinear form, a hypercubic tensor is symmetric if its coordinate hypermatrix $\mathcal{A}$ defined in [6.41] is symmetric, which means that:

$$a_{i_{\pi(1)},\cdots,i_{\pi(N)}} = a_{i_1,\cdots,i_N} \qquad [6.47]$$

for any permutation $\pi \in \mathcal{S}_{\mathcal{N}}$.

EXAMPLE 6.18.– Consider the tensor $\mathcal{X} \in \mathbb{K}^{I\times J\times K}$ with $I = J = K = 2$, expanded in the canonical basis as:

$$\mathcal{X} = \mathcal{E}_{221}^{(2\times 2\times 2)} + \mathcal{E}_{122}^{(2\times 2\times 2)} + \mathcal{E}_{212}^{(2\times 2\times 2)} - \mathcal{E}_{111}^{(2\times 2\times 2)}. \qquad [6.48]$$

The coordinate hypermatrix being such that:

$$x_{121} = x_{211} = x_{112} = x_{222} = 0 \ , \ x_{221} = x_{122} = x_{212} = 1 \text{ and } x_{111} = -1,$$

it satisfies constraints [6.47], which means that $\mathcal{X}$ is a symmetric tensor.

### 6.6.6.2. *Universal property*

The tensor product $\bigotimes_{n=1}^{N} E_n$ satisfies the following universal property.

PROPOSITION 6.19.– *For any multilinear map* $\varphi \in \mathcal{ML}(E_1, \cdots, E_N; F) : \overset{N}{\underset{n=1}{\times}} E_n \xrightarrow{\varphi} F$, *there exists a unique linear map* $f \in \mathcal{L}(\overset{N}{\underset{n=1}{\otimes}} E_n; F) : \overset{N}{\underset{n=1}{\otimes}} E_n \to F$, *such that the multilinear map* $\varphi$ *can be broken down as:*

$$\overset{N}{\underset{n=1}{\times}} E_n \xrightarrow{\psi} \overset{N}{\underset{n=1}{\otimes}} E_n \xrightarrow{f} F, \tag{6.49}$$

*that is,* $\varphi = f \circ \psi$, *which is equivalent to the following diagram, called commutative diagram:*

$$\begin{array}{ccc} \overset{N}{\underset{n=1}{\times}} E_n & \xrightarrow{\varphi} & F \\ \psi \searrow & & \nearrow f \\ & \overset{N}{\underset{n=1}{\otimes}} E_n & \end{array}$$

*So, for* $\mathbf{u}_n \in E_n$, *with* $n \in \langle N \rangle$, *we have:*

$$\varphi(\mathbf{u}_1, \cdots, \mathbf{u}_N) = f \circ \psi(\mathbf{u}_1, \cdots, \mathbf{u}_N) = f(\psi(\mathbf{u}_1, \cdots, \mathbf{u}_N)) = f(\overset{N}{\underset{n=1}{\otimes}} \mathbf{u}_n).$$

It should be noted that since a vector in the tensor space $\overset{N}{\underset{n=1}{\otimes}} E_n$ is expressed as a linear combination of elementary tensors, the linear function $f$ can be determined uniquely by transformation of the vectors of a basis $\mathcal{B}$ such as defined in [6.38], leading to the equation $\varphi(\mathbf{b}_{i_1}^{(I_1)}, \cdots, \mathbf{b}_{i_N}^{(I_N)}) = f(\mathbf{b}_{i_1}^{(I_1)} \otimes \cdots \otimes \mathbf{b}_{i_N}^{(I_N)})$, for all $\mathbf{b}_{i_n}^{(I_n)} \in \mathcal{B}^{(I_n)}, n \in \langle N \rangle$.

This universal property reflects the fact that there exists an isomorphism between the v.s. $\mathcal{ML}(E_1, \cdots, E_N; F)$ of multilinear maps over the Cartesian product $\overset{N}{\underset{n=1}{\times}} E_n$ and the v.s. $\mathcal{L}(\overset{N}{\underset{n=1}{\otimes}} E_n; F)$ of the linear maps built over the tensor product $\overset{N}{\underset{n=1}{\otimes}} E_n$. This property is often used to define the tensor product of two v.s., namely, when $N = 2$ corresponding to bilinear maps. The tensor product of $N$ v.s., with $N > 2$, then constitutes a simple extension to $N$-linear maps (Broomfield; Conrad; Greub 1978; Lang 2002).

FACT 6.20.– (Universal property when $E_n = \mathbb{K}^{I_n}$): When $E_n = \mathbb{K}^{I_n}$, for $n \in \langle N \rangle$, the commutative diagram becomes (de Silva and Lim 2008):

$$\begin{array}{ccc} \overset{N}{\underset{n=1}{\times}} \mathbb{K}^{I_n} & \xrightarrow{\varphi} & \mathbb{K}^{I_1 \times \cdots \times I_N} \\ \psi \searrow & & \nearrow f \\ & \overset{N}{\underset{n=1}{\otimes}} \mathbb{K}^{I_n} & \end{array}$$

In this diagram, the multilinear map $\varphi$ transforms the $N$-tuple $(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) \in \overset{N}{\underset{n=1}{\times}} \mathbb{K}^{I_n}$ into the hypermatrix $\overset{N}{\underset{n=1}{\circ}} \mathbf{x}^{(n)} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$, while the map $\psi$ transforms this $N$-tuple into an elementary tensor $\overset{N}{\underset{n=1}{\otimes}} \mathbf{x}^{(n)}$ of the tensor space $\overset{N}{\underset{n=1}{\otimes}} \mathbb{K}^{I_n}$, and $f$ assigns to this tensor its coordinate hypermatrix:

$$\varphi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) = \overset{N}{\underset{n=1}{\circ}} \mathbf{x}^{(n)} = [x_{i_1}^{(1)} \cdots x_{i_N}^{(N)}] \tag{6.50a}$$

$$\psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) = \overset{N}{\underset{n=1}{\otimes}} \mathbf{x}^{(n)} \tag{6.50b}$$

$$f(\overset{N}{\underset{n=1}{\otimes}} \mathbf{x}^{(n)}) = [x_{i_1}^{(1)} \cdots x_{i_N}^{(N)}] \tag{6.50c}$$

Expanding each vector $\mathbf{x}^{(n)}$ in the canonical basis of $\mathbb{K}^{I_n}$, that is, $\mathbf{x}^{(n)} = \sum_{i_n=1}^{I_n} x_{i_n}^{(n)} \mathbf{e}_{i_n}^{(I_n)}$, the map $\varphi$ breaks down into:

$$\begin{aligned}
\varphi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) &= f \circ \psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}) = f\big(\psi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)})\big) \\
&= f(\overset{N}{\underset{n=1}{\otimes}} \mathbf{x}^{(n)}) = f(x_{i_1}^{(1)} \cdots x_{i_N}^{(N)} \mathcal{E}_{i_1 \cdots i_N}^{(I_1 \times \cdots \times I_N)}) \text{ according to [6.46]}, \\
&= [x_{i_1}^{(1)} \cdots x_{i_N}^{(N)}] \text{ by definition of } f,
\end{aligned}$$

that is, the coordinate hypermatrix of the tensor $\overset{N}{\underset{n=1}{\otimes}} \mathbf{x}^{(n)}$ in the canonical basis. This coordinate hypermatrix being also given by $\overset{N}{\underset{n=1}{\circ}} \mathbf{x}^{(n)}$, the notations $\overset{N}{\underset{n=1}{\otimes}} \mathbf{x}^{(n)}$ and $\overset{N}{\underset{n=1}{\circ}} \mathbf{x}^{(n)}$ will be used indistinctly, which amount to interchanging the tensor and its coordinate hypermatrix, as previously discussed for vectors and matrices. The tensor product of vectors is then equivalent to the outer product of these vectors. We have:

$$\dim(\overset{N}{\underset{n=1}{\times}} \mathbb{K}^{I_n}) = \dim(\overset{N}{\underset{n=1}{\otimes}} \mathbb{K}^{I_n}) = \dim(\mathbb{K}^{I_1 \times \cdots \times I_N}) = \prod_{n=1}^{N} I_n.$$

### 6.6.6.3. *Illustration of the universal property*

To illustrate the universal property, we present below three examples concerning the Khatri–Rao product of two vectors, the multiple vector Khatri–Rao product, and the dot product of two vectors.

For the first two examples, $\varphi$ is a bilinear map from $\mathbb{R}^I \times \mathbb{R}^J$ to $\mathbb{R}^{IJ}$ and a multilinear map from $\overset{N}{\underset{n=1}{\times}} \mathbb{R}^{I_n}$ to $\mathbb{R}^{I_1 \cdots I_N}$, respectively. The image of $f$ is then a vectorized form of a matrix of $\mathbb{R}^{I \times J}$ and of a hypermatrix of the space $\mathbb{R}^{I_1 \times \cdots \times I_N}$, respectively. For the third example, $\varphi$ is a bilinear form from $\mathbb{R}^I \times \mathbb{R}^I$ to $\mathbb{R}$ and $f$ corresponds to the trace of a matrix of $\mathbb{R}^{I \times I}$.

*Khatri–Rao product of two vectors:* Consider the case $N = 2$, with $E_1 = \mathbb{R}^I$ and $E_2 = \mathbb{R}^J$, and the bilinear map corresponding to the Khatri–Rao product of two vectors:

$$\varphi \in \mathcal{BL}(\mathbb{R}^I, \mathbb{R}^J; \mathbb{R}^{IJ}) : \ \mathbb{R}^I \times \mathbb{R}^J \ni (\mathbf{u}, \mathbf{v}) \mapsto \varphi(\mathbf{u}, \mathbf{v}) = \mathbf{u} \diamond \mathbf{v} \in \mathbb{R}^{IJ},$$

where the Khatri–Rao product[8], denoted by $\diamond$, was defined in section 5.4.4 as:

$$\mathbf{u} \diamond \mathbf{v} = [u_1 v_1, \cdots, u_1 v_J, \cdots, u_I v_1, \cdots, u_I v_J]^T.$$

The tensor product, in the sense of the outer product, being defined by:

$$\psi : \mathbb{R}^I \times \mathbb{R}^J \to \mathbb{R}^{I \times J} \ , \ \mathbb{R}^I \times \mathbb{R}^J \ni (\mathbf{u}, \mathbf{v}) \mapsto \psi(\mathbf{u}, \mathbf{v}) = \mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^T \in \mathbb{R}^{I \times J},$$

the linear map $f$ then corresponds to the vectorization operation that transforms a matrix of $\mathbb{R}^{I \times J}$ into a column vector of $\mathbb{R}^{IJ}$ by stacking the row vectors[9] of the matrix one above the other. We have:

$$f(\mathbf{u} \otimes \mathbf{v}) = \text{vec}(\mathbf{u} \otimes \mathbf{v}) = \text{vec}(\mathbf{u}\mathbf{v}^T) = \mathbf{u} \diamond \mathbf{v},$$

and subsequently, the bilinear map corresponding to the Khatri–Rao product of two vectors is equivalent to the composition of the tensor product of these vectors with the vectorization of the matrix which results from the tensor product, or more specifically:

$$\begin{aligned} &\varphi = f \circ \psi \\ &\quad \Updownarrow \\ &\diamond = (\text{vec}) \circ (\otimes). \end{aligned}$$

This means that the computation of the Khatri–Rao product of two vectors can be carried out by applying the vectorization operator to the tensor product (in the sense of outer product) of these two vectors.

