de Gruyter Studies in Mathematics 17

de Gruyter Studies in Mathematics

1 Riemannian Geometry, *Wilhelm Klingenberg*
2 Semimartingales, *Michel Métivier*
3 Holomorphic Functions of Several Variables, *Ludger Kaup and Burchard Kaup*
4 Spaces of Measures, *Corneliu Constantinescu*
5 Knots, *Gerhard Burde and Heiner Zieschang*
6 Ergodic Theorems, *Ulrich Krengel*
7 Mathematical Theory of Statistics, *Helmut Strasser*
8 Transformation Groups, *Tammo tom Dieck*
9 Gibbs Measures and Phase Transitions, *Hans-Otto Georgii*
10 Analyticity in Infinite Dimensional Spaces, *Michel Hervé*
11 Elementary Geometry in Hyperbolic Space, *Werner Fenchel*
12 Transcendental Numbers, *Andrei B. Shidlovskii*
13 Ordinary Differential Equations, *Herbert Amann*
14 Dirichlet Forms and Analysis on Wiener Space, *Nicolas Bouleau and Francis Hirsch*
15 Nevanlinna Theory and Complex Differential Equations, *Ilpo Laine*
16 Rational Iteration, *Norbert Steinmetz*

Francesco Altomare · Michele Campiti

Korovkin-type Approximation Theory and its Applications

Walter de Gruyter
Berlin · New York 1994

Authors
Francesco Altomare
Michele Campiti
Dipartimento di Matematica
Università degli Studi di Bari
I-70125 Bari, Italy

1991 Mathematics Subject Classification: 41-02; 41A10, 41A25, 41A35, 41A36, 41A63, 41A65, 42A10, 46A55, 46E15, 46E27, 47A58, 47B38, 47B65; 46J10, 46L89, 47D06, 47B48, 47C10, 60F, 60J25, 60J35, 60J60

Key words: Positive Radon measures, Choquet boundaries, Korovkin-type theorems, positive linear projections, finitely defined operators, lattice homomorphisms, approximation by positive linear operators, rate of convergence, probabilistic methods in approximation theory, positive approximation processes, Feller semigroups, degenerate diffusion equations.

Printed on acid-free paper which falls within the guidelines of the ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication Data

Korovkin-type approximation theory and its applications / Francesco Altomare, Michele Campiti.
p. cm. – (De Gruyter studies in mathematics : 17)
Includes bibliographical references and index.
ISBN 3-11-014178-7
1. Approximation theory. I. Altomare, Francesco, 1951– II. Campiti, Michele, 1959– . III. Series.
QA221.K68 1994
511'.42–dc20 94-16853
CIP

Die Deutsche Bibliothek – Cataloging-in-Publication Data

Altomare, Francesco: Korovkin type approximation theory and its applications / Francesco Altomare ; Michele Campiti. – Berlin ; New York: de Gruyter, 1994
(De Gruyter studies in mathematics ; 17)
ISBN 3-11-014178-7
NE: Campiti, Michele:; GT

Printed in Germany.
Typesetting: ASCO Trade Ltd. Hong Kong. Printing: Gerike GmbH, Berlin.
Binding: Lüderitz & Bauer GmbH, Berlin. Cover design: Rudolf Hübler, Berlin.

Affectionately and gratefully dedicated to
Raffaella, Bianca Maria and Gianluigi
and to Giusy and my parents

Preface

Since their discovery the simplicity and, at the same time, the power of the classical theorems of Korovkin impressed several mathematicians.

During the last thirty years a considerable amount of research extended these theorems to the setting of different function spaces or more general abstract spaces such as Banach lattices, Banach algebras, Banach spaces and so on.

This work, in fact, delineated a new theory that we may now call Korovkin-type approximation theory.

At the same time, strong and fruitful connections of this theory have been revealed not only with classical approximation theory, but also with other fields such as functional analysis, harmonic analysis, measure theory, probability theory and partial differential equations.

This has been accomplished by a large number of mathematicians ranging from specialists in approximation theory to functional analysts.

A selected part of the theory is already documented in the monographs of Donner [1982] and Keimel and Roth [1992].

With this book we hope to contribute further to the subject by presenting a modern and comprehensive exposition of the main aspects of the theory in spaces of continuous functions (vanishing at infinity, respectively) defined on a compact space (a locally compact space, respectively), together with its main applications.

We have chosen to treat these function spaces since they play a central role in the whole theory and are the most useful for the applications.

Besides surveying both classical and recent results in the field, the book also contains a certain amount of new material. In any case, the majority of the results appears in a book for the first time.

We are happy to acknowledge our indebtedness to several friends and colleagues.

First, we would like to thank Hubert Berens, Heinz H. Gonska, Silvia Romanelli and Yurji A. Shashkin for reading a large part of the manuscript, for their fruitful suggestions and for their help in correcting mistakes.

We are also grateful to Ferdinand Beckhoff and Michael Pannenberg, to whom we asked to write Appendices B and A, respectively, for their collaboration in outlining the development of the theory in the setting of Banach algebras.

We are particularly indebted to Ferdinand Beckhoff, George Maltese, Rainer Nagel, Ioan Rasa and Rousłan K. Vasil'ev who read and checked the entire manuscript, gave valuable advice and criticisms and kindly corrected several mistakes and inaccuracies as well as our poor English. To them we extend our particular warm thanks.

We want to express our deep gratitude to Heinz Bauer not only for his interest in this work, for reading several chapters and for making several remarks, but also for inviting us to publish the book in the prestigious series De Gruyter Studies in Mathematics of which he is co-editor.

We thank him and the other editors of the series for accepting the monograph and Walter De Gruyter & Co. for producing it according to their usual high standard quality.

Finally we express our great affection and gratitude to Raffaella, Bianca Maria and Gianluigi and to Giusy, for their patience and understanding as well as for their constant encouragement over all these years without which this monograph would have never been completed.

We dedicate the book to them.

Bari, October 1993

Francesco Altomare
Michele Campiti

Contents

Introduction

Positive approximation processes play a fundamental role in approximation theory and appear in a very natural way in many problems dealing with the approximation of continuous functions, especially when one requires further qualitative properties, such as monotonicity, convexity, shape preservation and so on.

In 1953, P.P. Korovkin discovered the, perhaps, most powerful and, at the same time, simplest criterion in order to decide whether a given sequence $(L_n)_{n \in \mathbb{N}}$ of positive linear operators on the space $\mathscr{C}([0,1])$ is an approximation process, i.e., $L_n(f) \to f$ uniformly on $[0,1]$ for every $f \in \mathscr{C}([0,1])$.

In fact, it is sufficient to verify that $L_n(f) \to f$ uniformly on $[0,1]$ only for $f \in \{\mathbf{1}, x, x^2\}$.

Starting with this result, during the last thirty years a considerable number of mathematicians have extended Korovkin's theorem to other function spaces or, more generally, to abstract spaces, such as Banach lattices, Banach algebras, Banach spaces and so on.

This work, in fact, delineated a new theory that we may now call Korovkin-type approximation theory (in short, KAT).

At the same time, strong and fruitful connections of this theory have also been revealed not only with classical approximation theory, but also with other fields such as functional analysis (abstract Choquet boundaries and convexity theory, uniqueness of extensions of positive linear forms, convergence of sequences of positive linear operators in Banach lattices, structure theory of Banach lattices, convergence of sequences of linear operators in Banach algebras and in C^*-algebras, structure theory of Banach algebras, approximation problems in function algebras), harmonic analysis (convergence of sequences of convolution operators on function spaces and function algebras on (locally) compact topological groups, structure theory of topological groups), measure theory and probability theory (weak convergence of sequences of positive Radon measures and positive approximation processes constructed by probabilistic methods), and partial differential equations (approximation of solutions of Dirichlet problems and of diffusion equations).

After the pioneer work of P.P. Korovkin and his students E.N. Morozov and V.I. Volkov, that came to light in the fifties, a decisive step toward the modern development of Korovkin-type approximation theory was carried out by Yu.A. Shashkin when, in the sixties, he characterized the finite Korovkin sets in the space $\mathscr{C}(X)$, X compact metric space, in many respects and, in particular, in terms of geometric properties of state spaces.

The development of KAT in $\mathscr{C}(X)$-spaces was also pursued and amplified by Wulbert [1968], Berens and Lorentz [1973], [1975], and Bauer [1973]. In particular, Bauer expanded the investigation of Korovkin subspaces by using, in a systematic way, suitable enveloping functions, previously considered in connection with abstract Dirichlet problems.

As a matter of fact, these methods allowed Bauer [1973], [1974] to characterize Korovkin subspaces also in the framework of adapted spaces.

This was the first systematic study of Korovkin subspaces carried out in spaces of continuous functions on locally compact Hausdorff spaces.

This line of investigation led Bauer and Donner [1978], [1986] to the development of a satisfactory parallel theory in the space $\mathscr{C}_0(X)$ of all real-valued continuous functions vanishing at infinity on a locally compact space X.

On the other hand, in the seventies and eighties, Korovkin-type approximation theory rapidly grew along many other directions including other classical function Banach spaces, such as $L^p(X, \mu)$-spaces, and more abstract spaces such as locally convex ordered spaces and Banach lattices, Banach algebras, Banach spaces and so on.

In the specific setting of $L^p(X, \mu)$-spaces nowadays we have very satisfactory results. Noteworthy achievements were obtained by several mathematicians and culminated in the (in many respects) conclusive results of Donner [1980], [1981], [1982].

KAT has been well developed also in the framework of Banach lattices and locally convex vector lattices as is documented, for instance, in Donner's book [1982]; there, theorems on the extensions of positive linear operators are fruitfully used as a main tool.

In this context, fundamental contributions have been carried out by the Russian school (notably, M.A. Krasnosel'skii, E.A. Lifshits, S.S. Kutateladze, A.M. Rubinov, R.K. Vasil'ev) and by the German school (especially, K. Donner, H.O. Flösser, E. Scheffold and M. Wolff).

As far as we know, the development of the theory is still incomplete in the context of Banach algebras (especially in the non commutative case) and, even more so, in Banach spaces, although some attempts have been made to frame the different problems in a more systematic way, for instance, by Altomare [1982a], [1982c], [1984], [1986], [1987a], Pannenberg [1985], [1992], Limaye and Namboodiri [1979], [1986], Labsker [1971], [1972], [1982], [1985], [1989a].

Very recently, Keimel and Roth [1992] presented a unified approach to Korovkin-type approximation theory in the framework of so-called locally convex cones.

The reader will find a quite complete picture of what has been achieved in these fields in Appendix D, where we present a subject classification of KAT, which reflects the main lines of its development.

All references in the final bibliography concerned with KAT, are classified according to this subject classification; the classification numbers are indicated by the prefix SC.

Furthermore, in the same Appendix D we also include a subject index with a list of all references pertaining to every subsection of the subject classification.

However, in spite of our efforts, we are sure that the list of references is not complete. We apologize for possible errors and omissions due to lack of accurate information.

The main purpose of this book is to present a modern and comprehensive exposition of the main aspects of Korovkin-type approximation theory in the spaces $\mathscr{C}_0(X)$ (X locally compact non-compact space) and $\mathscr{C}(X)$ (X compact space), together with its main applications.

The function spaces we have chosen to treat play a central role in the whole theory and are the most useful for the applications in the various univariate, multivariate and infinite dimensional settings.

However we occasionally give some results concerning $L^p(X, \mu)$-spaces too.

The book is mainly intended as a reference text for research workers in the field; a large part of it can also serve as a textbook for a graduate level course.