8 The Khatri–Rao product will be considered in more detail in Volume 2.

9 In general, the vectorization operation of a matrix is defined as a stack of column vectors instead of row vectors.

EXAMPLE 6.21.– To illustrate this result, consider the case $I = J = 2$. We have:

$$\mathbf{u} \diamond \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \diamond \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 v_1 \\ u_1 v_2 \\ u_2 v_1 \\ u_2 v_2 \end{bmatrix}$$

$$\mathbf{u} \otimes \mathbf{v} = \begin{bmatrix} u_1 v_1 & u_1 v_2 \\ u_2 v_1 & u_2 v_2 \end{bmatrix}$$

$$\text{vec}(\mathbf{u} \otimes \mathbf{v}) = \begin{bmatrix} u_1 v_1 \\ u_1 v_2 \\ u_2 v_1 \\ u_2 v_2 \end{bmatrix} = \mathbf{u} \diamond \mathbf{v}.$$

*Multiple Khatri–Rao product:* The previous example can be generalized to a multilinear map $\varphi \in \mathcal{ML}(\mathbb{R}^{I_1}, \cdots, \mathbb{R}^{I_N}; \mathbb{R}^{I_1 \cdots I_N})$ corresponding to the multiple Khatri–Rao product of $N$ vectors. The tensor product $\psi : \overset{N}{\underset{n=1}{\times}} \mathbb{R}^{I_n} \to \mathbb{R}^{I_1 \times \cdots \times I_N}$ is then an $N$th-order tensor of rank one:

$$\overset{N}{\underset{n=1}{\times}} \mathbb{R}^{I_n} \ni (\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(N)}) \mapsto \varphi(\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(N)}) = \overset{N}{\underset{n=1}{\diamond}} \mathbf{u}^{(n)} \in \mathbb{R}^{I_1 \cdots I_N},$$

$$\overset{N}{\underset{n=1}{\times}} \mathbb{R}^{I_n} \ni (\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(N)}) \mapsto \psi(\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(N)}) = \overset{N}{\underset{n=1}{\otimes}} \mathbf{u}^{(n)} \in \mathbb{R}^{I_1 \times \cdots \times I_N}.$$

The multiple Khatri–Rao product gives a column vector such that the $i$th component is provided by:

$$(\overset{N}{\underset{n=1}{\diamond}} \mathbf{u}^{(n)})_i = \prod_{n=1}^{N} u_{i_n}^{(n)}, \quad \text{with} \quad i = i_N + (i_{N-1} - 1) I_N + \cdots + (i_1 - 1) I_2 \cdots I_N,$$

whereas the tensor product, in the sense of outer product, gives an $N$th-order tensor, of rank one, such that:

$$(\overset{N}{\underset{n=1}{\otimes}} \mathbf{u}^{(n)})_{i_1, \cdots, i_N} = \prod_{n=1}^{N} u_{i_n}^{(n)}. \qquad [6.51]$$

It is therefore easy to deduce the relation:

$$f[\psi(\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(N)})] = f[\overset{N}{\underset{n=1}{\otimes}} \mathbf{u}^{(n)}] = \text{vec}(\overset{N}{\underset{n=1}{\otimes}} \mathbf{u}^{(n)}) = \overset{N}{\underset{n=1}{\diamond}} \mathbf{u}^{(n)} = \varphi(\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(N)}).$$

This means that vectorization of the rank-one tensor, resulting from the outer product of $N$ vectors, is equivalent to computing the Khatri–Rao product of these vectors.

This vectorization operation that transforms a tensor into a vector will be detailed in Volume 2.

*Dot product of two vectors:* Another example is provided by the dot product, which, as we have seen in section 3.4.1.1, is a symmetric bilinear form. For $N = 2$, with $E_1 = E_2 = \mathbb{R}^I$, we have:

$$\varphi \in \mathcal{BL}(\mathbb{R}^I, \mathbb{R}^I; \mathbb{R}) : \ \mathbb{R}^I \times \mathbb{R}^I \ni (\mathbf{u}, \mathbf{v}) \mapsto \varphi(\mathbf{u}, \mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{I} u_i v_i \in \mathbb{R},$$

$$\psi : \mathbb{R}^I \times \mathbb{R}^I \to \mathbb{R}^{I \times I} : \ \mathbb{R}^I \times \mathbb{R}^I \ni (\mathbf{u}, \mathbf{v}) \mapsto \psi(\mathbf{u}, \mathbf{v}) = \mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^T \in \mathbb{R}^{I \times I}.$$

The linear map $f$ such that $\varphi = f \circ \psi$ is then the matrix trace, that is, the sum of the diagonal terms (see section 4.8.1):

$$f(\mathbf{u} \otimes \mathbf{v}) = \text{tr}(\mathbf{u} \otimes \mathbf{v}) = \text{tr}(\mathbf{u}\mathbf{v}^T) = \sum_{i=1}^{I} u_i v_i = \langle \mathbf{u}, \mathbf{v} \rangle = \varphi(\mathbf{u}, \mathbf{v}).$$

### 6.6.7. *Change of basis formula*

Let us consider the tensor $\mathcal{X} \in \bigotimes_{n=1}^{N} E_n$ defined in [6.41], and changes of basis in the v.s. $E_n$ , for $n \in \langle N \rangle$, such that:

$$\mathbf{b}_{i_n}^{(I_n)} = \sum_{j_n=1}^{I_n} p_{i_n, j_n}^{(n)} \mathbf{b}_{j_n}^{'(I_n)} \ , \ i_n \in \langle I_n \rangle. \qquad [6.52]$$

By replacing the basis vectors $\mathbf{b}_{i_n}^{(I_n)}$ by their expressions [6.52] in [6.41], and using the multilinearity property of the tensor product, we get:

$$\begin{aligned}
\mathcal{X} &= \sum_{i_1, \cdots, i_N=1}^{I_1, \cdots, I_N} a_{i_1, \cdots, i_N} \bigotimes_{n=1}^{N} \mathbf{b}_{i_n}^{(I_n)} = a_{i_1, \cdots, i_N} \bigotimes_{n=1}^{N} \mathbf{b}_{i_n}^{(I_n)} \\
&= \sum_{i_1, \cdots, i_N=1}^{I_1, \cdots, I_N} a_{i_1, \cdots, i_N} \bigotimes_{n=1}^{N} \left( \sum_{j_n=1}^{I_n} p_{i_n, j_n}^{(n)} \mathbf{b}_{j_n}^{'(I_n)} \right) \\
&= \sum_{i_1, \cdots, i_N=1}^{I_1, \cdots, I_N} a_{i_1, \cdots, i_N} \sum_{j_1, \cdots, j_N=1}^{I_1, \cdots, I_N} p_{i_1, j_1}^{(1)} \cdots p_{i_N, j_N}^{(N)} \bigotimes_{n=1}^{N} \mathbf{b}_{j_n}^{'(I_n)} \\
&= \sum_{j_1, \cdots, j_N=1}^{I_1, \cdots, I_N} a'_{j_1, \cdots, j_N} \bigotimes_{n=1}^{N} \mathbf{b}_{j_n}^{'(I_n)} = a'_{j_1, \cdots, j_N} \bigotimes_{n=1}^{N} \mathbf{b}_{j_n}^{'(I_n)},
\end{aligned} \qquad [6.53]$$

with:

$$a'_{j_1,\cdots,j_N} = \sum_{i_1,\cdots,i_N=1}^{I_1,\cdots,I_N} a_{i_1,\cdots,i_N} p^{(1)}_{i_1,j_1} \cdots p^{(N)}_{i_N,j_N}$$
$$= a_{i_1,\cdots,i_N} p^{(1)}_{i_1,j_1} \cdots p^{(N)}_{i_N,j_N} \text{ (with Einstein's convention).} \quad [6.54]$$

The hypermatrix $\mathcal{A}' = [a'_{j_1,\cdots,j_N}] \in \mathbb{K}^{I_1\times\cdots\times I_N}$ contains the coordinates of the tensor $\mathcal{X}$ in the new bases $\mathcal{B}'^{(I_n)} = \{\mathbf{b}'^{(I_n)}_1, \cdots, \mathbf{b}'^{(I_n)}_{I_n}\}$, for $n \in \langle N\rangle$. Let us define the matrices of changes of basis $\mathbf{P}^{(n)} = [p^{(n)}_{i_n,j_n}] \in \mathbb{K}^{I_n\times I_n}$. By comparison with [6.7] and [6.8], the equation of definition [6.54] of the hypermatrix $\mathcal{A}'$ can be written in the two following compact equivalent forms:

$$\mathcal{A}' = \mathcal{A} \times_1 \mathbf{P}^{(1)} \times_2 \cdots \times_N \mathbf{P}^{(N)} = \mathcal{A} \overset{N}{\underset{n=1}{\times}} \mathbf{P}^{(n)} \quad [6.55a]$$
$$= (\mathbf{P}^{(1)}, \cdots, \mathbf{P}^{(N)}) \,.\, \mathcal{A} \quad [6.55b]$$

These two equations express the fact that a linear transformation of matrix $\mathbf{P}^{(n)}$ is applied to each v.s. $E_n$, for $n \in \langle N\rangle$. They involve $N$ hypermatrix-matrix multiplications, that is, the $n$-mode products of the hypermatrix $\mathcal{A}$ with $\mathbf{P}^{(n)} \in \mathbb{K}^{I_n\times I_n}$, $n \in \langle N\rangle$ or, equivalently, an $N$-linear (or multilinear) multiplication.

The term multilinear multiplication comes from the fact that, for all matrices $P^{(n)}, Q^{(n)} \in \mathbb{K}^{I_n\times I_n}$, and for all $\alpha, \beta \in \mathbb{K}$, we have (de Silva and Lim 2008):

$$(\mathbf{P}^{(1)}, \cdots, \alpha\mathbf{P}^{(n)} + \beta\mathbf{Q}^{(n)}, \cdots, \mathbf{P}^{(N)}) \,.\, \mathcal{A} = \alpha(\mathbf{P}^{(1)}, \cdots, \mathbf{P}^{(n)}, \cdots, \mathbf{P}^{(N)}) \,.\, \mathcal{A}$$
$$+ \beta(\mathbf{P}^{(1)}, \cdots, \mathbf{Q}^{(n)}, \cdots, \mathbf{P}^{(N)}) \,.\, \mathcal{A}.$$

*Matrix case:* Let $\mathbf{A} \in \mathbb{R}^{I_1\times I_2}$ be the coordinate matrix and $(\mathbf{Q}, \mathbf{P}) = (\mathbf{P}^{(1)}, \mathbf{P}^{(2)})$ the matrices of changes of basis, with $\mathbf{Q} \in \mathbb{K}^{I_1\times I_1}$ and $\mathbf{P} \in \mathbb{K}^{I_2\times I_2}$. After the changes of basis, relations [6.54] and [6.55a] are then written as:

$$a'_{j_1 j_2} = a_{i_1 i_2} q_{i_1 j_1} p_{i_2 j_2} \ , \ j_n \in \langle I_n\rangle, \ n = 1, 2 \quad [6.56a]$$
$$\mathbf{A}' = \mathbf{A} \times_1 \mathbf{Q} \times_2 \mathbf{P} = \mathbf{Q}^T \mathbf{A} \mathbf{P}. \quad [6.56b]$$

This transformation [6.56b] of the coordinate matrix is to be compared with formula [4.53] highlighted for the matrix associated with a bilinear form after a change of bases.