The organization of the material does not follow the historical development of the subject and allows us to present the most important part of the theory in a concise way.

As a prerequisite, we require a basic knowledge of the theory of Radon measures on locally compact spaces as well as some standard topics from functional analysis such as various Hahn-Banach extension and separation theorems, the Krein-Milman theorem and Milman's converse theorem.

For the reader's convenience and to make the exposition self-contained, we collect all these prerequisites in Chapter 1.

However in some few sections, such as Sections 4.3, 5.2, 6.1 and 6.2, in order to present some significant applications of Korovkin approximation theory, we have also required a solid background on measures on topological spaces and the Riesz representation theorem, on some basic principles of probability theory, on Choquet's integral representation theory and on C_0-semigroups of bounded linear operators.

The definitions and the results pertaining to these topics are briefly reviewed also in Chapter 1 in some starred sections.

Thus a starred section or subsection in principle is not essential for the whole of the book but it serves only for a particular (notable) application that will be indicated in the same section.

Chapters 2, 3 and 4 are devoted to the main aspects of Korovkin-type approximation theory in $\mathscr{C}_0(X)$ and $\mathscr{C}(X)$-spaces.

The fundamental problem consists in studying, for a given positive linear operator $T\colon \mathscr{C}_0(X) \to \mathscr{C}_0(Y)$, those subspaces H of $\mathscr{C}_0(X)$ (if any) which have the remarkable property that every arbitrary equicontinuous net of positive linear operators (or positive contractions) from $\mathscr{C}_0(X)$ into $\mathscr{C}_0(Y)$ converges strongly to T whenever it converges to T on H.

Such subspaces are called Korovkin subspaces for T.

Historically, this problem (and related ones) was first considered when T is the identity operator; this classical case is developed in many respects in Chapter 4.

In the same chapter we also point out the strong interplay between KAT and Choquet's integral representation theory, as well as Stone-Weierstrass-type theorems.

Furthermore we present a detailed analysis of the existence of finite dimensional Korovkin subspaces and we give some estimates of the minimal dimension of such subspaces in terms of the small inductive dimension of the underlying space as well as of other topological parameters.

In Chapter 3 we characterize Korovkin subspaces for an arbitrary positive linear operator by emphasizing, among other things, additional properties, such as universal Korovkin-type properties with respect to positive linear operators, monotone operators and linear contractions.

We also consider other important classes of operators, such as positive projections, finitely defined operators and lattice homomorphisms.

The results concerning positive projections lead to some applications to Bauer simplices and to potential theory.

Finitely defined operators are important in this context because they are the only positive linear operators which can admit finite dimensional Korovkin subspaces.

Several characterizations are provided for this case and the interplay between Korovkin subspaces for finitely defined operators and Chebyshev systems is stressed.

Our main approach in developing the theory uses the basic idea, whose quintessence goes back to Korovkin, of studying approximation problems for equicontinuous nets of positive linear forms (Radon measures). This study, in fact, is carried out in Chapter 2. We deal with both the general case when the limit functionals are arbitrary bounded positive Radon measures, and the case when they are discrete or Dirac measures. The latter leads directly to the study of Choquet boundaries.

Chapters 5 and 6 are mainly concerned with applications to:

- Approximation of continuous functions by means of positive linear operators.
- Approximation and representation of the solutions of particular partial differential equations of diffusion type, by means of powers of positive linear operators.

More precisely, in Chapter 5 we give the first and best-known applications of Korovkin-type approximation theory, namely to the approximation of continuous functions defined on real intervals (bounded or not).

Throughout the chapter we describe different kinds of positive approximation processes.

Particular care is devoted to probabilistic-type operators, discrete-type operators, convolution operators for periodic functions and summation methods.

In general our results concern the uniform convergence on the whole interval or on compact subsets of it.

However in some cases we also investigate the convergence in L^p-spaces or in suitable weighted function spaces.

For almost all the specific approximation processes we consider in Chapter 5, we give estimates of the rate of convergence in terms of the classical modulus of continuity and, in some cases, of the second modulus of smoothness.

These estimates are not the sharpest possible but, on the other hand, an adequate analysis of improving them or of using more suitable moduli of smoothness would have gone too far for the purpose of this book.

For more details concerning rates of convergence of the specific approximation processes considered in Chapter 5 or of other more general ones, we refer, for instance, other than to the pioneering book of Korovkin [1960], also to the excellent books of Butzer and Nessel [1971], De Vore [1972], Ditzian and Totik [1987], Lorentz [1986 a] and Sendov and Popov [1988], for the univariate case as well as to the articles of Censor [1971] and Nishishiraho [1977], [1982b], [1983], [1987] for the multivariate and the infinite dimensional cases, respectively (see also Keimel and Roth [1992]).

In the final Chapter 6 we present a detailed analysis of some further sequences of positive linear operators that have been studied recently. These operators seem to play a non negligible role in some fine aspects of approximation theory. They connect the theory of C_0-semigroups of operators, partial differential equations and Markov processes.

The main examples we consider are the Bernstein-Schnabl operators, the Stancu-Schnabl operators and the Lototsky-Schnabl operators.

All these operators are constructed by means of a positive projection acting on the space of continuous functions on a convex compact set.

This general framework has the advantage of unifying the presentation of various well-known approximation processes and, at the same time, of providing new ones both in univariate and multivariate settings and in the infinite dimensional case, e.g., Bauer simplices.

After a careful analysis of the approximation properties of these operators, both from a qualitative and a quantitative point of view, a discussion follows of their monotonicity properties as well as their preservation of some global smoothness properties of functions, e.g., Hölder continuity.

Subsequently we show how these operators are strongly connected with initial and (Wentcel-type) boundary value problems in the theory of partial differential equations.

In fact, we prove that there exists a uniquely determined Feller semigroup that can be represented in terms of powers of the operators with which we are dealing.

The infinitesimal generator of the semigroup is explicitly determined in a core of its domain and, in the finite dimensional case, it turns out to be an elliptic second-order differential operator which degenerates on the Choquet boundary of the range of the projection.

Consequently we derive a representation and some qualitative properties of the solutions of the Cauchy problems which correspond to these diffusion equations.

We also emphasize the probabilistic meaning of our results by describing the transition function and the asymptotic behavior of the Markov processes governed by the above mentioned diffusion equations.

In Appendices A and B, written by M. Pannenberg and F. Beckhoff respectively, some of the main developments of Korovkin-type approximation theory in the setting of Banach algebras (commutative or not) are outlined essentially without proofs.

There the reader will have the opportunity to realize once again how Korovkin-type approximation theory, besides having its own interest, may also be used for solving problems of other important fields, such as Banach algebras and particular function algebras on locally compact abelian groups.

In Appendix C we list several concrete examples of Korovkin sets and determining sets. This list could be useful for rapidly checking those Korovkin sets that are most appropriate for the applications.

Finally we close every section with historical notes, giving credit and detailed references to supplementary results, so that, except in a few cases, we do not give references in the text. However, any inaccuracy or omission for historical details or in assigning priorities is unintentional and we apologize for possible errors.

In a diagram we also indicate some of the main connections among the various sections of the book.

After looking closely at the above mentioned subject classification in Appendix D, the reader will clearly see that the topics we have selected are not exhaustive with respect to the complete theory.

We have not dealt with certain other aspects of the theory, some of which have been indicated at the beginning of this introduction.

We also have to mention, for their particular interest and value of further investigations, those results concerning spaces of differentiable functions and spaces of continuous affine functions on convex compact subsets. The latter subject has been recently studied by Dieckmann [1993]. There the reader will also find a rather complete survey on this topic.

Although the aim of the book is to survey both classical and recent results in the field, the reader will find a certain amount of new material. In any case, the majority of the results presented here appears in a book for the first time.

We hope this monograph may serve not only to illustrate how effectively Korovkin-type approximation theory acts as a contact point between approximation theory and other areas of researches, notably functional analysis, but it may also lead to further investigations and to new applications to the theory of approximation by positive linear operators.

Interdependence of sections

Legend:

1. On the line corresponding to a given section on the vertical axis we have indicated all the sections on which the section depends.

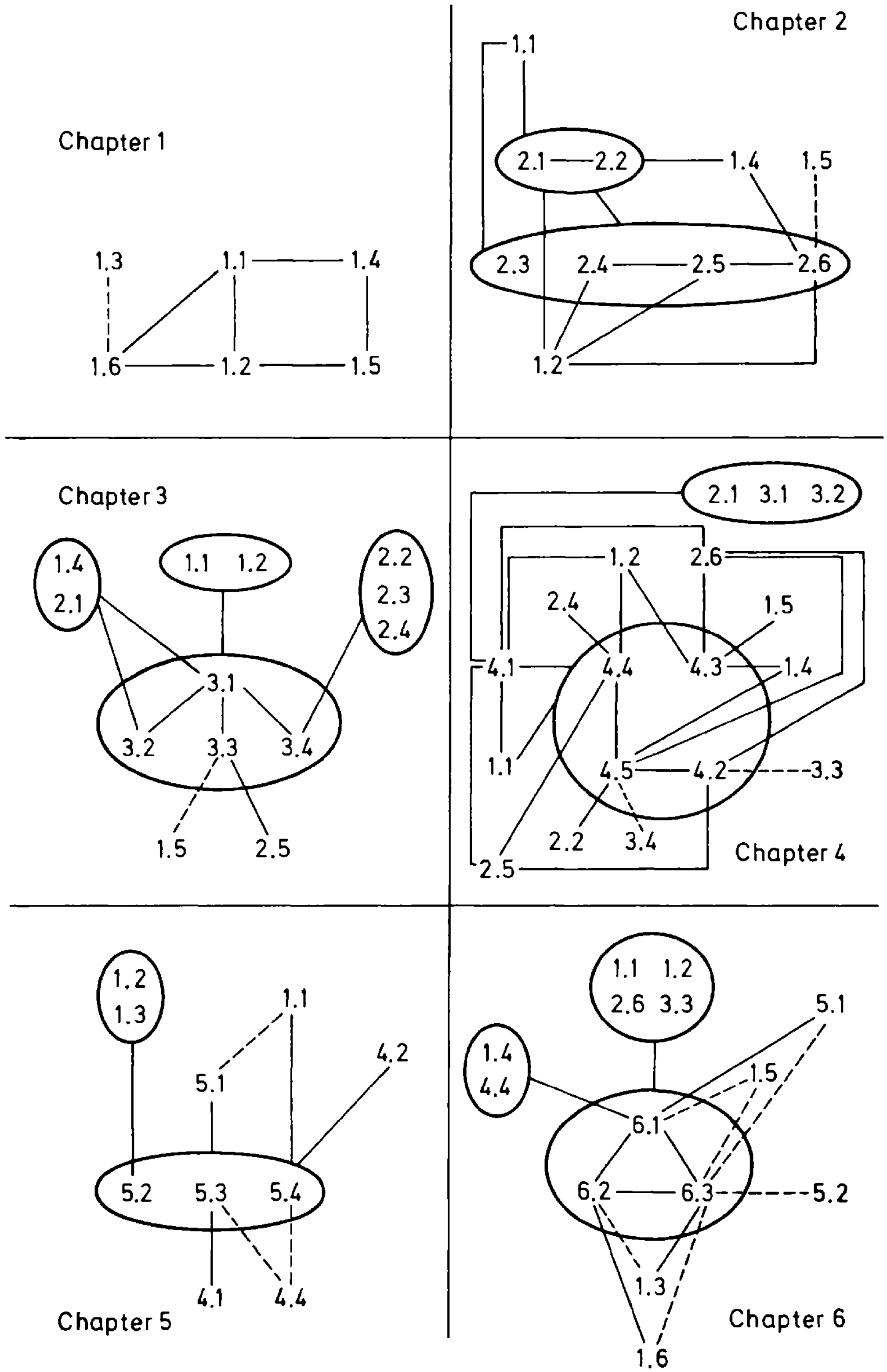
2. On the column corresponding to a given section on the horizontal axis we have indicated all the sections which depend on it.