## 6.7. Tensor rank and tensor decompositions

### 6.7.1. *Matrix rank*

In the matrix case, a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ is of rank one if and only if it can be written as the outer product of two non-zero vectors:

$$\begin{aligned} &\mathbf{A} = \mathbf{u} \circ \mathbf{v} \ , \ \mathbf{u} \in \mathbb{K}^I, \mathbf{v} \in \mathbb{K}^J \\ &\quad \Updownarrow \\ &a_{ij} = u_i v_j. \end{aligned}$$

The rank of a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$, denoted $\mathrm{r}_{\mathbf{A}}$ or $\mathrm{r}(\mathbf{A})$, is defined as the smallest integer $R$ such that $\mathbf{A}$ be written as the sum of $R$ matrices of rank one:

$$\mathrm{r}(\mathbf{A}) = \min\{R : \mathbf{A} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r, \mathbf{u}_r \in \mathbb{K}^I, \mathbf{v}_r \in \mathbb{K}^J\}. \quad [6.57]$$

The writing of $\mathbf{A}$ in [6.57] corresponds to a dyadic decomposition of $\mathbf{A}$. By defining the matrices $\mathbf{U} \in \mathbb{K}^{I \times R}$ and $\mathbf{V} \in \mathbb{K}^{J \times R}$ whose columns are the vectors $\mathbf{u}_r$ and $\mathbf{v}_r$, respectively, the dyadic decomposition of $\mathbf{A}$ can be written as the following matrix product:

$$\mathbf{A} = \mathbf{U}\mathbf{V}^T, \quad [6.58]$$

where

$$\mathbf{U} = \left[\mathbf{u}_1 \ \mathbf{u}_2 \cdots \mathbf{u}_R\right] \ , \ \mathbf{V} = \left[\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_R\right].$$

In the case of a tensor of order higher than two, there are several ways to define the rank. In the following, we consider the definition of tensor rank from its decomposition into a linear combination of elementary tensors, that is, a sum of rank-one tensors (Harshman 1970; Hitchcock 1927; Kruskal 1977), which constitutes a generalization of the matrix rank to orders higher than two. The multilinear rank of an $N$-order tensor, related to the Tucker decomposition (Tucker 1966) and to the $N$ modal matricizations of the tensor, will be introduced in Volume 2.

### 6.7.2. *Hypermatrix rank*

Consider a hypermatrix $\mathcal{A} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ of order $N$. This hypermatrix is of rank one if and only if there exists $N$ non-zero vectors $\mathbf{u}^{(n)} \in \mathbb{K}^{I_n}, n \in \langle N \rangle$, such that $\mathcal{A} = \overset{N}{\underset{n=1}{\circ}} \mathbf{u}^{(n)}$, that is, $\mathcal{A}$ is written as the outer product of $N$ vectors.

The rank of $\mathcal{A} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$, denoted by $\mathrm{r}_{\mathcal{A}}$ or $\mathrm{r}(\mathcal{A})$, is defined as the smallest integer $R$ such that $\mathcal{A}$ is written as a sum of $R$ rank-one hypermatrices (Lim 2013):

$$\mathrm{r}(\mathcal{A}) = \min\{R : \mathcal{A} = \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(N)}, \mathbf{u}_r^{(n)} \in \mathbb{K}^{I_n}, n \in \langle N \rangle\}. \quad [6.59]$$

This equation defines a canonical polyadic decomposition (CPD) of the hypermatrix $\mathcal{A}$ (Hitchcock 1927). This decomposition is also called PARAFAC [for parallel factors (Harshman 1970)] or CANDECOMP [for canonical decomposition (Carroll and Chang 1970)]. The CPD introduced in [6.59] involves $N$ factor matrices $\mathbf{U}^{(n)} = [\mathbf{u}_1^{(n)} \cdots \mathbf{u}_R^{(n)}] \in \mathbb{K}^{I_n \times R}$, $n \in \langle N \rangle$.

This decomposition will be studied in more detail in Volume 2 where other tensor decompositions will be presented.

### 6.7.3. *Symmetric rank of a hypermatrix*

In the case of a symmetric hypermatrix $\mathcal{A} \in \mathbb{K}^{I \times \cdots \times I}$ of order $N$, the symmetric rank of $\mathcal{A}$, denoted by $\mathrm{r}_S(\mathcal{A})$, is defined as the smallest integer $R_S$ such that $\mathcal{A}$ can be written as a sum of $R_S$ symmetric rank-one hypermatrices:

$$\mathrm{r}_S(\mathcal{A}) = \min\{R_S : \mathcal{A} = \sum_{r=1}^{R_S} \mathbf{u}_r \circ \cdots \circ \mathbf{u}_r = \sum_{r=1}^{R_S} \mathbf{u}_r^{\circ N}, \mathbf{u}_r \in \mathbb{K}^I\}. \quad [6.60]$$

For a complex symmetric hypermatrix, the symmetric rank $R_S$ and the tensor rank $R$, as defined in [6.59], are such that:

$$R \leq R_S. \quad [6.61]$$

In Comon *et al.* (2008), it is shown that $R = R_S$ generically (i.e. with probability one), when $R_s \leq I$ and $N$ is large enough compared to $I$.

### 6.7.4. *Comparative properties of hypermatrices and matrices*

Several properties differentiate hypermatrices of order $N \geq 3$ from matrices.

– The rank of a hypermatrix $\mathcal{A} \in \mathbb{K}^{I_1 \times \cdots \times I_N}$ of order higher than two depends on the field $\mathbb{K}$ over which its elements are defined. Indeed, since $\mathbb{R}^{I_1 \times \cdots \times I_N} \subseteq \mathbb{C}^{I_1 \times \cdots \times I_N}$ for $N \geq 3$, the tensor ranks over $\mathbb{K} = \mathbb{R}$ and over $\mathbb{K} = \mathbb{C}$, respectively, denoted by $\mathrm{r}_{\mathbb{R}}$ and $\mathrm{r}_{\mathbb{C}}$, are such that $\mathrm{r}_{\mathbb{C}}(\mathcal{A}) \leq \mathrm{r}_{\mathbb{R}}(\mathcal{A})$. This is not the case for matrices for which we have $\mathrm{r}_{\mathbb{C}}(\mathbf{A}) = \mathrm{r}_{\mathbb{R}}(\mathbf{A})$.

EXAMPLE 6.22.– (Comon *et al.* 2008): Let the hypermatrix $\mathcal{A} \in \mathbb{R}^{2\times 2\times 2}$ be real symmetric, of symmetric rank three over $\mathbb{R}$, defined as:

$$\mathcal{A} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}^{\circ 3} + \frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix}^{\circ 3} - 2\begin{bmatrix}1\\0\end{bmatrix}^{\circ 3}. \qquad [6.62]$$

It is easy to verify that the elements of this hypermatrix are given by:

$$a_{121} = a_{211} = a_{112} = 0\ ,\ a_{221} = a_{122} = a_{212} = 1\ ,\ a_{111} = -1 \text{ and } a_{222} = 0,$$

that is, this hypermatrix is that of the coordinates of the symmetric tensor $\mathcal{X}$ of example [6.48].

In $\mathbb{C}$, this hypermatrix can be written as:

$$\mathcal{A} = \frac{i}{2}\begin{bmatrix}-i\\1\end{bmatrix}^{\circ 3} - \frac{i}{2}\begin{bmatrix}i\\1\end{bmatrix}^{\circ 3}\ ,\ \text{where } i^2 = -1. \qquad [6.63]$$

Indeed, exploiting the results [6.14]–[6.16] and [6.17] and [6.18], we have:

$$a_{ijk} = \frac{i}{2}c_{ijk} - \frac{i}{2}b_{ijk},$$

which gives the above coefficients $a_{ijk}$.

Subsequently, $\mathcal{A}$ is of symmetric rank two over $\mathbb{C}$, and therefore, the symmetric ranks of $\mathcal{A}$ over $\mathbb{R}$ and $\mathbb{C}$ are different.

– The rank of a real hypermatrix $\mathcal{A} \in \mathbb{R}^{I_1\times\cdots\times I_N}$ may be greater than the largest of its dimensions, that is, one can have $R > \max\{I_1, \cdots, I_N\}$, whereas for a matrix $\mathbf{A} \in \mathbb{K}^{I\times J}$, the rank is at most equal to the smallest of its dimensions, that is, $R \leq \min(I, J)$.

– While a random matrix is of full rank with probability one, a real random hypermatrix of order higher than two can have several ranks called typical ranks. So, Kruskal (1989) showed that a third-order real hypermatrix of dimensions $2 \times 2 \times 2$, has two typical ranks two and three. For complex hypermatrices, there is only one typical rank, called a generic rank, with probability one.

– A fundamental question concerns the determination of an upper bound for the rank of a hypermatrix. Thus, for a third-order hypermatrix $\mathcal{A} \in \mathbb{K}^{I\times J\times K}$, the rank is bounded by (Kruskal 1989):

$$R \leq \min\{IJ, JK, KI\}. \qquad [6.64]$$

As we have just seen, there are several notions of rank for a hypermatrix, and therefore for a tensor. One can define the non-negative rank of a non-negative real hypermatrix $\mathcal{A}$, that is, a hypermatrix whose all elements are positive or zero. This rank, denoted $\mathrm{r}_+(\mathcal{A})$, is the minimal number of rank-one non-negative terms that form the CPD. Similarly to the matrix case, we have $\mathrm{r}(\mathcal{A}) \leq \mathrm{r}_+(\mathcal{A})$ (Comon 2014). For a review of the main results on the different notions of rank of a hypermatrix, refer to Sidiropoulos *et al.* (2017).

### 6.7.5. *CPD and dimensionality reduction*

In the era of big data, complexity reduction is a major issue for storage, visualization, representation, analysis and classification of massive data, or megadata such as, for example, in recommendation systems, or for processing medical, astronomical, climatic observation, or social networks databases. This complexity reduction can be obtained in two different ways, either at the level of the data representation model or through the use of numerically efficient processing methods.

The previously introduced CPD plays a fundamental role in reducing the dimensionality of a hypermatrix. Thus, for a hypermatrix $\mathcal{A} \in \mathbb{K}^{I\times\cdots\times I}$, of order $N$ and of the same dimension for each mode ($I_n = I, \forall n \in \langle N\rangle$), the number of elements and, therefore, the memory needed to store all the data contained in $\mathcal{A}$, is $I^N$, while for CPD [6.59] the amount of data to be stored, that is, the elements of the $N$ factor matrices $\mathbf{U}^{(n)} \in \mathbb{K}^{I\times R}$, $n \in \langle N\rangle$, is equal to $NRI$. This constitutes a very significant dimensionality reduction for large values of $N$ and $I$, that is, for very large hypermatrices.