3. The symbol ● indicates a main dependence while the symbol ○ denotes a minor dependence.

4. A section which depends on another one is also dependent on all sections depending on it, even though this is not explicitly indicated.
So, an empty box corresponding to two sections does not mean necessarily that the two sections are independent.

	1.1	1.2	1.3	1.4	1.5	1.6	2.1	2.2	2.3	2.4	2.5	2.6	3.1	3.2	3.3	3.4	4.1	4.2	4.3	4.4	4.5	5.1	5.2	5.3	5.4	6.1	6.2	6.3
1.1																												
1.2	●																											
1.3																												
1.4	●																											
1.5	●	●		●																								
1.6	●	●	○																									
2.1	●	●		●																								
2.2	●	●		●			●																					
2.3		●					●	●																				
2.4	●	●					●	●																				
2.5	●	●					●	●		●																		
2.6	●	●		●	○		●	●			●																	
3.1	●	●		●			●																					
3.2	●	●		●			●						●															
3.3	●	●			○						●	●	●															
3.4	●	●						●	●	●		●	●															
4.1	●	●					●				●	●	●	●														
4.2	●										●	●			○		●											
4.3	●	●		●	●							●					●											
4.4	●	●								●	●						●											
4.5	●			●				●				●				○	●	●		●								
5.1	○																											
5.2	●	●	●															●				●						
5.3	●																●	●		○		●						
5.4	●																	●		○		●						
6.1	●	●		●	○							●			●					●		●						
6.2	●	●	○			●						●			●											●		
6.3	●	●	●		○	○						●			●							○	○			●	●	

Legend: In each square we have considered the interdependence among the sections of a chapter and all the preceding ones.
A dotted line indicates a minor dependence.



Notation

We denote by $\mathbb{N}$ the set of natural numbers $1, 2, \ldots$, $\mathbb{Z}$ the set of integers, $\mathbb{Q}$ the set of rational numbers, $\mathbb{R}$ the set of real numbers and $\mathbb{C}$ the set of complex numbers, endowed with their usual topology.

The letter $\mathbb{K}$ stands either for the field $\mathbb{R}$ or for the field $\mathbb{C}$. We also denote by $\mathbb{N}_0$ the set $\mathbb{N} \cup \{0\}$.

When $+\infty$ and $-\infty$ are added to $\mathbb{R}$, then we obtain the extended real line $\tilde{\mathbb{R}}$.

If $x_1, \ldots, x_p$ are in $\mathbb{K}$, the product $x_1 \ldots x_p$ is sometimes denoted by $\prod_{i=1}^{p} x_i$.

A real number x is called *positive* (*strictly positive*, respectively) if $x \geq 0$ ($x > 0$, respectively). The symbols $\mathbb{Z}_+$, $\mathbb{Q}_+$ and $\mathbb{R}_+$ denote the subsets formed by the positive elements of the respective sets. If $z = x + iy \in \mathbb{C}$, then $\bar{z} := x - iy$, $\mathcal{R}e\, z := x$, $\mathcal{I}m\, z := y$ and $|z| := \sqrt{x^2 + y^2}$ denote the *conjugate*, the *real part*, the *imaginary part* and the *modulus* of z, respectively. Here and in the sequel we adopt the notation $A := B$ to signify that the symbol A is used to denote the object B or that the objects A and B are equal by definition.

For given real numbers a and b, $a < b$, the *intervals* $[a, b]$ ($[a, b[$, $]a, b]$, $]a, b[$, respectively) are the subsets of all $x \in \mathbb{R}$ satisfying $a \leq x \leq b$ ($a \leq x < b$, $a < x \leq b$, $a < x < b$, respectively).

The *empty set* is denoted by $\emptyset$. If X and Y are sets, then the notation $X \subset Y$ means that X is a subset of Y and the case $X = Y$ is not excluded.

As usual, the symbols $X \cup Y$ and $X \cap Y$, or $\bigcup_{i \in I} X_i$ and $\bigcap_{i \in I} X_i$, respectively, are used to denote the *union* and the *intersection* of two sets X and Y or of a family $(X_i)_{i \in I}$ of sets, respectively.

Given two sets X and Y, the symbols $X \backslash Y$ stands for the set of all elements $x \in X$ such that $x \notin Y$. $\mathrm{Card}(X)$ denotes the *cardinality* of a finite set X.

We denote by $(a_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}$ a *matrix* of objects with n rows and m columns; if $m = n$ and $a_{ij} \in \mathbb{K}$ for every $i, j = 1, \ldots, n$, we denote the *determinant* of the matrix by $\begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix}$ or by $\det(a_{ij})$.

The *Kronecker symbol* δ_{ij} stands for 1 if $i = j$ and 0 otherwise.

If X is a topological space and $A \subset X$, we use the symbols $\bar{A}$, $\mathring{A}$ and ∂A to denote the *closure*, the *interior* and the *boundary* of A, respectively. Thus, $\partial A := \bar{A} \cap (\overline{X \backslash A}) = \bar{A} \backslash \mathring{A}$.

If $\bar{A} = X$, then we say that A is *dense* in X or that A is *everywhere dense*.

If X is a vector space over the field $\mathbb{K}$ and $A \subset X$, we denote by $\mathscr{L}(A)$ the *linear subspace generated by* A, i.e., the intersection of all linear subspaces containing A.

If $A \subset X$ and $a \in X$, the set $A + a := \{x + a | x \in A\}$ denotes a *translate* of A. Moreover $-A := \{-x | x \in A\}$ and, if $B \subset X$, $A + B := \{x + y | x \in A, y \in B\}$ and $A - B := \{x - y | x \in A, y \in B\}$.

If E is a function space and $A \subset E$, we also put $A + \mathbb{R}_+ := \{f + \alpha | f \in A, \alpha \in \mathbb{R}_+\}$.

The symbol $\dim(E)$ denotes the algebraic dimension of a vector space E.

The *cartesian product* $X \times Y$ of two sets X and Y is the set of all ordered pairs (x, y) with $x \in X$ and $y \in Y$.

More generally, if $(X_i)_{i \in I}$ is a family of sets, the *cartesian product* $\prod_{i \in I} X_i$ of the family consists of all families $(x_i)_{i \in I}$ where $x_i \in X_i$ for every $i \in I$.

If I is finite, say $I = \{1, \ldots, p\}$, then the cartesian product $\prod_{i=1}^{p} X_i$ is often identified with the set of all p-uples $(x_i, \ldots, x_p)$ of elements where $x_i \in X_i$ $(1 \leq i \leq p)$.

The cartesian product of p copies of the same set X is denoted by X^p.

The *unit circle* $\mathbb{T}$ and the *unit disk* $\mathbb{D}$ are, respectively, the subsets $\mathbb{T} := \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 = 1\} = \{z \in \mathbb{C} | |z| = 1\}$ and $\mathbb{D} := \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\} = \{z \in \mathbb{C} | |z| \leq 1\}$. The *unit sphere* in $\mathbb{R}^{p+1}$ is denoted by $\mathbb{S}_p$.

If $(X_i)_{i \in I}$ is a family of topological spaces, then the cartesian product $\prod_{i \in I} X_i$ endowed with the product topology is called the *product space* of the family $(X_i)_{i \in I}$.

As usual, the symbol $f: X \to Y$ denotes a *mapping* from a set X into a set Y. Sometimes we also use the symbol $x \mapsto f(x)$. If $f: X \to Y$ is a mapping, for every $A \subset X$ and $B \subset Y$ we set

$$f(A) := \{y \in Y | \text{there exists } x \in A \text{ such that } f(x) = y\}$$

and

$$f^{-1}(B) := \{x \in X | f(x) \in B\}.$$

The subsets $f(A)$ and $f^{-1}(B)$ are called the *image* of A and the *inverse image of* B *under* f, respectively.

More generally, the symbol $f: D_X(f) \to Y$ stands for a mapping from a subset $D_X(f)$ of X into Y. The subset $D_X(f)$ is called the *domain* of the mapping f and the image $f(D_X(f))$ of $D_X(f)$ is called the *range* of f. If $Y = X$, $D_X(f)$ will be simply denoted by $D(f)$.

Given a mapping $f: X \to Y$ and a subset A of X, the restriction of f to A is denoted by $f_{|A}$. Moreover, considering another mapping $g: Y \to Z$ from Y into a set Z, the *composition* of f and g is denoted by $g \circ f$. The p-th power $(p \geq 1)$ of a mapping $f: X \to X$ is defined as

$$f^p = \begin{cases} f, & \text{if } p = 1, \\ f^{p-1} \circ f, & \text{if } p \geq 2. \end{cases}$$

Generally, throughout the book the mappings are denoted by small letters, say $f, g, h, \ldots$. However, when we deal with random variables acting on probability spaces as well as with linear mappings (almost always called *linear operators*) acting on vector spaces, we use capital letters, say $Y, Z, \ldots$, and $L, S, T, \ldots$, respectively.

Sometimes for a given linear mapping $T: E \to F$ acting from a vector space E into a vector space F, the value of T at a point $f \in E$ is denoted by Tf instead of $T(f)$, if no confusion can arise. Furthermore the composition of T with another linear mapping $S: F \to G$ is also denoted by ST instead of $S \circ T$.

A linear mapping from a vector space (over $\mathbb{K}$) into $\mathbb{K}$ is also called a *linear form* or a *functional* on E.

A mapping $f: X \to Y$ is said to be *injective* if $f(x) = f(y)$ implies $x = y$ for every x, $y \in X$. If $f(X) = Y$, i.e., for every $y \in Y$ there exists $x \in X$ such that $f(x) = y$, we say that f is *surjective.*

If f is both injective and surjective, we say that f is *bijective.*

A mapping $f: X \to Y$ is bijective if and only if is *invertible*, i.e., there exists a (unique) mapping $g: Y \to X$ such that $g(f(x)) = x$ and $f(g(y)) = y$ for every $x \in X$ and $y \in Y$.

The mapping g is called the *inverse* of f and it is denoted by f^{-1}.

A mapping from a set X into $\mathbb{R}$ ($\tilde{\mathbb{R}}$, $\mathbb{C}$, respectively) is called a *real-valued* (*numerical*, *complex-valued*, respectively) *function on* X. When we simply speak of a *function on* X we always mean a real-valued function on X.

If A is contained in a set X, then the *characteristic function* of A is the function $\mathbf{1}_A: X \to \mathbb{R}$ defined by putting for every $x \in X$

$$\mathbf{1}_A(x) := \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

The constant function on X of constant value 1 is denoted by $\mathbf{1}$.

Giving two functions $f: X \to \mathbb{R}$ and $g: X \to \mathbb{R}$, we use the symbols $\sup(f, g)$ and $\inf(f, g)$ to denote the functions on X defined by putting

$$\sup(f, g)(x) := \sup(f(x), g(x)),$$

$$\inf(f, g)(x) := \inf(f(x), g(x))$$

for every $x \in X$. These functions are also denoted by $f \vee g$ and $f \wedge g$, respectively.

More generally, if $f_1, \ldots, f_p$ are real functions on X we define the functions $\sup_{1 \le i \le p} f_i \colon X \to \mathbb{R}$ and $\inf_{1 \le i \le p} f_i \colon X \to \mathbb{R}$ by

$$\left(\sup_{1 \le i \le p} f_i\right)(x) := \sup_{1 \le i \le p} f_i(x), \quad \left(\inf_{1 \le i \le p} f_i\right)(x) := \inf_{1 \le i \le p} f_i(x) \quad (x \in X).$$

These functions are also denoted by $f_1 \vee \cdots \vee f_p$ and $f_1 \wedge \cdots \wedge f_p$, respectively.

If $f\colon X \to \mathbb{R}$, the *positive part* f^+ of f, the *negative* part f^- of f and the *absolute value* $|f|$ of f are defined as

$$f^+ := \sup(f, 0), \quad f^- := \sup(-f, 0), \quad |f| := \sup(f, -f).$$

Clearly we have $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

If $f_1, \ldots, f_p$ are functions on X, sometimes we use the symbol $\prod_{i=1}^{p} f_i$ to denote the real-valued function on X defined by

$$\left(\prod_{i=1}^{p} f_i\right)(x) := \prod_{i=1}^{p} f_i(x) = f_1(x) \ldots f_p(x) \quad (x \in X).$$

A function $f\colon X \to \mathbb{R}$ is said to be *positive* if $f(x) \ge 0$ for every $x \in X$. Moreover we say that f is *strictly positive* if $f(x) > 0$ for every $x \in X$.

If $f\colon X \to \mathbb{R}$ and $g\colon X \to \mathbb{R}$ are functions on X, we write $f \le g$ ($f < g$, respectively) if $f(x) \le g(x)$ ($f(x) < g(x)$, respectively) for every $x \in X$.

A real function $f\colon I \to \mathbb{R}$ defined on a real interval I is called *increasing* (*decreasing, strictly increasing, strictly decreasing*, respectively) if $f(x) \le f(y)$ ($f(y) \le f(x)$, $f(x) < f(y)$, $f(y) < f(x)$, respectively) for every $x, y \in I$ satisfying $x < y$. A *monotone* (*strictly monotone*, respectively) function is a function which is indifferently increasing or decreasing (strictly increasing or strictly decreasing, respectively).

The symbols o and O are the usual Landau symbols. Thus, if $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ are sequences of real numbers, the symbol $x_n = o(y_n)$, $n \to \infty$, means that $x_n/y_n \to 0$ as $n \to \infty$, while the symbol $x_n = O(y_n)$, $n \to \infty$, means that there exists a constant $M > 0$ such that $|x_n/y_n| \le M$ for every $n \in \mathbb{N}$.

A similar meaning must be attributed to the symbols $f(x) = o(g(x))$, $x \to x_0$, and $f(x) = O(g(x))$, $x \to x_0$, where f and g are functions defined on a subset X of a topological space and x_0 is a limit point for X.

The symbols $\Rightarrow$, $\Leftarrow$, $\Leftrightarrow$ stand for the usual logical implication symbols. Thus (a) $\Rightarrow$ (b), (a) $\Leftarrow$ (b) and (a) $\Leftrightarrow$ (b) mean that statement (a) implies statement (b), statement (b) implies statement (a) and statement (a) is equivalent to statement (b), respectively.

The section numbered, say, by a.b is the b-th section of Chapter a.

Definitions, lemmas, propositions and theorems are numbered by three digits, say a.b.c, where a denotes the number of the chapter, b that of the section and c is the progressive number within the section.

Formulas are numbered by an index of the form (a.b.c) where the digits a, b, c have the above specified meaning.

Sections, formulas, theorems, definitions, etc., are referred to by their corresponding numbers.

The end of a proof is indicated by the symbol $\square$.

Chapter 1

Preliminaries

The main aim of this introductory chapter is to present the general notation and definitions we shall use throughout the book.

To make the exposition self-contained we also review those prerequisites which are necessary for a full understanding.

The topics are Radon measures, locally convex spaces and some basic aspects of general topology.

They have been selected primarily in view of our needs and are presented without pretence of completeness and without proofs.

Throughout the book we shall attempt to give various applications of Korovkin-type approximation theory. However some of them require a solid background also from other branches of analysis, e.g., measures on topological spaces, integration theory with respect to Radon measures, basic principles of probability theory, Choquet's integral representation theory and C_0-semigroups of bounded linear operators.

For the sake of completeness we also review the definitions and results pertaining to these topics and we include them in starred sections.

Thus a starred section or subsection is not essential for the whole book but will be only used for a particular (important) application of the Korovkin-type approximation theory.

1.1 Topology and analysis

In this section we present the definitions and the main properties of compact and locally compact spaces. We shall also introduce the main function spaces which we shall be concerned with in the sequel. For more details see Bourbaki [1965] and Engelking [1980] or the short and elegant Section 7.4 of Bauer [1981].

We begin by recalling the notions of net and filter.

A *filter* on a set I is a collection $\mathscr{F}$ of non-empty subsets of I, which is closed under the formation of finite intersections and such that, if $F \in \mathscr{F}$ and $F \subset G \subset I$, then $G \in \mathscr{F}$.

If $(I, \leq)$ is a *directed set*, i.e., $\leq$ is a partial ordering on I such that for every $i, j \in I$ there exists $\lambda \in I$ satisfying $i \leq \lambda$ and $j \leq \lambda$, then the set $\mathscr{F}_{\leq}$ of all subsets F of I for which there exists $i_0 \in I$ such that $\{i \in I \mid i_0 \leq i\} \subset F$, is a filter on I and it is called the *filter of sections on* I.

If $\mathscr{F}_1$ and $\mathscr{F}_2$ are filters on I, we shall say that $\mathscr{F}_2$ is *finer* than $\mathscr{F}_1$ if $\mathscr{F}_1 \subset \mathscr{F}_2$ or, equivalently, if for every $F_1 \in \mathscr{F}_1$ there exists $F_2 \in \mathscr{F}_2$ such that $F_2 \subset F_1$.

A *filtered family* $(x_i)_{i \in I}^{\mathscr{F}}$ of a set X is a family $(x_i)_{i \in I}$ of elements of X such that on the index set I there is fixed a filter $\mathscr{F}$.

A *net* (or *generalized sequence*) on X is a family $(x_i)_{i \in I}^{\leq}$ of elements of X such that on the set I there is a partial ordering $\leq$ with respect to which $(I, \leq)$ is a directed set.

Given a topological space X, we say that a filtered family $(x_i)_{i \in I}^{\mathscr{F}}$ *converges to a point* $x_0 \in X$ if for every neighborhood V of x_0 there exists $F \in \mathscr{F}$ such that $x_i \in V$ for each $i \in F$. In this case, x_0 is called a *limit* of the filtered family.

A filtered family is said to be *convergent* if it converges to some point.

If X is a *Hausdorff space*, i.e., for every pair of distinct points $x_1, x_2 \in X$ there exist neighborhoods V_1 and V_2 of x_1 and x_2, respectively, such that $V_1 \cap V_2 = \varnothing$, then every convergent filtered family $(x_i)_{i \in I}^{\mathscr{F}}$ converges to a unique limit $x_0 \in X$. In this case we shall write

$$\lim_{\substack{\mathscr{F} \\ i \in I}} x_i = x_0. \tag{1.1.1}$$

Sometimes, the notation $x_i \xrightarrow{\mathfrak{F}} x_0$ will be also used, if no confusion can arise.

If $(x_i)_{i \in I}^{\leq}$ is a net, we say that $(x_i)_{i \in I}^{\leq}$ converges to a point $x_0 \in X$, if $(x_i)_{i \in I}^{\mathscr{F}_{\leq}}$ converges to x_0. Explicitly this means that for every neighborhood V of x_0 there exists $i_0 \in I$ such that $x_i \in V$ for every $i \in I$, $i \geq i_0$. The point x_0 is called a *limit* of the net $(x_i)_{i \in I}^{\mathfrak{F}}$.

If X is Hausdorff, then x_0 is unique and we shall write

$$\lim_{\substack{\leq \\ i \in I}} x_i = x_0, \quad \text{or} \quad x_i \to x_0. \tag{1.1.2}$$

If X and Y are topological spaces, then a mapping $f\colon X \to Y$ is continuous at a point $x_0 \in X$ if and only if for every net $(x_i)_{i \in I}^{\leq}$ in X which converges to x_0, $(f(x_i))_{i \in I}^{\leq}$ converges to $f(x_0)$.

Given a topological space X, a numerical function $f\colon X \to \mathbb{R} \cup \{+\infty\}$ ($f\colon X \to \mathbb{R} \cup \{-\infty\}$, respectively) is said to be *lower semi-continuous* (*upper semi-continuous*, respectively) *at a point* $x_0 \in X$ if for every $\lambda \in \mathbb{R}$ satisfying $f(x_0) > \lambda$ ($f(x_0) < \lambda$, respectively) there exists a neighborhood V of x_0 such that $f(x) > \lambda$ ($f(x) < \lambda$, respectively) for every $x \in V$.

A function f is said to be *lower semi-continuous* (*upper semi-continuous*, respectively) if it is lower semi-continuous (upper semi-continuous, respectively) at every point of X.

In fact, f is lower semi-continuous (upper semi-continuous, respectively) if and only if the subset $\{x \in X \mid f(x) > \lambda\}$ is open for every $\lambda \in \mathbb{R}$ (the subset $\{x \in X \mid f(x) < \lambda\}$ is open for every $\lambda \in \mathbb{R}$, respectively) or, equivalently, the subset $\{x \in X \mid f(x) \leq \lambda\}$ is closed for every $\lambda \in \mathbb{R}$ (the subset $\{x \in X \mid f(x) \geq \lambda\}$ is closed for every $\lambda \in \mathbb{R}$, respectively).

Moreover, a function $f: X \to \mathbb{R}$ is continuous if and only if it is both lower and upper semi-continuous.

A topological space X is called a *compact space* if every open cover of X has a finite subcover.

A subset of a topological space is said to be *compact* (*relatively compact*, respectively) if it is compact in the relative topology (if its closure is compact, respectively).

A useful characterization of compact spaces may be stated in terms of filtered families. More precisely, a topological space X is compact if and only if for every filtered family $(x_i)_{i \in I}^{\mathscr{F}}$ in X there exists a filter $\mathscr{F}'$ on I finer than $\mathscr{F}$ such that $(x_i)_{i \in I}^{\mathscr{F}'}$ is convergent in X.

If X is *metrizable*, i.e., the topology of X is induced by a metric on X, then X is compact if and only if every sequence of points of X admits a convergent subsequence.

Every compact metrizable space is *separable*, i.e., X contains a dense countable subset.

Let X be a compact Hausdorff space and denote by $\mathscr{T}_1$ its topology. If $\mathscr{T}_2$ is another Hausdorff topology on X such that $\mathscr{T}_2 \subset \mathscr{T}_1$, then necessarily $\mathscr{T}_2 = \mathscr{T}_1$.

Every compact Hausdorff space X is *normal*, i.e., each closed subset of X possesses a fundamental system of closed neighborhoods.

In this case, the *Tietze's extension theorem* holds (see Choquet [1969, Theorem 6.1]).