In the case of a third-order hypermatrix $\mathcal{X} \in \mathbb{K}^{I\times J\times K}$ of rank $R$, a CPD is written as:

$$x_{ijk} = \sum_{r=1}^{R} a_{ir}b_{jr}c_{kr} \qquad [6.65a]$$

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{A}_{.r} \circ \mathbf{B}_{.r} \circ \mathbf{C}_{.r} \qquad [6.65b]$$

$$= \mathcal{I}_{3,R}\times_1\mathbf{A}\times_2\mathbf{B}\times_3\mathbf{C}, \qquad [6.65c]$$

where $\mathbf{A}_{.r} \in \mathbb{K}^I, \mathbf{B}_{.r} \in \mathbb{K}^J, \mathbf{C}_{.r} \in \mathbb{K}^K$ are the $r$th column vectors of the factor matrices

$$\mathbf{A} = [\mathbf{A}_{.1}, \cdots, \mathbf{A}_{.R}] \ , \ \mathbf{B} = [\mathbf{B}_{.1}, \cdots, \mathbf{B}_{.R}] \ , \ \mathbf{C} = [\mathbf{C}_{.1}, \cdots, \mathbf{C}_{.R}].$$

This PARAFAC model, of rank $R$, is denoted by $\|\mathbf{A}, \mathbf{B}, \mathbf{C}; R\|$.

NOTE 6.23.– The form [6.65c] based on $n$-mode products, with $n \in \{1, 2.3\}$, can be obtained from the scalar from [6.65a] in employing the generalized Kronecker delta $\delta_{r_1,r_2,r_3} = \begin{cases} 1 & \text{if} \quad r_1 = r_2 = r_3 = r \\ 0 & \text{otherwise} \end{cases}$, which gives:

$$x_{ijk} = \sum_{r_1,r_2,r_3=1}^{R} \delta_{r_1,r_2,r_3}a_{i,r_1}b_{j,r_2}c_{k,r_3}. \qquad [6.66]$$

Taking into account the definition of the third-order identity hypermatrix $\mathcal{I}_{3,R} = [\delta_{r_1,r_2,r_3}]$, with $r_1, r_2, r_3 \in \langle R \rangle$, it is easy to verify that this equation is identical to the scalar writing [6.65c] of the hypermatrix, the summations over $r_1$, $r_2$, and $r_3$ corresponding to mode-1, -2, and -3 products of the identity hypermatrix with the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively.

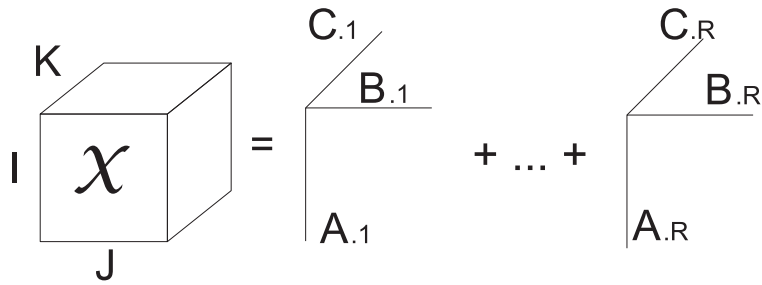



**Figure 6.1.** *Third-order PARAFAC model*

Figure 6.1 illustrates the PARAFAC model $\|\mathbf{A}, \mathbf{B}, \mathbf{C}; R\|$ in the form [6.65b] of a sum of $R$ rank-one hypermatrices, each one being equal to the outer product of three column vectors $(\mathbf{A}_{.r}, \mathbf{B}_{.r}, \mathbf{C}_{.r})$, with $r \in \langle R \rangle$.

A fundamental difference between matrices and hypermatrices concerns the uniqueness properties of their dyadic and polyadic decomposition, respectively.

Indeed, using the identity $(\mathbf{V}\mathbf{\Lambda}^{-T})^T = \mathbf{\Lambda}^{-1}\mathbf{V}^T$, it is easy to verify that the dyadic decomposition [6.58] of $\mathbf{A}$ is unique up to a non-singular matrix (namely, invertible), $\mathbf{\Lambda} \in \mathbb{K}^{R \times R}$, because:

$$\mathbf{A} = (\mathbf{U}\mathbf{\Lambda})(\mathbf{V}\mathbf{\Lambda}^{-T})^T = \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}^{-1}\mathbf{V}^T = \mathbf{U}\mathbf{V}^T.$$

In Volume 2, we shall see that the CPD of a hypermatrix of order higher than two is unique up to simple column permutation and scaling ambiguities in the factor matrices. One then speaks of essential uniqueness. We shall also see that such uniqueness can be ensured for a CPD model under (sufficient but not necessary) conditions that are not very constraining.

It should be noted that when $\mathbf{A}$ and $\mathcal{A}$ are of rank one, that is, $\mathbf{A} = \mathbf{u} \circ \mathbf{v}$ and $\mathcal{A} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$, then ambiguities are scalar, and all matrices and hypermatrices equivalent to $\mathbf{A}$ and $\mathcal{A}$ can be written as: $(\alpha\mathbf{u}) \circ (\frac{1}{\alpha}\mathbf{v})$ and $(\alpha\mathbf{a}) \circ (\beta\mathbf{b}) \circ (\frac{1}{\alpha\beta}\mathbf{c})$, respectively, with $\alpha, \beta \in \mathbb{K}$.

Determining the parameters of a decomposition from a data hypermatrix is a very important problem for tensor-based applications. In Volume 2, we shall present a standard parameter estimation method for solving this so-called inverse problem, in the case of a CPD model, namely, the alternating least-squares (ALS) algorithm.

The HOSVD method (for high-order singular value decomposition, de Lathauwer *et al.* 2000), linked to the Tucker (1966) model, will also be presented. This method is very useful for determining a low-rank approximation of a data hypermatrix in the form of a truncated HOSVD (THOSVD). It generalizes to hypermatrices of order higher than two, the truncated SVD which is utilized for low-rank matrix approximation (Eckart and Young 1936). Nonetheless, it is important to note that, unlike truncated SVD, THOSVD is not optimal. This results from the fact that, unlike the matrix case, deflationary-type methods which consist of calculating and successively subtracting rank-one approximations do not necessarily induce a decrease in the rank of a higher-order hypermatrix (Stegeman and Comon 2010).

The tensor approximation problem using a lower-rank tensor is a fundamental problem from the point of view of complexity reduction. In contrast to the matrix case, for tensors of order higher than two, this is an ill-posed problem due to the fact that the set of tensors of rank equal to $R$ is not generally closed. This means that a sequence of tensors of rank $R$ can converge to a tensor of higher rank, as illustrated by the example below.

EXAMPLE 6.24.– Let the sequence of rank-two tensors defined by Sidiropoulos *et al.* (2017):

$$\mathcal{X}_n = n(\mathbf{u} + \frac{1}{n}\mathbf{v}) \circ (\mathbf{u} + \frac{1}{n}\mathbf{v}) \circ (\mathbf{u} + \frac{1}{n}\mathbf{v}) - n\mathbf{u} \circ \mathbf{u} \circ \mathbf{u},$$

with $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$. Using the associativity property of the outer product over addition, $\mathcal{X}_n$ can be developed as:

$$\begin{aligned}\mathcal{X}_n &= \mathbf{u} \circ \mathbf{u} \circ \mathbf{v} + \mathbf{u} \circ \mathbf{v} \circ \mathbf{u} + \mathbf{v} \circ \mathbf{u} \circ \mathbf{u} + \frac{1}{n}\Big(\mathbf{v} \circ \mathbf{v} \circ \mathbf{u} \\ &\quad + \mathbf{v} \circ \mathbf{u} \circ \mathbf{v} + \mathbf{u} \circ \mathbf{v} \circ \mathbf{v}\Big) + \frac{1}{n^2}\mathbf{v} \circ \mathbf{v} \circ \mathbf{v} \\ &= \mathcal{X} + \text{terms in } 0(\frac{1}{n}).\end{aligned}$$

When $n \to \infty$, the sequence $\mathcal{X}_n$ tends toward the tensor $\mathcal{X}$ of higher rank than $\mathcal{X}_n$, equal to three, defined by:

$$\mathcal{X} = \mathbf{u} \circ \mathbf{u} \circ \mathbf{v} + \mathbf{u} \circ \mathbf{v} \circ \mathbf{u} + \mathbf{v} \circ \mathbf{u} \circ \mathbf{u}.$$

### 6.7.6. *Tensor rank*

In the case of a tensor $\mathcal{X} \in \bigotimes_{n=1}^{N} E_n$, the rank of $\mathcal{X}$, denoted by $\mathrm{r}(\mathcal{X})$, is defined as the minimum number $R$ of elementary tensors whose sum can represent $\mathcal{X}$:

$$\mathrm{r}(\mathcal{X}) = \min\{R : \mathcal{X} = \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \otimes \cdots \otimes \mathbf{u}_r^{(N)}, \mathbf{u}_r^{(n)} \in E_n, n \in \langle N \rangle\} \qquad [6.67]$$

where $\mathbf{u}_r^{(1)} \otimes \cdots \otimes \mathbf{u}_r^{(N)}$ is to be considered as a tensor product. It should be noted that when one chooses the basis $\mathcal{B}$ defined in [6.38] for the tensor space $\overset{N}{\underset{n=1}{\otimes}} E_n$, then $\text{rank}(\mathcal{A}) = \text{rank}(\mathcal{X})$, which means that the tensor has the same rank as the hypermatrix of its coordinates in the basis $\mathcal{B}$.

The determination of the rank of a tensor of order higher than two and, thus, the computation of an exact PARAFAC decomposition are difficult problems because they are ill-posed (Hastad 1990; Hillar and Lim 2013).

## 6.8. Eigenvalues and singular values of a hypermatrix

In section 4.16, we presented the concept of matrix eigenvalue/eigenvector, and in Volume 2, we shall present that of singular value/singular vector, as well as the associated eigenvalue and singular value decompositions (EVD and SVD).

Here, we briefly present the notions of eigenvalue and singular value of a third-order tensor, pointing out the link to the matrix case. The notion of tensor eigenvalue was independently introduced by Lim (2005) and Qi (2005).

In Table 6.2, we summarize the equations for calculating eigenvalues and singular values for a matrix and a third-order tensor.

The pair $(\lambda, \mathbf{u})$ is called an eigenvalue–eigenvector pair, or eigenpair, of matrix $\mathbf{A}$, or of tensor $\mathcal{A}$. It is important to note that for matrices, if $(\lambda, \mathbf{u})$ is an eigenpair, then $(\lambda, \alpha\mathbf{u})$ is also an eigenpair for any $\alpha \neq 0$. It is said that the eigenvectors of a matrix are invariant up to a scaling factor, which justifies the normalization of eigenvectors such that $\|\mathbf{u}\|_2^2 = 1$.

In the matrix case, the triplet $(\sigma, \mathbf{u}, \mathbf{v})$ represents a singular value $(\sigma)$ and left- $(\mathbf{u})$ and right- $(\mathbf{v})$ singular vectors, while in the tensor case, the quadruplet $(\sigma, \mathbf{u}, \mathbf{v}, \mathbf{w})$ represents a singular value and mode -1, -2, and -3 singular vectors.

NOTE 6.25.– Note that the right singular vectors of $\mathbf{A}$ are the eigenvectors of the symmetric matrix $\mathbf{A}^T\mathbf{A}$, while the left singular vectors are the eigenvectors of the symmetric matrix $\mathbf{A}\mathbf{A}^T$.

In section 4.16.10, we saw that for a symmetric matrix $\mathbf{A}$, the eigenpair $(\lambda, \mathbf{u})$ can be obtained by maximizing the quadratic form $\mathbf{u}^T\mathbf{A}\mathbf{u}$ under the constraint $\|\mathbf{u}\|_2^2 = 1$, which amounts to optimizing the Lagrangian $L(\mathbf{u}, \lambda) = \mathbf{u}^T\mathbf{A}\mathbf{u} - \lambda(\|\mathbf{u}\|_2^2 - 1)$. The pair $(\lambda, \mathbf{u})$ is called $l_2$-eigenvalue and $l_2$-eigenvector.