1.1.1 Theorem (*Tietze*). *If X is a normal space, then every continuous function from a closed subset of X is continuously extendable to X.*

Finally, we recall that, if X is compact, then every increasing (decreasing, respectively) sequence of lower semi-continuous (upper semi-continuous, respectively) functions on X converging pointwise to a continuous function, converges uniformly on X as well (*Dini's theorem*) (for a proof see Engelking [1989, Lemma 3.2.18]).

A topological space X is said to be *locally compact* if each of its point possesses a compact neighborhood.

In fact, if X is locally compact and Hausdorff, then each point of X has a fundamental system of compact neighborhoods.

The spaces $\mathbb{R}^p$, $p \geq 1$, as well as discrete spaces and compact spaces are examples of locally compact spaces.

Every locally compact Hausdorff space which is *countable at infinity* (i.e., it is the union of a sequence of compact subsets of X), is normal.

Furthermore, a locally compact Hausdorff space which has a *countable base* (i.e., there exists a countable family of open subsets such that every open subset is the union of some subfamily of this countable family) is necessarily metrizable, complete (and, hence, normal) and separable.

Conversely, a metrizable locally compact Hausdorff space which is countable at infinity has a countable base and, hence, is separable.

For technical reasons, it will be often useful to consider the (*Alexandrov*) *one-point-compactification* X_ω *of* X, which is defined as $X_\omega := X \cup \{\omega\}$, where ω is an object which does not belong to X (ω is often called the *point at infinity of* X).

The topology $\mathscr{T}_\omega$ on X_ω is defined as

$$\mathscr{T}_\omega := \mathscr{T} \cup \{X_\omega \setminus K \mid K \subset X, K \text{ compact}\}, \tag{1.1.3}$$

where $\mathscr{T}$ denotes the topology on X.

The topological space $(X_\omega, \mathscr{T}_\omega)$ is a Hausdorff compact space and X is open in X_ω. Furthermore, if X is non compact, then X is dense in X_ω.

For example, the one-point-compactification of $\mathbb{R}^p$, $p \geq 1$, is homeomorphic to the unit sphere of $\mathbb{R}^{p+1}$.

If a net $(x_i)_{i \in I}^{\leq}$ of elements of X converges to ω in X_ω, we say that the net $(x_i)_{i \in I}^{\leq}$ *converges to the point at infinity of* X. This means that for every compact subset K of X there exists $i_0 \in I$ such that $x_i \in X \setminus K$ for every $i \in I$, $i \geq i_0$.

Analogously, if $f\colon X \to \mathbb{R}$ and $\alpha \in \mathbb{R}$, we say that f *converges to* α *at the point at infinity*, if for every $\varepsilon > 0$ there exists a compact subset K of X such that $|f(x) - \alpha| \leq \varepsilon$ for every $x \in X \setminus K$. If this is the case, we shall write

$$\lim_{x \to \omega} f(x) = \alpha. \tag{1.1.4}$$

Given a set X, we shall denote by $\mathscr{B}(X)$ the Banach space of all real-valued bounded functions defined on X, endowed with the norm of the *uniform convergence* (briefly, the *sup-norm*) defined by

$$\|f\| := \sup_{x \in X} |f(x)| \quad \text{for every } f \in \mathscr{B}(X). \tag{1.1.5}$$

If X is a topological space, $\mathscr{C}(X)$ denotes the space of all real-valued continuous functions on X. Furthermore, we set

$$\mathscr{C}_b(X) := \mathscr{C}(X) \cap \mathscr{B}(X). \tag{1.1.6}$$

The space $\mathscr{C}_b(X)$, endowed with the sup-norm, is a Banach space.

Given a locally compact space X, we shall denote by $\mathscr{C}_0(X)$ the space of all functions $f \in \mathscr{C}(X)$ which *vanish at infinity*.

A function $f \in \mathscr{C}(X)$ vanishes at infinity if for every real number $\varepsilon > 0$ the set $\{x \in X \mid |f(x)| \geq \varepsilon\}$ is compact or, equivalently, if for every real number $\varepsilon > 0$ there exists a compact subset K of X such that $|f(x)| \leq \varepsilon$ for every $x \in X \setminus K$.

In other words, $f \in \mathscr{C}_0(X)$ if and only if $\lim_{x \to \omega} f(x) = 0$ or, what is the same, the function $\tilde{f}: X_\omega \to \mathbb{R}$ defined by

$$\tilde{f}(x) := \begin{cases} f(x), & \text{if } x \in X, \\ 0, & \text{if } x = \omega, \end{cases} \tag{1.1.7}$$

is continuous in X_ω.

Sometimes, the above function $\tilde{f}$ is called the *canonical extension* of f to X_ω.

Note that if $f \in \mathscr{C}_0(X)$ and $(x_i)_{i \in I}^{\leq}$ is a net of elements of X converging to the point at infinity of X, then $(f(x_i))_{i \in I}^{\leq}$ converges to 0.

Clearly, if X is compact, then $\mathscr{C}_0(X) = \mathscr{C}(X) = \mathscr{C}_b(X)$.

In general, $\mathscr{C}_0(X)$ is a closed subspace of $\mathscr{C}_b(X)$ and, hence, endowed with the sup-norm, it is a Banach space. Unless otherwise stated, we shall always consider the space $\mathscr{C}_0(X)$ endowed with this norm.

On $\mathscr{C}_0(X)$ we shall also consider the natural ordering induced by the cone

$$\mathscr{C}_0^+(X) := \{f \in \mathscr{C}_0(X) \mid f(x) \geq 0 \text{ for every } x \in X\}. \tag{1.1.8}$$

If $f \in \mathscr{C}_0(X)$, then $f^+, f^-, |f| \in \mathscr{C}_0(X)$ and

$$\|f\| = \||f|\| = \max\{\|f^+\|, \|f^-\|\}. \tag{1.1.9}$$

Thus $\mathscr{C}_0(X)$ is a Banach lattice. We recall, indeed, that a *normed lattice* E is a vector lattice (see Section 1.4) endowed with a *lattice norm* $\|\cdot\|$, i.e.,

$$|f| \leq |g| \quad \Rightarrow \quad \|f\| \leq \|g\| \quad \text{for every } f, g \in E. \tag{1.1.10}$$

If E is a Banach space for a lattice norm, then we say that E is a *Banach lattice*.

Other than $\mathscr{C}_0(X)$, standard examples of Banach lattices are the spaces $\mathscr{C}_b(X)$, $\mathscr{B}(X)$ and $L^p(X, \mu)$, $1 \leq p \leq +\infty$ (see Section 1.2).

An important property of Banach lattices concerns the automatic continuity of positive linear operators acting on them.

More precisely, if E is a Banach lattice and if F is another normed lattice, then every *positive* linear operator $T: E \to F$ (i.e., $T(f) \geq 0$ for every $f \in E$, $f \geq 0$) is continuous.

Furthermore, if $E = \mathscr{C}(X)$, X compact, then

$$\|T\| = \|T(\mathbf{1})\|, \tag{1.1.11}$$

where $\mathbf{1}$ denotes the constant function 1.

In particular every *positive linear form* on a Banach lattice E, i.e., every positive linear mapping from E into $\mathbb{R}$, is continuous.

Another function space playing a fundamental role in this book is the space $\mathscr{K}(X)$ of all real-valued continuous functions $f\colon X \to \mathbb{R}$ whose *support*

$$\operatorname{Supp}(f) := \overline{\{x \in X \mid f(x) \neq 0\}} \tag{1.1.12}$$

is compact.

In other words, a continuous function $f\colon X \to \mathbb{R}$ belongs to $\mathscr{K}(X)$ if it vanishes on the complement of a suitable compact subset of X.

The space $\mathscr{K}(X)$ is dense in $\mathscr{C}_0(X)$ and, if X is compact, obviously coincides with $\mathscr{C}(X)$.

Besides the topology induced by the sup-norm, on $\mathscr{K}(X)$ there are other important locally convex topologies, such as the inductive topology and the projective topology. However, throughout the book we shall only make use of the first one. We refer to Choquet [1969, Section 16] for a detailed study of the properties of the other topologies.

If X is a locally compact Hausdorff space, then there are sufficiently many functions in $\mathscr{K}(X)$. This is a consequence of the following result, whose proof can be found for instance in Bauer [1981, Lemma 7.4.2].

1.1.2 Theorem. *For every compact subset K of X and for every open subset U containing K, there exists $g \in \mathscr{K}(X)$ such that $0 \leq g \leq 1$, $g = 1$ on K and $\operatorname{Supp}(g) \subset U$.*

1.2 Radon measures

Radon measures are a powerful tool which is fruitfully used in several branches of analysis such as probability theory, potential theory and integral representation theory. In this book they play a central role.

For more details the reader is referred to Bourbaki [1969] or Choquet [1969]. For a modern approach see also Bauer [1992] and Anger and Portenier [1992].

A *Radon measure* on a locally compact Hausdorff space X is a linear form $\mu\colon \mathscr{K}(X) \to \mathbb{R}$ satisfying the following property:

> For any compact subset K of X there exists $M_K \geq 0$ such that $|\mu(f)| \leq M_K \|f\|$ for every $f \in \mathscr{K}(X)$ having its support contained in K.

The space of all Radon measures on X will be denoted by $\mathscr{M}(X)$. Thus $\mathscr{M}(X)$ is the dual space of the locally convex space $\mathscr{K}(X)$ endowed with the inductive topology.

As a matter of fact, in this book we shall restrict ourselves mainly to *bounded Radon measures*. They are those Radon measures $\mu \in \mathscr{M}(X)$ which are continuous with respect to the sup-norm. In this case the *norm* of μ is defined to be the number

$$\|\mu\| := \sup\{|\mu(f)| \mid f \in \mathscr{K}(X), \|f\| \leq 1\}. \tag{1.2.1}$$

A bounded Radon measure μ is said to be *contractive* if $\|\mu\| \leq 1$.

The space of all bounded Radon measures will be denoted by $\mathscr{M}_b(X)$.

Furthermore, we shall denote by $\mathscr{M}^+(X)$ the cone of all positive Radon measures. Thus, $\mu \in \mathscr{M}^+(X)$ if $\mu \in \mathscr{M}(X)$ and $\mu(f) \geq 0$ for every $f \in \mathscr{K}(X)$, $f \geq 0$. In fact, every positive linear form on $\mathscr{K}(X)$ is automatically in $\mathscr{M}^+(X)$.

Finally we set

$$\mathscr{M}_b^+(X) := \mathscr{M}^+(X) \cap \mathscr{M}_b(X), \tag{1.2.2}$$

and

$$\mathscr{M}_1^+(X) := \{\mu \in \mathscr{M}_b^+(X) \mid \|\mu\| = 1\}. \tag{1.2.3}$$

The elements of $\mathscr{M}_1^+(X)$ are also called *probability Radon measures*.

It is easy to see that every $\mu \in \mathscr{M}_b(X)$ ($\mu \in \mathscr{M}_b^+(X)$, respectively) can be extended to a (unique) continuous (positive, respectively) linear form on $\mathscr{C}_0(X)$ that we shall continue to denote by μ.

If X is compact, then $\mathscr{M}(X) = \mathscr{M}_b(X)$ (and, hence, $\mathscr{M}^+(X) = \mathscr{M}_b^+(X)$). Moreover, for every $\mu \in \mathscr{M}^+(X)$, $\|\mu\| = \mu(\mathbf{1})$. Conversely, if $\mu \in \mathscr{M}(X)$ and $\|\mu\| = \mu(\mathbf{1})$, then $\mu \in \mathscr{M}^+(X)$.