Similarly, the triplet $(\sigma, \mathbf{u}, \mathbf{v})$ can be obtained by maximizing the bilinear form $\mathbf{u}^T\mathbf{A}\mathbf{v}$ under the constraints $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1$, or equivalently by optimizing the Lagrangian $L(\mathbf{u}, \mathbf{v}, \lambda) = \mathbf{u}^T\mathbf{A}\mathbf{v} - \lambda(\|\mathbf{u}\|_2\|\mathbf{v}\|_2 - 1)$.

| |
|---|
| $\mathbf{A} \in \mathbb{R}^{I \times I}\ ,\ \mathbf{u} \in \mathbb{R}^{I}, \mathbf{u} \neq \mathbf{0}_I$ |
| Eigenvalues $\lambda$ |
| $\sum_{j=1}^{I} a_{ij}u_j = \lambda u_i\ ,\ i \in \langle I \rangle\ \Leftrightarrow\ \mathbf{Au} = \lambda\mathbf{u}$ |
| $\mathbf{A} \in \mathbb{R}^{I \times J}\ ,\ \mathbf{u} \in \mathbb{R}^{I}\ ,\ \mathbf{v} \in \mathbb{R}^{J}$ |
| Singular values $\sigma$ |
| $\sum_{j=1}^{J} a_{ij}v_j = \sigma u_i\ ,\ i \in \langle I \rangle\ \Leftrightarrow\ \mathbf{Av} = \sigma\mathbf{u}\ \Leftrightarrow\ \mathbf{A}^T\mathbf{Av} = \sigma^2\mathbf{v}$<br>$\sum_{i=1}^{I} a_{ij}u_i = \sigma v_j\ ,\ j \in \langle J \rangle\ \Leftrightarrow\ \mathbf{A}^T\mathbf{u} = \sigma\mathbf{v}\ \Leftrightarrow\ \mathbf{AA}^T\mathbf{u} = \sigma^2\mathbf{u}$ |
| $\mathcal{A} \in \mathbb{R}^{I \times I \times I}\ ,\ \mathbf{u} \in \mathbb{R}^{I}\ ,\ \mathbf{u} \neq \mathbf{0}_I$ |
| Eigenvalues $\lambda$ |
| $\sum_{i,j=1}^{I} a_{ijk}u_iu_j = \lambda u_k\ ,\ k \in \langle I \rangle$ |
| $\mathcal{A} \in \mathbb{R}^{I \times J \times K}\ ,\ \mathbf{u} \in \mathbb{R}^{I}\ ,\ \mathbf{v} \in \mathbb{R}^{J}\ ,\ \mathbf{w} \in \mathbb{R}^{K}$ |
| Singular values $\sigma$ |
| $\sum_{j,k=1}^{J,K} a_{ijk}v_jw_k = \sigma u_i\ ,\ i \in \langle I \rangle$<br>$\sum_{i,k=1}^{I,K} a_{ijk}u_iw_k = \sigma v_j\ ,\ j \in \langle J \rangle$<br>$\sum_{i,j=1}^{I,J} a_{ijk}u_iv_j = \sigma w_k\ ,\ k \in \langle K \rangle$ |

**Table 6.2.** *Matrix and third-order tensor eigenvalues and singular values*

It is then easy to derive the equations in Table 6.2 by writing the Karush–Kuhn–Tucker optimality conditions, or more specifically, by cancelling the partial derivatives of the Lagrangian with respect to every variable to be optimized (Lim 2005).

For a symmetric third-order tensor, the equations for the computation of eigenvalues given in Table 6.2 can be obtained by optimizing the Lagrangian $L(\mathbf{u}, \lambda) = \sum_{i,j,k=1}^{I} a_{ijk}u_iu_ju_k - \lambda(\|\mathbf{u}\|_2^2 - 1)$. It can be observed that the invariance property of the eigenvectors up to a scaling factor is no longer satisfied with the constraint $\|\mathbf{u}\|_2^2 = 1$.

To satisfy this invariance property, it is necessary to consider the norm $l_3$ for the constraint $\|\mathbf{u}\|_3 = (\sum_{i=1}^{I} |u_i|^3)^{1/3} = 1$, which gives the following computation equations:

$$\sum_{i,j=1}^{I} a_{ijk}u_iu_j = \lambda u_k^2\ ,\ k \in \langle I \rangle, \qquad [6.68]$$

and the pair $(\lambda, \mathbf{u})$ is then called $l_3$-eigenvalue and $l_3$-eigenvector.

Similarly, for singular values, the Lagrangian to optimize is written as $L(\mathbf{u}, \mathbf{v}, \mathbf{w}, \lambda) = \sum_{i,j,k=1}^{I,J,K} a_{ijk}u_iv_jw_k - \lambda(\|\mathbf{u}\|_3\|\mathbf{v}\|_3\|\mathbf{w}\|_3 - 1)$, and the equations

of Table 6.2 become:

$$\sum_{j,k=1}^{J,K} a_{ijk} v_j w_k = \sigma u_i^2 \ , \ i \in \langle I \rangle \qquad [6.69a]$$

$$\sum_{i,k=1}^{I,K} a_{ijk} u_i w_k = \sigma v_j^2 \ , \ j \in \langle J \rangle \qquad [6.69b]$$

$$\sum_{i,j=1}^{I,J} a_{ijk} u_i v_j = \sigma w_k^2 \ , \ k \in \langle K \rangle, \qquad [6.69c]$$

where $\sigma$ is a singular value of $\mathcal{A}$, and $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are modes-1, -2 and -3 singular vectors, respectively. The above results can easily be generalized to an $N$th-order tensor.

So, for a symmetric tensor $\mathcal{A} \in \mathbb{R}^{I \times \cdots \times I}$, of order $N$, an eigenvector $\mathbf{u} \in \mathbb{R}^I$ is a non-zero vector satisfying the following polynomial equations (Qi 2005):

$$\sum_{i_1,\cdots,i_{N-1}=1}^{I,\cdots,I} a_{i_1,\cdots,i_N} u_{i_1} \cdots u_{i_{N-1}} = \lambda (u_{i_N})^{N-1} \ , \ i_N \in \langle I \rangle.$$

For $N = 3$, equation [6.68] is found again.

Similarly, for $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, mode-$n$ singular vectors, denoted by $\mathbf{u}^{(n)}$, with $n \in \langle N \rangle$, and associated with a singular value $\sigma$, are solutions to the following polynomial equations:

$$\sum_{i_1,\cdots,i_{n-1},i_{n+1}\cdots,i_N=1}^{I_1,\cdots,I_{n-1},I_{n+1},\cdots,I_N} a_{i_1,\cdots,i_n,\cdots,i_N} u_{i_1}^{(1)} \cdots u_{i_{n-1}}^{(n-1)} u_{i_{n+1}}^{(n+1)} \cdots u_{i_N}^{(N)} = \sigma (u_{i_n}^{(n)})^{N-1}$$

each equation being defined by the factor $u_{i_n}^{(n)}$, with $i_n \in \langle I_n \rangle$ and $n \in \langle N \rangle$.

For $N = 3$, we find again equations [6.69a]–[6.69c], with $(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \mathbf{u}^{(3)}) = (\mathbf{u}, \mathbf{v}, \mathbf{w})$, and $(I_1, I_2, I_3) = (I, J, K)$.

## 6.9. Isomorphisms of tensor spaces

Given that two $\mathbb{K}$-v.s. are isomorphic if they have the same dimension, it is possible to decompose a tensor space in different ways. This means that the elements of a hypermatrix of the space $\mathbb{K}^{I_1 \times \cdots \times I_N}$ can be stored in different ways by combination of modes. Two standard unfolding methods are vectorization and matricization (also called matrix unfolding), which consist of storing the elements in a column vector and a matrix, respectively. More generally, one can store the elements of an $N$th order

hypermatrix in a hypermatrix of order $P < N$, the cases $P = 1$ and $P = 2$ corresponding to vectorization and matricization, respectively.

For example, a fourth-order hypermatrix $\mathcal{A} \in \mathbb{K}^{I_1 \times I_2 \times I_3 \times I_4}$ can be vectorized into a vector $\mathbf{v} \in \mathbb{K}^{I_1 I_2 I_3 I_4}$. The order of dimensions in the product $I_1 I_2 I_3 I_4$ is directly related to the order of variation of the indices $(i_1, i_2, i_3, i_4)$. By choosing to vary $i_1$ the most slowly and $i_4$ the most rapidly, we have $v_i = a_{i_1 i_2 i_3 i_4}$ with $i = i_4 + (i_3 - 1)I_4 + (i_2 - 1)I_3 I_4 + (i_1 - 1)I_2 I_3 I_4$.

The hypermatrix $\mathcal{A}$ can also be unfolded into a matrix as, for example, $\mathbf{M} \in \mathbb{K}^{I_1 \times I_2 I_3 I_4}$ such that $m_{i_1, j} = a_{i_1 i_2 i_3 i_4}$ with $j = i_4 + (i_3 - 1)I_4 + (i_2 - 1)I_3 I_4$.

One can also unfold $\mathcal{A}$ into a third-order hypermatrix by combining, for example, the last two modes. The unfolded hypermatrix $\mathcal{B} \in \mathbb{K}^{I_1 \times I_2 \times I_3 I_4}$ is such that $b_{i_1, i_2, k} = a_{i_1 i_2 i_3 i_4}$ with $k = i_4 + (i_3 - 1)I_4$.

NOTE 6.26.– To highlight how the modes are combined to build the unfoldings, we shall use their dimensions to distinguish them. So, for the three above examples, we shall write the unfoldings as: $\mathbf{a}_{I_1 I_2 I_3 I_4}$, $\mathbf{A}_{I_1 \times I_2 I_3 I_4}$, and $\mathcal{A}_{I_1 \times I_2 \times I_3 I_4}$, instead of $\mathbf{v}$, $\mathbf{M}$, and $\mathcal{B}$, respectively.

Generally speaking, using a $P$-partition of the set $\mathcal{J} = \{i_1, \cdots, i_N\}$ into $P$ disjoint subsets of $n_p$ indices each $\mathcal{J}^{(p)} = \{j_1^{(p)}, \cdots, j_{n_p}^{(p)}\} \subset I$, with $j_k^{(p)} \in \mathcal{J}$, $k \in \langle n_p \rangle$, $p \in \langle P \rangle$, and $\sum_{p=1}^{P} n_p = N$, an $N$th-order hypermatrix of the space $\mathbb{K}^{I_1 \times \cdots \times I_N}$ can be unfolded in the form of a $P$th-order hypermatrix as:

$$\begin{aligned} \mathbb{K}^{I_1} \otimes \cdots \otimes \mathbb{K}^{I_N} = \mathbb{K}^{I_1 \times \cdots \times I_N} &\simeq \mathbb{K}^{J_1^{(1)} \cdots J_{n_1}^{(1)}} \otimes \cdots \otimes \mathbb{K}^{J_1^{(P)} \cdots J_{n_P}^{(P)}} \\ &= \mathbb{K}^{J_1^{(1)} \cdots J_{n_1}^{(1)} \times \cdots \times J_1^{(P)} \cdots J_{n_P}^{(P)}}, \end{aligned} \qquad [6.70]$$

where $J_k^{(p)}$ is the dimension of mode $j_k^{(p)}$, and each space $\mathbb{K}^{J_1^{(p)} \cdots J_{n_p}^{(p)}}$, for $p \in \langle P \rangle$, is associated with a combination of the modes $\{j_1^{(p)}, \cdots, j_{n_p}^{(p)}\}$ of the original hypermatrix. It should be noted that tensor spaces $\mathbb{K}^{I_1 \times \cdots \times I_N}$ and $\mathbb{K}^{J_1^{(1)} \cdots J_{n_1}^{(1)} \times \cdots \times J_1^{(P)} \cdots J_{n_P}^{(P)}}$ having the same dimension $\prod_{n=1}^{N} I_n$ are isomorphic but do not have the same structure.