Another simple but useful property which is satisfied by every positive Radon measure $\mu \in \mathscr{M}^+(X)$ is the so-called *Cauchy-Schwarz inequality*, i.e.,

$$\mu(|fg|) \leq \sqrt{\mu(f^2)\mu(g^2)}, \tag{1.2.4}$$

which holds for every $f, g \in \mathscr{K}(X)$ (respectively, for every $f, g \in \mathscr{C}_0(X)$ provided $\mu \in \mathscr{M}_b^+(X)$).

As a matter of fact the same inequality holds by replacing $\mathscr{K}(X)$ (or $\mathscr{C}_0(X)$) with an arbitrary vector sublattice E of continuous functions on X and by considering a positive linear form $L: E \to \mathbb{R}$ on E (we recall that a linear subspace E of a vector lattice F is said to be a *sublattice* of F if for every $f, g \in E$ the supremum and the infimum of f and g in F lie in E). In this case one has

$$L(|fg|) \leq \sqrt{L(f^2)L(g^2)}, \tag{1.2.5}$$

for every $f, g \in E$ such that $fg, f^2, g^2 \in E$.

The simplest examples of Radon measures on X are the Dirac measures.

More precisely, given $x \in X$, the *Dirac measure* at x is the (bounded) Radon measure ε_x defined by

$$\varepsilon_x(f) := f(x) \qquad \text{for every } f \in \mathscr{K}(X) \text{ (or, } f \in \mathscr{C}_0(X)). \tag{1.2.6}$$

In fact, $\varepsilon_x \in \mathscr{M}_1^+(X)$.

A linear combination of Dirac measures is called a *discrete measure* on X. Thus, discrete measures are those bounded Radon measures on X of the form

$$\mu := \sum_{i=1}^{n} \lambda_i \varepsilon_{x_i}, \tag{1.2.7}$$

where $n \geq 1$, $x_1, \ldots, x_n \in X$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. In this case μ is positive if and only if every λ_i is positive.

Furthermore

$$\|\mu\| = \sum_{i=1}^{n} |\lambda_i|. \tag{1.2.8}$$

Other important examples of Radon measures can be constructed as follows.

Let $\mu \in \mathscr{M}(X)$ and consider $g \in \mathscr{C}(X)$. Then, the linear form $v: \mathscr{K}(X) \to \mathbb{R}$ defined by

$$v(f) := \mu(f \cdot g) \quad \text{for every } f \in \mathscr{K}(X), \tag{1.2.9}$$

is a Radon measure on X. It is called the *measure with density g relative to μ* and it is denoted by $g \cdot \mu$.

If μ and g are bounded (respectively, positive), then $g \cdot \mu$ is bounded (respectively, positive) and

$$\|g \cdot \mu\| \leq \|g\| \, \|\mu\|. \tag{1.2.10}$$

By a simple method it is possible to extend every bounded positive Radon measure $\mu \in \mathscr{M}_b^+(X)$ to a positive Radon measure $\tilde{\mu}$ on the one-point-compactification X_ω of X. The measure $\tilde{\mu}$ is defined by

$$\tilde{\mu}(g) := \mu(g_{|X} - g(\omega)) + g(\omega)\,\|\mu\| \quad \text{for every } g \in \mathscr{C}(X_\omega). \tag{1.2.11}$$

Clearly we have $\|\tilde{\mu}\| = \|\mu\|$ and $\tilde{\mu}(\tilde{f}) = \mu(f)$ for every $f \in \mathscr{C}_0(X)$.

We say that a Radon measure $\mu \in \mathscr{M}(X)$ is *zero* on an open subset U of X if $\mu(f) = 0$ for every function $f \in \mathscr{K}(X)$ whose support is contained in U.

We shall denote by $\mathfrak{A}(\mu)$ the collection of all open subsets of X on which μ is zero.

The *support* of the measure μ is then defined to be the subset

$$\mathrm{Supp}(\mu) := X \setminus \bigcup_{U \in \mathfrak{A}(\mu)} U. \tag{1.2.12}$$

Thus $\mathrm{Supp}(\mu)$ is a closed subset of X and, in fact, is the complement of the largest open subset of X on which μ is zero.

Clearly a point $x_0 \in X$ belongs to $\mathrm{Supp}(\mu)$ if for every neighborhood V of x_0 there exists $f \in \mathscr{K}(X)$ such that $\mathrm{Supp}(f) \subset V$ and $\mu(f) \neq 0$.

On the other hand, $x_0 \notin \mathrm{Supp}(\mu)$ if there exists an open neighborhood V of x_0 on which μ is zero.

It is also clear that $\mathrm{Supp}(\mu) = \varnothing$ if and only if $\mu = 0$.

Here, we list some of the main properties of supports which we shall use later. For a proof see Bourbaki [1969, Chapter III, Section 2] and Choquet [1969, Section 11].

1.2.1 Theorem. *Let X be a locally compact Hausdorff space and let $\mu \in \mathscr{M}(X)$. Then*

(1) $\mathrm{Supp}(\mu + \upsilon) \subset \mathrm{Supp}(\mu) \cup \mathrm{Supp}(\upsilon)$ *for every $\upsilon \in \mathscr{M}(X)$. If μ and υ are positive we have equality in the above inclusion.*
(2) *If $f \in \mathscr{K}(X)$ and $f = 0$ on* $\mathrm{Supp}(\mu)$, *then $\mu(f) = 0$. If $\mu \in \mathscr{M}_b(X)$, then the same property holds for every $f \in \mathscr{C}_0(X)$.*
(3) *If $f, g \in \mathscr{K}(X)$ (or $f, g \in \mathscr{C}_0(X)$ provided $\mu \in \mathscr{M}_b(X)$) and $f = g$ on* $\mathrm{Supp}(\mu)$, *then $\mu(f) = \mu(g)$.*
(4) *If $\mu \in \mathscr{M}^+(X)$ and $f \in \mathscr{K}(X)$ (or $f \in \mathscr{C}_0(X)$ provided $\mu \in \mathscr{M}_b^+(X)$) then $\mu(f) \geq 0$ if $f \geq 0$ on* $\mathrm{Supp}(\mu)$.
(5) *If $\mu \in \mathscr{M}^+(X)$ and $f \in \mathscr{K}(X)$, $f \geq 0$ (or $f \in \mathscr{C}_0^+(X)$ provided $\mu \in \mathscr{M}_b^+(X)$) and if $\mu(f) = 0$, then $f = 0$ on* $\mathrm{Supp}(\mu)$.
(6) *For every $g \in \mathscr{C}(X)$,* $\mathrm{Supp}(g \cdot \mu) = \overline{\{x \in \mathrm{Supp}(\mu) | g(x) \neq 0\}} \subset \mathrm{Supp}(g) \cap \mathrm{Supp}(\mu)$.
(7) *If $x_1, \ldots, x_n$ are distinct points of X, $n \geq 1$, then* $\mathrm{Supp}(\mu) = \{x_1, \ldots, x_n\}$ *if and only if $\mu = \sum_{i=1}^{n} \lambda_i \varepsilon_{x_i}$ for some $\lambda_1, \ldots, \lambda_n \in \mathbb{R} \setminus \{0\}$.*
(8) *If* $\mathrm{Supp}(\mu)$ *is compact, then μ is bounded.*

On the space $\mathscr{M}(X)$ we shall consider the *vague topology* which is, by definition, the coarsest topology on $\mathscr{M}(X)$ for which all the mappings $\varphi_f (f \in \mathscr{K}(X))$ are continuous, where

$$\varphi_f(\mu) := \mu(f) \quad \text{for every } f \in \mathscr{K}(X) \text{ and } \mu \in \mathscr{M}(X). \tag{1.2.13}$$

It is a locally convex topology and, in fact, is the weak*-topology of the dual space of the locally convex space $\mathscr{K}(X)$ (see Section 1.4).

A net $(\mu_i)_{i\in I}^{\leq}$ in $\mathscr{M}(X)$ converges to a Radon measure $\mu \in \mathscr{M}(X)$ with respect to the vague topology if $\lim_{\substack{\leq \\ i\in I}} \mu_i(f) = \mu(f)$ for every $f \in \mathscr{K}(X)$. In this case we also say that $(\mu_i)_{i\in I}^{\leq}$ *converges vaguely* to μ.

Endowed with the vague topology, $\mathscr{M}(X)$ is a locally convex Hausdorff space. Furthermore in general $\mathscr{M}(X)$ is not metrizable. If X has a countable base, then $\mathscr{M}^+(X)$ is metrizable and separable.

A useful characterization of *vaguely compact subsets of* $\mathscr{M}(X)$ (i.e., those subsets which are compact with respect to the vague topology) is indicated below.

We also notice that a subset $\mathfrak{A}$ of $\mathscr{M}(X)$ is *vaguely bounded* if the set $\{\mu(f) | \mu \in \mathfrak{A}\}$ is bounded for every $f \in \mathscr{K}(X)$.

Moreover, $\mathfrak{A}$ is *strongly bounded* if for every compact subset K of X there exists $M_K \geq 0$ such that $|\mu(f)| \leq M_K \|f\|$ for every $\mu \in \mathfrak{A}$ and for every $f \in \mathscr{K}(X)$ satisfying $\mathrm{Supp}(f) \subset K$.

For a proof of the next result see Choquet [1969, Theorem 12.6].

1.2.2 Theorem. *A subset $\mathfrak{A}$ of $\mathscr{M}(X)$ is relatively vaguely compact in $\mathscr{M}(X)$ if and only if it is vaguely bounded or, equivalently, if and only if it is strongly bounded.*

From this criterion it follows that for every $r > 0$ the set $\{\mu \in \mathscr{M}_b(X) | \|\mu\| \leq r\}$ is vaguely compact in $\mathscr{M}(X)$. Moreover, if X is compact, then $\{\mu \in \mathscr{M}^+(X) | \|\mu\| = r\}$ is vaguely compact too.

Via the vague topology every Radon measure can be approximated by discrete measures. More precisely we have the following result, whose proof can be found again in Choquet [1969, Theorem 12.11].

1.2.3 Theorem (*Approximation theorem*). *Let X be a locally compact Hausdorff space. Then the following assertions hold*:

(1) *For every $\mu \in \mathscr{M}(X)$ there exists a net $(\mu_i)_{i\in I}^{\leq}$ of discrete Radon measures which converges vaguely to μ. Moreover, if μ is positive, every μ_i can be chosen positive too.*

(2) *If $\mu \in \mathscr{M}(X)$ and* $\mathrm{Supp}(\mu)$ *is compact, then there exists a net $(\mu_i)_{i\in I}^{\leq}$ of discrete Radon measures vaguely convergent to μ such that $\|\mu_i\| = \|\mu\|$ and* $\mathrm{Supp}(\mu_i) \subset \mathrm{Supp}(\mu)$ *for every $i \in I$.*

Furthermore, if μ is positive, the measures μ_i can also be chosen positive.

*Measures on topological spaces. The Riesz representation theorem

In this subsection we briefly discuss some properties of Borel and Baire measures on a topological space together with Riesz's representation theorem.

This last theorem will be used only in Section 5.2 to give a probabilistic interpretation of Korovkin's theorem.

In the sequel it is assumed that the reader has a thorough familiarity with the essential notions of measure theory and integration theory. As a reference the reader could consult Bauer [1981, part I], [1992].

If X is a topological space, we shall denote by $\mathfrak{B}(X)$ the σ-algebra of all *Borel sets* in X, i.e., the σ-algebra generated by the open subsets of X.