EXAMPLE 6.27.– Let a third-order hypermatrix $\mathcal{A} = [a_{ijk}] \in \mathbb{K}^{2 \times 2 \times 2}$. We have:

$$\mathbf{a}_{IJK} = \begin{bmatrix} a_{111} & a_{112} & a_{121} & a_{122} & a_{211} & a_{212} & a_{221} & a_{222} \end{bmatrix}^T \in \mathbb{K}^{IJK}$$

$$\mathbf{a}_{JKI} = \begin{bmatrix} a_{111} & a_{211} & a_{112} & a_{212} & a_{121} & a_{221} & a_{122} & a_{222} \end{bmatrix}^T \in \mathbb{K}^{JKI}$$

$$\mathbf{A}_{IJ \times K} = \begin{bmatrix} a_{111} & a_{112} \\ a_{121} & a_{122} \\ a_{211} & a_{212} \\ a_{221} & a_{222} \end{bmatrix} \in \mathbb{K}^{IJ \times K}, \quad \mathbf{A}_{IK \times J} = \begin{bmatrix} a_{111} & a_{121} \\ a_{112} & a_{122} \\ a_{211} & a_{221} \\ a_{212} & a_{222} \end{bmatrix} \in \mathbb{K}^{IK \times J}.$$

The vectorized forms $\mathbf{a}_{IJK}$ and $\mathbf{a}_{JKI}$, as well as the matrix unfoldings $\mathbf{A}_{IJ\times K}$ and $\mathbf{A}_{IK\times J}$, contain all the elements of hypermatrix $\mathcal{A}$, in two different structures, that is, vectors of $\mathbb{K}^8$ and matrices of $\mathbb{K}^{4\times 2}$. However, it is important to note that for a given structure, the order of the dimensions modifies the arrangement of the elements of the hypermatrix, as it can be verified with the examples above. So, the vector spaces $\mathbb{K}^{IJK}$ and $\mathbb{K}^{JKI}$ both correspond to the same structure $\mathbb{K}^8$ of vectors of dimension 8, but are associated with two different arrangements of the elements $a_{ijk}$. Similarly, the matrix spaces $\mathbb{K}^{IJ\times K}$ and $\mathbb{K}^{IK\times J}$ both correspond to the same structure $\mathbb{K}^{4\times 2}$ due to the fact that $I = J = K = 2$, but to different arrangements.

So, the hypermatrix space $\mathbb{K}^{2\times 2\times 2}$ is isomorphic to the vector space $\mathbb{K}^8$ and to the matrix space $\mathbb{K}^{4\times 2}$. This means that the addition of two hypermatrices and the multiplication of a hypermatrix by a scalar can be indifferently performed using the vectorized or matricized forms of the hypermatrices.

EXAMPLE 6.28.– For the tensor $\mathcal{X} \in \mathbb{K}^{I\times J\times K}$ defined in [6.43] in the canonical basis, with $I = J = K = 2$:

$$\mathcal{X} = x_{111}\mathcal{E}_{111}^{(2\times 2\times 2)} + x_{212}\mathcal{E}_{212}^{(2\times 2\times 2)} + x_{221}\mathcal{E}_{221}^{(2\times 2\times 2)} + x_{222}\mathcal{E}_{222}^{(2\times 2\times 2)},$$

the coefficient hypermatrix $[x_{ijk}]$ can be vectorized and matricized, for example, as:

$$\mathbf{x}_{IJK} = \begin{bmatrix} x_{111} & 0 & 0 & 0 & 0 & x_{212} & x_{221} & x_{222} \end{bmatrix}^T \in \mathbb{K}^8$$

$$\mathbf{x}_{IKJ} = \begin{bmatrix} x_{111} & 0 & 0 & 0 & 0 & x_{221} & x_{212} & x_{222} \end{bmatrix}^T \in \mathbb{K}^8$$

$$\mathbf{X}_{I\times JK} = \left[\begin{array}{cc|cc} x_{111} & 0 & 0 & 0 \\ 0 & x_{212} & x_{221} & x_{222} \end{array}\right] \in \mathbb{K}^{2\times 4}$$

$$\mathbf{X}_{I\times KJ} = \left[\begin{array}{cc|cc} x_{111} & 0 & 0 & 0 \\ 0 & x_{221} & x_{212} & x_{222} \end{array}\right] \in \mathbb{K}^{2\times 4}.$$

EXAMPLE 6.29.– Similarly, for example [6.62], we have:

$$\mathbf{a}_{IJK} = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}^T \in \mathbb{K}^8$$

$$\mathbf{A}_{I\times JK} = \left[\begin{array}{cc|cc} -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right] \in \mathbb{K}^{2\times 4}$$

$$= \mathbf{A}_{I\times KJ} = \mathbf{A}_{J\times IK} = \mathbf{A}_{J\times KI} = \mathbf{A}_{K\times IJ} = \mathbf{A}_{K\times JI}.$$

Note that the above equalities result from the symmetry of the corresponding hypermatrix, which is such that:

$a_{121} = a_{211} = a_{112} = a_{222} = 0$ , $a_{221} = a_{122} = a_{212} = 1$ and $a_{111} = -1$.

Tensor vectorization and matricization operations will be detailed in Volume 2 and illustrated along with the CPD/PARAFAC and Tucker models. Matrix representations of tensors will be exploited for solving the parameter estimation problem for these models using the alternating least squares method. In the other books of the collection, the distinction between hypermatrix and tensor will no longer be made, which implies that the bases of the v.s. are implicitly chosen.

# References


Abadir, K.M. and Magnus, J.R. (2005). *Matrix Algebra*. Cambridge University Press, New York.

Acar, E. and Yener, B. (2009). Unsupervised multiway data analysis: a literature survey. *IEEE Transactions on Knowledge and Data Engineering*, 21(1), 6–20.

Allen, R.L. and Mills, D.W. (2004). *Signal Analysis. Time, Frequency, Scale, and Structure*. IEEE Press, John Wiley & Sons, Hoboken.

Aubry, E. (2012). Algèbre linéaire et bilinéaire. Available at: https://math.unice.fr/ eaubry/ Enseignement/Agreg/UE4.pdf.

Baksalary, J.K. and Styan, G.P.H. (2002). Generalized inverses of partitioned matrices in Banachiewicz-Schur form. *Linear Algebra and its Applications*, 354, 41–47.

Barnett, S. (1990). *Matrices: Methods and Applications*. Oxford University Press, Oxford.

Beckmann, P. (1973). *Orthogonal Polynomials for Engineers and Physicists.* The Golem Press, Boulder, CO.

Bellman, R. (1970). *Introduction to Matrix Analysis.* McGraw-Hill, New York.

Ben-Israel, A. and Greville, T.N.E. (2001). *Generalized Inverses: Theory and Applications.* Rutgers University, Piscataway, NJ.

Bernstein, D.S. (2005). *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, Princeton and Oxford.

Bjerhammar, A. (1973). *Theory of Errors and Generalized Matrix Inverses.* Elsevier, Oxford.

Bouilloc, T. and Favier, G. (2012). Nonlinear channel modeling and identification using bandpass Volterra-Parafac models. *Signal Processing*, 92(6), 1492–1498.

Bourennane, S., Fossati, C., and Cailly, A. (2010). Improvement of classification for hyperspectral images based on tensor modeling. *IEEE Geoscience and Remote Sensing Letters*, 7(4), 801–805.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38, 149–171.



*From Algebraic Structures to Tensors,*
First Edition. Edited by Gérard Favier.

Bro, R. (1998). Multi-way analysis in the food industry. Models, algorithms, and applications. PhD thesis, University of Amsterdam, Amsterdam.

Broomfield, J. Construction of the tensor product $V \otimes W$. Available at: http://www-users.math.umn.edu/ broom010/doc/TensorProduct.pdf.

Burns, F., Carlson, D., Haynsworth, E., and Markham, T. (1974). Generalized inverse formulas using the Schur complement. *SIAM Journal on Applied Mathematics*, 26(2), 254–259.

Cardoso, J.-F. (1990). Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem. *IEEE ICASSP'90*, Albuquerque, NM, USA, 2655–2658.

Cardoso, J.-F. (1991). Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors. *Proceedings of IEEE ICASSP'91*, Toronto, ON, Canada, 3109–3112.

Cardoso, J.-F. and Comon P. (1990). Tensor-based independent component analysis. *EUSIPCO'90*, Barcelona, Spain, 673–676.

Carroll, J.D. and Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3), 283–319.

Carroll, J.D., Pruzansky, S., and Kruskal, J.B. (1980). Candelinc: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 45(1), 3–24.

Cattell, R. (1944). Parallel proportional profiles and other principles for determining the choice of factors by rotation. *Psychometrika*, 9, 267–283.

Cichocki, A. (2013). Era of big data processing: a new approach via tensor networks and tensor decompositions. Dans International Workshop on Smart Info-Media Systems in Asia (SISA-2013), Nagoya.

Cichocki, A., Mandic, D., de Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C., and Phan, A.H. (2015). Tensor decompositions for signal applications: from two-way to multiway component analysis. *IEEE Signal Processing Magazine*, 32(2), 145–163.

Cichocki, A., Zdunek, R., Phan, A.H., and Amari, S.-I. (2009). *Nonnegative Matrix and Tensor Factorizations. Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation.* John Wiley & Sons, New York.

Comon, P. (2014). Tensors: a brief introduction. *IEEE Signal Processing Magazine*, 31, 44–53.

Comon, P., Golub, G., Lim, L.-H., and Mourrain, B. (2008). Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3), 1254–1279.

Comon, P. and Jutten, C. (2010). *Handbook of Blind Source Separation. Independent Component Analysis and Applications.* Elsevier, Oxford.

Cong, F., Lin, Q.-H., Kuang, L.-D., Gong, X.-F., Astikainen, P., and Ristaniemi, T. (2015). Tensor decomposition of EEG signals: a brief review. *Journal of Neuroscience Methods*, 248, 59–69.

Conrad, K. Tensor products. Available at: http://www.math.uconn.edu/ kconrad/blurbs/ linmultialg/tensorprod.pdf.

Coppi, R. and Bolasco, S. (1989). *Multiway Data Analysis.* Elsevier, Amsterdam.

Coste, S. (2016). Espaces métriques et topologie. ENS Cachan. Available at: www.scoste.fr/topologie2.0.pdf.

Cullen, C.G. (1997). *Linear Algebra with Applications*. Addison-Wesley, Reading.

da Costa, M.N., Favier, G., and Romano, J.-M. (2018). Tensor modelling of MIMO communication systems with performance analysis and Kronecker receivers. *Signal Processing*, 145, 304–316.

de Almeida, A.L.F., Favier, G., and Mota, J.C.M. (2008). A constrained factor decomposition with application to MIMO antenna systems. *IEEE Transactions on Signal Processing*, 56(6), 2429–2442.

de Almeida, A.L.F. and Favier, G. (2013). Double Khatri-Rao space-time-frequency coding using semi-blind PARAFAC based receiver. *IEEE Signal Processing Letters*, 20(5), 471–474.

de Lathauwer, L. (1997). Signal processing based on multilinear algebra. PhD thesis, KU Leuven, Leuven, Belgium.

de Lathauwer, L., De Moor, B., and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4), 1253–1278.

de Launey, W. and Seberry, J. (1994). The strong Kronecker product. *Journal of Combinatorial Theory*, 66(2), 192–213.

de Silva, V. and Lim, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3), 1084–1127.

de Vos, M., de Lathauwer, L., Vanrumste, B., Van Huffel, S., and Van Paesschen, W. (2007). Canonical decomposition of ictal scalp EEG and accurate source localisation: principles and simulation study. *Computational Intelligence and Neuroscience*, 2007, Article 58253.

Durbin, J. (1960). The fitting of time series. *Rev. Institut Int. de Statistique*, 28(3), 233–244.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), 211–218.

Favier, G. and Bouilloc, T. (2009a). Identification de modèles de Volterra basée sur la décomposition PARAFAC. *22ème Colloque GRETSI*, Dijon.

Favier, G. and Bouilloc, T. (2009b). Parametric complexity reduction of Volterra models using tensor decompositions. *17th European Signal Processing Conference (EUSIPCO 2009)*, Glasgow, Scotland.

Favier, G. and Bouilloc, T. (2010). Identification de modèles de Volterra basée sur la décomposition PARAFAC de leurs noyaux et le filtre de Kalman étendu. *Traitement du Signal*, 27(1), 27–51.

Favier, G. and de Almeida, A.L.F. (2014a). Overview of constrained PARAFAC models. *EURASIP Journal on Advances in Signal Processing*, 2014, 142.

Favier, G. and de Almeida, A.L.F. (2014b). Tensor space-time-frequency coding with semi-blind receivers for MIMO wireless communication systems. *IEEE Transactions on Signal Processing*, 62(22), 5987–6002.

Favier, G., Fernandes, C.A.R., and de Almeida, A.L.F. (2016). Nested Tucker tensor decomposition with application to MIMO relay systems using tensor space-time Coding (TSTC). *Signal Processing*, 128, 318–331.

Favier, G. and Kibangou, A.Y. (2009a). Tensor-based methods for system identification. Part 1: Tensor tools. *International Journal on Sciences and Techniques of Automatic Control*, 3(1), 840–869.

Favier, G. and Kibangou, A.Y. (2009b). Tensor-based methods for system identification. Part 2: Three examples of tensor-based system identification methods. *International Journal on Sciences and Techniques of Automatic Control*, 3(1), 870–889.

Favier, G., da Costa, M.N., de Almeida, A.L.F., and Romano, J.M.T. (2012a). Tensor space-time (TST) coding for MIMO wireless communication systems. *Signal Processing*, 92(4), 1079–1092.

Favier, G., Kibangou, A.Y., and Bouilloc, T. (2012b). Nonlinear system modeling and identification using Volterra-PARAFAC models. *International Journal of Adaptive Control and Signal Processing*, 26, 30–53.

Fernandes, C.E.R, Favier, G., and Mota, J.C.M. (2008). Blind channel identification algorithms based on the Parafac decomposition of cumulant tensors: the single and multiuser cases. *Signal Processing*, 88, 1382–1401.

Fernandes, C.E.R, Favier, G., and Mota, J.C.M. (2009a). Parafac-based blind identification of convolutive MIMO linear systems. *Proceedings of 15th IFAC Symposium on System Identification (SYSID'2009)*, Saint-Malo.

Fernandes, C.A.R, Favier, G., and Mota, J.C.M. (2009b). Blind identification of multiuser nonlinear channels using tensor decomposition and precoding. *Signal Processing*, 89(12), 2644–2656.

Fernandes, C.A.R, Favier, G., and Mota, J.C.M. (2011). PARAFAC-based channel estimation and data recovery in nonlinear MIMO spread spectrum communication systems. *Signal Processing*, 91(2), 311–322.

Gantmacher, F.R. (1959). *The Theory of Matrices*, Vol. 1. Chelsea Publishing Co., New York.

Garcia-Bayona, I. (2019). Traces of Schur and Kronecker products for block matrices. *Khayyam Journal of Mathematics*, doi: 10.22034/kjm.2019.84207.

Golub, G.H. and Van Loan, C.F. (1996). *Matrix Computations*, 3rd ed. John Hopkins University Press, Baltimore.

Gourdon, X. (2009). *Algèbre (Les maths en tête)*, 2nd ed. Ellipses, Paris.

Greub, W.H. (1967). *Linear Algebra*, 4th ed. Springer-Verlag, Berlin, Heidelberg, and New York.

Greub, W.H. (1978). *Multilinear Algebra*, 2nd ed. Springer-Verlag, New York.

Grifone, J. (2011). *Algèbre linéaire*, 4th ed. Cépaduès-Editions, Toulouse.

Guillopé, L. (2010). Espaces de Hilbert et fonctions spéciales. Ecole polytechnique de l'Université de Nantes. Available at: https://www.math.sciences.univ-nantes.fr/ guillope/etn3-hfs/.

Haardt, M., Roemer, F., and Del Galdo, G. (2008). Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems. *IEEE Transactions on Signal Processing*, 56(7), 3198–3213.

Hackbusch, W. (2012). *Tensor Spaces and Numerical Tensor Calculus* (Springer series in Computational Mathematics). Springer, Berlin and Heidelberg, 2012.

Harshman, R.A. (1970). Foundations of the Parafac procedure: models and conditions for an explanatory multimodal factor analysis. UCLA Working papers in Phonetics, 16, 1–84.

Hastad, J. (1990). Tensor rank is NP-complete. *Journal of Algorithms*, 11, 644–654.

Hillar, C.J. and Lim, L.-H. (2013). Most tensor problems are NP-hard. *Journal of the ACM*, 60(6), 45:1–45:39.

Hitchcock, F.L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematical Physics*, 6(1), 165–189.

Horn, R.A. and Johnson, C.A. (1985). *Matrix Analysis*. Cambridge University Press, Cambridge.

Horn, R.A. and Johnson, C.A. (1991). *Topics in Matrix Analysis*. Cambridge University Press, Cambridge.

Householder, A.S. (1958). Unitary triangularization of a nonsymmetric matrix. *Communications of the ACM*, 5, 339–342.

Hunyadi, B., Dupont, P., Van Paesschen, W., and Van Huffel, S. (2016). Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data. *WIREs Data Mining and Knowledge Discovery*, doi: 10.1002/widm.1197.

Jiang, T., Sidiropoulos, N.D., and ten Berge, J.M.F. (2001). Almost-sure identifiability of multidimensional harmonic retrieval. *IEEE Transactions on Signal Processing*, 49(9), 1849–1859.

Kibangou, A.Y. and Favier, G. (2007). Blind joint identification and equalization of Wiener-Hammerstein communication channels using PARATUCK-2 tensor decomposition. *Proceedings of EUSIPCO'2007*, Poznan, Poland, September.

Kibangou, A.Y. and Favier, G. (2008). Matrix and tensor decompositions for identification of block-structured nonlinear channels in digital transmission systems. *9th IEEE Signal Processing Advances in Wireless Communications Workshop (SPAWC 2008)*, Recife, Brazil.

Kibangou, A.Y. and Favier, G. (2009). Identification of parallel-cascade Wiener systems using joint diagonalization of third-order Volterra kernel slices. *IEEE Signal Processing Letters*, 16(3), 188–191.

Kibangou, A.Y. and Favier, G. (2010). Tensor analysis-based model structure determination and parameter estimation for block-oriented nonlinear systems. *IEEE Journal of Selected Topics in Signal Processing*, 4(3), 514–525.

Kiers, H.A.L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, 14(2), 105–122.

Kolda, T.G. and Bader, B.W. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3), 455–500.

Kroonenberg, P.M. (2008). *Applied Multiway Data Analysis*. John Wiley & Sons, New York.

Kruskal, J.B. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions. *Linear Algebra and its Applications,* 18, 95–138.

Kruskal, J.B. (1989). Rank, decomposition, and uniqueness for 3-way and N-way arrays. In *Multiway Data Analysis*, R. Coppi and S. Bolasco (eds). Elsevier, Amsterdam, 7–18.

Lahat, D., Adali, T., and Jutten, C. (2015). Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of IEEE*, 103(9), 1449–1477.

Lancaster, P. and Tismenetsky, M. (1985). *The Theory of Matrices with Applications.* Academic Press, New York.

Lang, S. (2002). *Algebra*. Springer, New York.

Lascaux, P. and Théodor, R. (2000). *Analyse numérique matricielle appliquée à l'art de l'ingénieur. Tome 1 - Méthodes directes*. Dunod, Paris.

Lee, N. and Cichocki, A. (2017). Fundamental tensor operations for large-scale data analysis using tensor network formats. *Multidimensional Systems and Signal Processing*, 29, 921–960.

Levinson, N. (1947). The Wiener r.m.s. (root mean-square) error criterion in filter design and prediction. *Journal of Mathematical Physics*, 25, 261–278.

Lim, L.-H. (2005). Singular values and eigenvalues of tensors: a variational approach. *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP'05)*, 129–132.

Lim, L.-H. (2013). Tensors and hypermatrices, 2nd edition. In *Handbook of Linear Algebra*, L. Hogben (ed.). Chapman and Hall/CRC Press. Available at: https://www.stat.uchicago.edu/lekheng/work/hla.pdf.

Liu, X., Jin, H., and Visnjic, J. (2016). Representations of generalized inverses and Drazin inverse of partitioned matrix with Banachiewicz-Schur forms. *Mathematical Problems in Engineering*, 2016, Article 9236281, 14 pages.

Lu, H., Plataniotis, K., and Venetsanopoulos, A. (2011). A survey of multilinear subspace learning for tensor data. *Pattern Recognition*, 44(7), 1540–1551.

Lu, T.-T. and Shiou, S.-H. (2002). Inverses of $2 \times 2$ block matrices. *Computers and Mathematics with Applications*, 43, 119–129.

Lütkepohl, H. (1996). *Handbook of Matrices.* Wiley, Chichester.

McCullagh, P. (1987). *Tensor Methods in Statistics.* Chapman and Hall, New York.

Magnus, J.R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, Chichester.

Mahalanobis, P.C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 49–55.

Marsaglia, G. and Styan, G.P.H. (1974a). Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra*, 2, 269–292.

Marsaglia, G. and Styan, G.P.H. (1974b). Rank conditions for generalized inverses of partitioned matrices. *Sankhyä Series A*, 36(4), 437–442.

Meyer, C.D. (2000). *Matrix Analysis and Applied Linear Algebra.* SIAM, Philadelphia.

Moore, E.H. (1935). *Generalized Analysis, Part I*. Memoirs of the American Philosophical Society, Philadelphia, 147–209.

Morup, M. (2011). Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 24–40.

Nion, D. and Sidiropoulos, N.D. (2010). Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar. *IEEE Transactions on Signal Processing*, 58(11), 5693–5705.