We shall also denote by $\mathfrak{B}_0(X)$ the σ-algebra of *Baire sets* of X, which is defined as the smallest σ-algebra in X with respect to which all continuous functions on X are measurable.

In general, $\mathfrak{B}_0(X) \subsetneqq \mathfrak{B}(X)$. If X is metrizable, then the two σ-algebras coincide.

A measure υ on $\mathfrak{B}(X)$ (on $\mathfrak{B}_0(X)$, respectively) is called *regular* if for every $B \in \mathfrak{B}(X)$

$$\upsilon(B) = \inf\{\upsilon(U) | B \subset U, U \text{ open}\} = \sup\{\upsilon(K) | K \subset B, K \text{ compact}\}, \quad (1.2.14)$$

(respectively, for every $B \in \mathfrak{B}_0(X)$

$$\begin{aligned}\upsilon(B) &= \inf\{\upsilon(U) | B \subset U, U \text{ open}, U \in \mathfrak{B}_0(X)\}\\ &= \sup\{\upsilon(K) | K \subset B, K \text{ compact}, K \in \mathfrak{B}_0(X)\}).\end{aligned} \quad (1.2.15)$$

Furthermore we say that a measure υ on $\mathfrak{B}(X)$ (on $\mathfrak{B}_0(X)$, respectively) is a *Borel measure* on X (a *Baire measure* on X, respectively) if

$$\upsilon(K) < +\infty \quad \text{for every compact subset } K \text{ of } X \quad (1.2.16)$$

(respectively,

$$\upsilon(K) < +\infty \quad \text{for every compact subset } K \in \mathfrak{B}_0(X)). \quad (1.2.17)$$

If X is a *Polish space*, i.e., it has a countable base and its topology is defined by a metric with respect to which it is complete, then every finite Borel measure on X is regular (note that locally compact Hausdorff spaces with a countable base are Polish spaces).

If X is a locally compact Hausdorff space which is countable at infinity, then every Baire measure on X is regular and σ-finite.

We now state the Riesz's representation theorem which establishes a one-to-one correspondence between Radon measures and Baire (Borel) measures.

For a proof see Bourbaki [1969] and Bauer [1981, Theorem 7.5.4].

1.2.4 Theorem (*Riesz's representation theorem*). *Let X be a locally compact Hausdorff space. Given $\mu \in \mathcal{M}^+(X)$, then*

(1) *If μ is bounded, there exists a unique regular finite Borel measure υ on X such that $\mathcal{K}(X) \subset \mathcal{L}^1(X, \mathfrak{B}(X), \upsilon)$ and $\mu(f) = \int f \, d\upsilon$ for every $f \in \mathcal{K}(X)$.*

(2) *If X is countable at infinity, there exists a unique (regular and σ-finite) Baire measure υ on X such that $\mathscr{K}(X) \subset \mathscr{L}^1(X, \mathfrak{B}_0(X), \upsilon)$ and $\mu(f) = \int f\,d\upsilon$ for every $f \in \mathscr{K}(X)$.*

If X is countable at infinity and metrizable, then an explicit construction of the measure υ is indicated in the next subsection.

From now on we shall assume that X is a locally compact Hausdorff space which is countable at infinity.

On the basis of the above theorem, by a common abuse of notation, we shall continue to denote by $\mathscr{M}^+(X)$ the cone of all Baire measures on X. Furthermore, the subset of all $\mu \in \mathscr{M}^+(X)$ satisfying $\mu(X) < +\infty$ ($\mu(X) = 1$, respectively) will be denoted by $\mathscr{M}_b^+(X)$ ($\mathscr{M}_1^+(X)$, respectively).

A net $(\mu_i)_{i \in I}^{\leq}$ in $\mathscr{M}_b^+(X)$ is said to be *weakly convergent* to a measure $\mu \in \mathscr{M}_b^+(X)$ if for every $f \in \mathscr{C}_b(X)$

$$\lim_{\substack{\leq \\ i \in I}} \int f\,d\mu_i = \int f\,d\mu. \tag{1.2.18}$$

If X is a Polish space and if $(\mu_i)_{i \in I}^{\leq}$ converges weakly to μ, then we obtain

$$\lim_{\substack{\leq \\ i \in I}} \int f\,d\mu_i = \int f\,d\mu \tag{1.2.19}$$

for every Borel-measurable, bounded, μ-a.e. continuous function $f: X \to \mathbb{R}$.

Finally note that, in general, the weak convergence in $\mathscr{M}_b^+(X)$ can be derived from a topology on $\mathscr{M}_b^+(X)$ that is called the *weak topology* on $\mathscr{M}_b^+(X)$.

As a simple criterion to decide if a sequence of measures with densities converges weakly, we mention the following result which is often referred to as *Scheffé's theorem*:

Let $\mu \in \mathscr{M}^+(X)$ and consider a sequence $(g_n)_{n \in \mathbb{N}}$ of positive μ-integrable numerical functions on X converging pointwise to a positive μ-integrable function g on X. If $\int g_n\,d\mu \to \int g\,d\mu$, then the sequence of measures with densities $(g_n \cdot \mu)_{n \in \mathbb{N}}$ converges weakly to $g \cdot \mu$.

In fact, for every $f \in \mathscr{C}_b(X)$

$$\left|\int f\,d(g_n \cdot \mu) - \int f\,d(g \cdot \mu)\right| = \left|\int (fg_n - fg)\,d\mu\right| \leq \|f\| \int |g_n - g|\,d\mu$$

$$= \|f\| \int (g_n + g - 2(g_n \wedge g))\,d\mu \to 0,$$

since $\int g_n \wedge g\,d\mu \to \int g\,d\mu$ by Lebesgue's dominated convergence theorem.

As a matter of fact the same result holds even if $g_n \to g$ *μ-stochastically*, i.e., for every $\varepsilon > 0$ and for every Baire subset A of finite measure $\mu(A)$ one has

$$\lim_{n\to\infty} \mu(\{|g_n - g| \geq \varepsilon\} \cap A) = 0$$

(see Bauer [1992, Lemma 21.6]).

*Integration with respect to Radon measures. $L^p(X, \mu)$-spaces

As we have seen in the above subsection, there is a one-to-one correspondence between positive Radon measures and regular Baire (Borel) measures.

Without using this correspondence, it is nevertheless possible to develop an integration theory with respect to Radon measures that leads, in particular, to the construction of the same $L^p(X, \mu)$-spaces that one obtains via the classical procedure by starting from the corresponding regular Baire (Borel) measures.

Below we shall briefly indicate the most salient points of this integration theory. For complete details see Bourbaki [1969, Chapter IV] or Choquet [1969, Section 11].

However the reader who is not interested in Korovkin-type theorems in $L^p(X, \mu)$-spaces can directly proceed to the other sections.

In the sequel we shall fix a locally compact Hausdorff space X and a positive Radon measure $\mu \in \mathscr{M}^+(X)$.

We set

$$\mathscr{K}^+(X) := \{f \in \mathscr{K}(X) | f \geq 0\} \tag{1.2.20}$$

and

$$\mathscr{I}^+(X) := \{f : X \to \tilde{\mathbb{R}} | f \text{ is lower semi-continuous and positive}\}. \tag{1.2.21}$$

For every $f \in \mathscr{I}^+(X)$ we put

$$\int^* f\, d\mu := \sup\{\mu(g) | g \in \mathscr{K}^+(X), g \leq f\} \in \mathbb{R}_+ \cup \{+\infty\}. \tag{1.2.22}$$

The number $\int^* f\, d\mu$ is called the *upper integral of f with respect to μ*. Of course,

$$\int^* g\, d\mu = \mu(g) \quad \text{if } g \in \mathscr{K}^+(X). \tag{1.2.23}$$

To assign an upper integral to every positive function we proceed as follows. If $f\colon X \to \mathbb{R}_+ \cup \{+\infty\}$ is an arbitrary function we set

$$\int^* f\,d\mu := \inf\left\{\int^* g\,d\mu \,|\, g \in \mathscr{I}^+(X), f \le g\right\} \in \mathbb{R}_+ \cup \{+\infty\}. \qquad (1.2.24)$$

Again, $\int^* f\,d\mu$ is called the *upper integral of f with respect to μ* and coincides with the one defined by (1.2.22) provided $f \in \mathscr{I}^+(X)$.

Now, fix $p \in \mathbb{R}$, $1 \le p$, and set for every $f\colon X \to \mathbb{R}$

$$N_p(f) := \left(\int^* |f|^p d\mu\right)^{1/p}. \qquad (1.2.25)$$

Furthermore let

$$\mathscr{F}^p(X,\mu) := \{f\colon X \to \mathbb{R} \,|\, N_p(f) < +\infty\}. \qquad (1.2.26)$$

Then, N_p is a seminorm on $\mathscr{F}^p(X,\mu)$ and $\mathscr{K}(X) \subset \mathscr{F}^p(X,\mu)$, because of (1.2.23) and the equality $\mathscr{K}(X) = \mathscr{K}^+(X) - \mathscr{K}^+(X)$.

We denote by $\mathscr{L}^p(X,\mu)$ the closure of $\mathscr{K}(X)$ in $\mathscr{F}^p(X,\mu)$ with respect to the seminorm N_p. Finally, we set

$$L^p(X,\mu) := \mathscr{L}^p(X,\mu)/\mathscr{N}_p, \qquad (1.2.27)$$

where $\mathscr{N}_p$ is the equivalence relation on $\mathscr{L}^p(X,\mu)$ defined by

$$f\mathscr{N}_p g \quad \Leftrightarrow \quad N_p(f-g) = 0 \quad (f, g \in \mathscr{L}^p(X,\mu).) \qquad (1.2.28)$$

The space $L^p(X,\mu)$ endowed with the norm $\|\cdot\|_p$ inherited from N_p is a Banach space.

By embedding $\mathscr{L}^p(X,\mu)$ in $L^p(X,\mu)$ we have that $\mathscr{K}(X)$ is dense in $L^p(X,\mu)$.

The functions in $\mathscr{L}^1(X,\mu)$ (or, in $L^1(X,\mu)$) are also called *μ-integrable functions.*

The value attained by the unique extension of μ to $\mathscr{L}^1(X,\mu)$ in $f \in \mathscr{L}^1(X,\mu)$ is called the *integral of f with respect to μ* and is denoted by one of the following symbols

$$\int f\,d\mu, \quad \int f(x)\,d\mu(x), \quad \int f\mu. \qquad (1.2.29)$$

If f, $g \in \mathscr{L}^1(X,\mu)$ and $f\mathscr{N}_1 g$, then $\int f\,d\mu = \int g\,d\mu$. So, the integral can be defined for every $f \in L^1(X,\mu)$.

For some classical criteria of integrability as well as for a complete treatment of the properties of the upper integral, the integral and $L^p(X, \mu)$-spaces see Bourbaki [1969, Chapter IV].

For example, if $f: X \to \mathbb{R}_+ \cup \{+\infty\}$ is lower semi-continuous and positive, then f is μ-integrable if and only if $\int^* f\,d\mu < +\infty$. If this is the case, then

$$\int f\,d\mu = \int^* f\,d\mu. \tag{1.2.30}$$

Because of (1.2.23) and (1.2.30), for every $f \in \mathscr{K}(X)$ we shall often denote the value $\mu(f)$ by $\int f\,d\mu$.

Given a subset A of X we shall set

$$\mu^*(A) := \int^* \mathbf{1}_A\,d\mu \in \mathbb{R}_+ \cup \{+\infty\}, \tag{1.2.31}$$

where $\mathbf{1}_A$ denotes the characteristic function of A.