Nion, D., Mokios, K.N., Sidiropoulos, N.D., and Potamianos, A. (2010). Batch and adaptive PARAFAC-based blind separation of convolutive speech mixtures. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6), 1193–1207.

Noble, B. and Daniel, J.W. (1988). *Applied Linear Algebra*, 3rd ed. Prentice-Hall International, London.

Nomakuchi, K. (1980). On the characterization of generalized inverses by bordered matrices. *Linear Algebra and its Applications*, 33, 1–8.

Padhy, S., Goovaerts, G., Boussé, M., de Lathauwer, L., and Van Huffel, S. (2018). The power of tensor-based approaches in cardiac applications. In *Biomedical Signal Processing – Advances in Theory, Algorithms and Applications*, R. Naik Ganesh (ed.). Springer-Verlag GmbH, 1–34.

Papalexakis, E.E., Faloutsos, C., Sidiropoulos, N.D., and Kolda, T.G. (2016). Tensors for data mining and data fusion: models, applications, and scalable algorithms. *ACM Transactions on Intelligent Systems and Technology*, 8(2), 16.1–16.44.

Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51, 406–413.

Perlis, S. (1958). *Theory of Matrices.* Addison-Wesley, Reading.

Pollock, D.S.G. (1999). *A Handbook of Time-Series Analysis, Signal Processing and Dynamics.* Academic Press, New York.

Powell, P.D. (2011). Calculating determinants of block matrices, arXiv:1112.4379 [math.RA], December.

Qi, L. (2005). Eigenvalues of a real supersymmetric tensor. *Journal of Symbolic Computation*, 40(6), 1302–1324.

Reinhard, H. (1997). *Eléments de mathématiques du signal. Tome 1 - Signaux déterministes*. Dunod, Paris.

Rotella, F. and Borne, P. (1995). *Théorie et pratique du calcul matriciel*. Technip, Paris.

Sage, M. (2005). Introduction aux espaces quotients. Available at: www.normalesup.org/sage/Enseignement/Cours/Quotient.pdf.

Schultz, T., Fuster, A., Ghosh, A., Deriche, R., Florack, L., and Lek-Heng, L. (2014). Higher-order tensors in diffusion imaging. In *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, C. F. Westin, A. Vilanova, and B. Burgeth (eds). Springer, 129–161. Available at: https://hal.inria.fr/hal-00848526.

Serre, D. (2002). *Matrices. Theory and Applications.* Springer-Verlag, New York.

Sidiropoulos, N.D., Giannakis, G.B., and Bro, R. (2000a). Blind PARAFAC receivers for DS-CDMA systems. *IEEE Transactions on Signal Processing,* 48(3), 810–823.

Sidiropoulos, N.D., Bro, R., and Giannakis, G.B. (2000b). PARAFAC factor analysis in sensor array processing. *IEEE Transactions on Signal Processing*, 48(8), 2377–2388.

Sidiropoulos, N.D. (2001). Generalizing Carathéodory's uniqueness of harmonic parameterization to N dimensions. *IEEE Transactions on Information Theory*, 47(4), 1687–1690.

Sidiropoulos, N.D., de Lathauwer, L., Fu, X., Huang, K., Papalexakis, E.E., and Faloutsos, C. (2017). Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13), 3551–3582.

Smilde, A., Bro, R., and Geladi, P. (2004). *Multi-way Analysis. Applications in the Chemical Sciences*. John Wiley and Sons, Chichester.

Stegeman, A. and Comon, P. (2010). Subtracting a best rank-1 approximation does not necessarily decrease tensor rank. *Linear Algebra and its Applications*, 433(7), 1276–1300.

Strang, G. (1980). *Linear Algebra and its Applications*, 3rd ed. Brooks/Cole, Thomson Learning, Boston.

ten Berge, J.M.F. (1991). Kruskal's polynomial for 2×2×2 arrays and a generalization to 2×n×n arrays. *Psychometrika*, 56, 631–636, doi: 10.1007/BF02294495.

Tian, Y. (2004). Rank equalities for block matrices and their Moore-Penrose inverses. *Houston Journal of Mathematics*, 30(2), 483–510.

Tian, Y. and Takane, Y. (2005). Schur complements and Banachiewicz-Schur forms. *Electronic Journal of Linear Algebra*, 13, 405–418.

Tian, Y. and Takane, Y. (2009). More on generalized inverses of partitioned matrices with Banachiewicz-Schur forms. *Linear Algebra and its Applications*, 430, 1641–1655.

Tracy, D.S. and Singh, R.P. (1972). A new matrix product and its applications in matrix differentiation. *Statistica Neerlandica*, 26, 143–157.

Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika,* 31, 279–311.

Vasilescu, M.A.O. and Terzopoulos, D. (2002). Multilinear analysis of image ensembles: TensorFaces. *Proceedings of the European Conference on Computer Vision (ECCV '02)*, Copenhagen, Denmark, May, 447–460.

Velasco-Forero, S. and Angulo, J. (2013). Classification of hyperspectral images by tensor modeling and additive morphological decomposition. *Pattern Recognition*, 46(2), 566–577.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis*, P.R. Krishnaiah (ed.). Academic Press, New York, 391–420.

Young N. (1988). *An Introduction to Hilbert Space.* Cambridge University Press, Cambridge.

Zhang, F. (1999). *Matrix Theory: Basic Results and Techniques*. Springer-Verlag, New York.

# Index




*From Algebraic Structures to Tensors,*
First Edition. Edited by Gérard Favier.

## D



## E

## N



## O

## P



## Q



## R



## S

## T



## U



## V



## Y

Other titles from

in

Digital Signal and Image Processing

## 2019

MEYER Fernand
*Topographical Tools for Filtering and Segmentation 1: Watersheds on Node- or Edge-weighted Graphs*
*Topographical Tools for Filtering and Segmentation 2: Flooding and Marker-based Segmentation on Node- or Edge-weighted Graphs*

## 2017

CESCHI Roger, GAUTIER Jean-Luc
*Fourier Analysis*

CHARBIT Maurice
*Digital Signal Processing with Python Programming*

CHAO Li, SOULEYMANE Bella-Arabe, YANG Fan
*Architecture-Aware Optimization Strategies in Real-time Image Processing*

FEMMAM Smain
*Fundamentals of Signals and Control Systems*
*Signals and Control Systems: Application for Home Health Monitoring*

MAÎTRE Henri
*From Photon to Pixel – 2nd edition*

PROVENZI Edoardo
*Computational Color Science: Variational Retinex-like Methods*

## 2015

BLANCHET Gérard, CHARBIT Maurice
*Digital Signal and Image Processing using MATLAB®*
*Volume 2 – Advances and Applications:The Deterministic Case – 2nd edition*
*Volume 3 – Advances and Applications: The Stochastic Case – 2nd edition*

CLARYSSE Patrick, FRIBOULET Denis
*Multi-modality Cardiac Imaging*

GIOVANNELLI Jean-François, IDIER Jérôme
*Regularization and Bayesian Methods for Inverse Problems in Signal and Image Processing*

## 2014

AUGER François
*Signal Processing with Free Software: Practical Experiments*

BLANCHET Gérard, CHARBIT Maurice
*Digital Signal and Image Processing using MATLAB®*
*Volume 1 – Fundamentals – 2nd edition*

DUBUISSON Séverine
*Tracking with Particle Filter for High-dimensional Observation and State Spaces*

ELL Todd A., LE BIHAN Nicolas, SANGWINE Stephen J.
*Quaternion Fourier Transforms for Signal and Image Processing*

FANET Hervé
*Medical Imaging Based on Magnetic Fields and Ultrasounds*

MOUKADEM Ali, OULD Abdeslam Djaffar, DIETERLEN Alain
*Time-Frequency Domain for Segmentation and Classification of Non-stationary Signals: The Stockwell Transform Applied on Bio-signals and Electric Signals*

NDAGIJIMANA Fabien
*Signal Integrity: From High Speed to Radiofrequency Applications*

PINOLI Jean-Charles
*Mathematical Foundations of Image Processing and Analysis Volumes 1 and 2*

TUPIN Florence, INGLADA Jordi, NICOLAS Jean-Marie
*Remote Sensing Imagery*

VLADEANU Calin, EL ASSAD Safwan
*Nonlinear Digital Encoders for Data Communications*

## 2013

GOVAERT Gérard, NADIF Mohamed
*Co-Clustering*

DAROLLES Serge, DUVAUT Patrick, JAY Emmanuelle
*Multi-factor Models and Signal Processing Techniques: Application to Quantitative Finance*

LUCAS Laurent, LOSCOS Céline, REMION Yannick
*3D Video: From Capture to Diffusion*

MOREAU Eric, ADALI Tulay
*Blind Identification and Separation of Complex-valued Signals*

PERRIN Vincent
*MRI Techniques*

WAGNER Kevin, DOROSLOVACKI Milos
*Proportionate-type Normalized Least Mean Square Algorithms*

FERNANDEZ Christine, MACAIRE Ludovic, ROBERT-INACIO Frédérique
*Digital Color Imaging*

FERNANDEZ Christine, MACAIRE Ludovic, ROBERT-INACIO Frédérique
*Digital Color: Acquisition, Perception, Coding and Rendering*

NAIT-ALI Amine, FOURNIER Régis
*Signal and Image Processing for Biometrics*

OUAHABI Abdeljalil
*Signal and Image Multiresolution Analysis*

## 2011

CASTANIÉ Francis
*Digital Spectral Analysis: Parametric, Non-parametric and Advanced Methods*

DESCOMBES Xavier
*Stochastic Geometry for Image Analysis*

FANET Hervé
*Photon-based Medical Imagery*

MOREAU Nicolas
*Tools for Signal Compression*

## 2010

NAJMAN Laurent, TALBOT Hugues
*Mathematical Morphology*

## 2009

BERTEIN Jean-Claude, CESCHI Roger
*Discrete Stochastic Processes and Optimal Filtering – 2nd edition*

CHANUSSOT Jocelyn *et al.*
*Multivariate Image Processing*

DHOME Michel
*Visual Perception through Video Imagery*

GOVAERT Gérard
*Data Analysis*

GRANGEAT Pierre
*Tomography*

MOHAMAD-DJAFARI Ali
*Inverse Problems in Vision and 3D Tomography*

SIARRY Patrick
*Optimization in Signal and Image Processing*

## 2008

ABRY Patrice *et al.*
*Scaling, Fractals and Wavelets*

GARELLO René
*Two-dimensional Signal Analysis*

HLAWATSCH Franz *et al.*
*Time-Frequency Analysis*

IDIER Jérôme
*Bayesian Approach to Inverse Problems*

MAÎTRE Henri
*Processing of Synthetic Aperture Radar (SAR) Images*

MAÎTRE Henri
*Image Processing*

NAIT-ALI Amine, CAVARO-MENARD Christine
*Compression of Biomedical Images and Signals*

NAJIM Mohamed
*Modeling, Estimation and Optimal Filtration in Signal Processing*

QUINQUIS André
*Digital Signal Processing Using Matlab*

## 2007

BLOCH Isabelle
*Information Fusion in Signal and Image Processing*

GLAVIEUX Alain
*Channel Coding in Communication Networks*

OPPENHEIM Georges *et al.*
*Wavelets and their Applications*

## 2006

CASTANIÉ Francis
*Spectral Analysis*

NAJIM Mohamed
*Digital Filters Design for Signal and Image Processing*