The value $\mu^*(A)$ is called the *outer measure* of A with respect to μ.

A subset A of X such that $\mu^*(A) = 0$ is called *negligible* (or *of measure zero*).

A property P of points of X is said to hold *μ-almost everywhere* (shortly, μ-a.e.) if the subsets of all points $x \in X$ for which $P(x)$ is false, is contained in a set of measure zero.

The *inner measure* of a subset A of X is defined as

$$\mu_*(A) := \sup\{\mu^*(K) \mid K \subset A,\ K \text{ compact}\}. \tag{1.2.32}$$

A subset A of X is called *μ-integrable* if $\mathbf{1}_A$ is μ-integrable. In this case, we have

$$\mu_*(A) = \mu^*(A) = \int^* \mathbf{1}_A\,d\mu. \tag{1.2.33}$$

The space X is μ-integrable, i.e., the constant functions are μ-integrable, if and only if μ is bounded. Then,

$$\|\mu\| = \mu^*(X) = \int \mathbf{1}\,d\mu. \tag{1.2.34}$$

Furthermore $\mathscr{C}_0(X)$ is a dense subspace of $\mathscr{L}^p(X, \mu)$ $(1 \le p < +\infty)$. Moreover for every $f \in \mathscr{C}_0(X)$

$$\left|\int f\,d\mu\right| \le \|f\|\,\|\mu\|. \tag{1.2.35}$$

A function $f: X \to \mathbb{R}_+ \cup \{+\infty\}$ is called *μ-measurable* if for every $\varepsilon > 0$ and for every compact subset Y of X there exists a compact subset K of X such that f is continuous on K and $\mu^*(Y \setminus K) < \varepsilon$.

The subspace of all μ-measurable functions $f: X \to \mathbb{R}$ such that

$$N_\infty(f) := \inf\{\alpha \in \mathbb{R}_+ \,|\, |f| \le \alpha \ \mu\text{-a.e.}\} < +\infty \tag{1.2.36}$$

is denoted by $\mathscr{L}^\infty(X, \mu)$.

The quotient space $\mathscr{L}^\infty(x, \mu)/\mathscr{N}_\infty$, where $\mathscr{N}_\infty$ is the equivalence relation on $\mathscr{L}^\infty(X, \mu)$ defined by

$$f \mathscr{N}_\infty g \quad \Leftrightarrow \quad N_\infty(f - g) = 0 \quad \Leftrightarrow \quad f = g \ \mu\text{-a.e.}, \tag{1.2.37}$$

will be denoted by $L^\infty(X, \mu)$.

The space $L^\infty(X, \mu)$ endowed with the norm

$$\|f\|_\infty := N_\infty(f) \tag{1.2.38}$$

(where every $f \in L^\infty(X, \mu)$ is identified with an arbitrary representing function), is a Banach space.

Moreover, all $L^p(X, \mu)$-spaces $(1 \le p \le +\infty)$ endowed with the ordering

$$f \le g \quad \text{if} \quad f \le g \quad \mu\text{-a.e.} \tag{1.2.39}$$

are Banach lattices.

We say that a subset A of X is *μ-measurable* if $\mathbf{1}_A$ is μ-measurable. The set of all μ-measurable subsets of X will be denoted by $\mathfrak{B}^*(X)$. In fact, $\mathfrak{B}^*(X)$ is a σ-algebra and μ^* is a measure on $\mathfrak{B}^*(X)$. Moreover, $\mathfrak{B}^*(X)$ contains the σ-algebra $\mathfrak{B}(X)$ of Borel sets in X.

If X is countable at infinity and metrizable, then the unique Baire (or, equivalently, Borel) measure υ on X determined from Riesz's representation theorem is, in fact, μ^*. Furthermore, $\mathscr{L}^1(X, \mathfrak{B}(X), \mu^*) = \mathscr{L}^1(X, \mu)$ and for every $f \in \mathscr{L}^1(X, \mu)$

$$\int f \, d\mu = \int f \, d\mu^* \tag{1.2.40}$$

(see Choquet [1969, Theorem 11.18]).

Restrictions and extensions of Radon measures

Again we fix a locally compact Hausdorff space X and a measure $\mu \in \mathscr{M}^+(X)$. Given a locally compact subset Y of X, there is a classical procedure to define a new Radon measure on Y, which is called the *restriction of* μ *to* Y and is denoted by $\mu_{|Y}$.

This restriction is defined by

$$\mu_{|Y}(f) := \int f^* \, d\mu \quad \text{for every } f \in \mathscr{K}(Y), \tag{1.2.41}$$

where

$$f^*(x) := \begin{cases} f(x), & \text{if } x \in Y, \\ 0, & \text{if } x \notin Y. \end{cases} \tag{1.2.42}$$

(Note that $(f^+)^*$ and $(f^-)^*$ are upper semi-continuous and, hence, μ-measurable; so, $f^* = (f^+)^* - (f^-)^*$ is μ-measurable, bounded and with compact support, so that it is μ-integrable).

When X is metrizable or countable at infinity, we may employ another procedure (which does not make use of integration theory) to define the restriction of μ to Y in the particular case when Y is compact and $\mathrm{Supp}(\mu) \subset Y$.

In fact, since X is normal, given $f \in \mathscr{K}(Y) = \mathscr{C}(Y)$, by Tietze's theorem there exists a continuous extension $f_1 \colon X \to \mathbb{R}$ of f. After choosing $f_2 \in \mathscr{K}(X)$ such that $f_2 = 1$ on Y, clearly the function $g := f_1 f_2$ is another extension of f which belongs to $\mathscr{K}(X)$.

Note that, if g_1 and g_2 are two extensions of f belonging to $\mathscr{K}(X)$, then $\mu(g_1) = \mu(g_2)$ since $\mathrm{Supp}(\mu) \subset Y$. So, we may define a Radon measure υ on Y by

$$\upsilon(f) := \mu(g) \quad \text{for every } f \in \mathscr{C}(Y), \tag{1.2.43}$$

where $g \in \mathscr{K}(X)$ is an arbitrary extension of f to X.

As it is easy to see, the measure υ coincides, in fact, with the measure $\mu_{|Y}$ defined by (1.2.41).

We shall use this approach mainly in the particular case when X is compact.

Finally let us make some remarks about a simple procedure to extend Radon measures.

Let Y be a closed subset of X and fix $\mu \in \mathscr{M}_b^+(Y)$. Then we may define a new bounded Radon measure $\bar{\mu}$ on X by

$$\bar{\mu}(f) := \mu(f_{|Y}) \quad \text{for every } f \in \mathscr{K}(X). \tag{1.2.44}$$

The measure $\bar{\mu}$ is called the *canonical extension of* μ *over* X and, by a common abuse of language, it is still denoted by μ.

Image Radon measures

Let X and Y be locally compact Hausdorff spaces. A continuous mapping $\varphi: X \to Y$ is called *proper* if for every compact subset K of Y, $\varphi^{-1}(K)$ is compact in X.

If X is compact, then every continuous mapping from X into Y is proper.

Furthermore, if $f: Y \to \mathbb{R}$, then $\mathrm{Supp}(f \circ \varphi) \subset \varphi^{-1}(\mathrm{Supp}(f))$ so that, if f is proper, we have

$$f \circ \varphi \in \mathscr{K}(X) \quad \text{for every } f \in \mathscr{K}(Y), \tag{1.2.45}$$

as well as

$$f \circ \varphi \in \mathscr{C}_0(X) \quad \text{for every } f \in \mathscr{C}_0(Y). \tag{1.2.46}$$

Given a Radon measure $\mu \in \mathscr{M}(X)$ and a proper mapping $\varphi: X \to Y$, we may consider the Radon measure υ on Y defined by

$$\upsilon(f) := \mu(f \circ \varphi) \quad \text{for every } f \in \mathscr{K}(Y). \tag{1.2.47}$$

The measure υ is called the *image of μ under the mapping φ* and it is denoted by $\varphi(\mu)$.

If μ is positive, $\varphi(\mu)$ is positive. If μ is bounded, then $\varphi(\mu)$ is bounded as well and $\|\varphi(\mu)\| \le \|\mu\|$. Moreover, if, in addition, μ is positive, then $\|\varphi(\mu)\| = \|\mu\|$.

In general, $\mathrm{Supp}(\varphi(\mu)) \subset \varphi(\mathrm{Supp}(\mu))$ and we have equality if μ is positive.

Furthermore the mapping $\mu \mapsto \varphi(\mu)$ from $\mathscr{M}(X)$ into $\mathscr{M}(Y)$ is vaguely continuous.

Note that, if $\varphi: X \to Y$ is an arbitrary continuous mapping (or, more generally, measurable with respect to the σ-algebras $\mathfrak{B}(X)$ and $\mathfrak{B}(Y)$) and $\mu \in \mathscr{M}_b^+(X)$, we could define a Radon measure $\varphi^*(\mu)$ on $\mathscr{M}(Y)$ by

$$\varphi^*(\mu)(f) := \int f \circ \varphi \, d\mu \quad \text{for every } f \in \mathscr{K}(Y). \tag{1.2.48}$$

But, in this case, the mapping $\mu \mapsto \varphi^*(\mu)$ fails to be continuous.

For more details on proper mappings and image measures see Bourbaki [1965, § 10, n.1], [1969] and Choquet [1969, Section 13].

Tensor products of Radon measures and of positive operators

Let $(X_i)_{1 \le i \le p}$ be a finite family of locally compact Hausdorff spaces and consider the product space $\prod_{i=1}^{p} X_i$ endowed with the product topology.

The product space is a locally compact Hausdorff space and is compact if each X_i is compact.

For each $j = 1, \ldots, p$ we shall denote by $\mathrm{pr}_j\colon \prod_{i=1}^{p} X_i \to X_j$ the *j-th projection* which is defined by

$$pr_j(x) := x_j \quad \text{for every } x = (x_i)_{1 \le i \le p} \in \prod_{i=1}^{p} X_i. \tag{1.2.49}$$

By a common abuse of notation, if $X \subset \prod_{i=1}^{p} X_i$, the restriction of each pr_j to X will be denoted by pr_j as well.

If finitely many functions $f_i\colon X_i \to \mathbb{R}$, $1 \le i \le p$, are given, we shall denote by $\bigotimes_{i=1}^{p} f_i\colon \prod_{i=1}^{p} X_i \to \mathbb{R}$ the new function defined by

$$\left(\bigotimes_{i=1}^{p} f_i\right)(x) := \prod_{i=1}^{p} f_i(x_i) \quad \text{for every } x = (x_i)_{1 \le i \le p} \in \prod_{i=1}^{p} X_i. \tag{1.2.50}$$

Thus we have

$$\bigotimes_{i=1}^{p} f_i = \prod_{i=1}^{p} f_i \circ pr_i. \tag{1.2.51}$$

Furthermore, if $j = 1, \ldots, p$ and $f_j\colon X_j \to \mathbb{R}$, then

$$f_j \circ pr_j = \bigotimes_{i=1}^{p} f_{i,j}, \tag{1.2.52}$$

where $f_{i,j} := \mathbf{1}$ if $i \ne j$, and $f_{i,j} := f_j$ if $i = j$.

Clearly, if for each $i = 1, \ldots, p$, $f_i \in \mathscr{K}(X_i)$, then $\bigotimes_{i=1}^{p} f_i \in \mathscr{K}\left(\prod_{i=1}^{p} X_i\right)$, because $\mathrm{Supp}\left(\bigotimes_{i=1}^{p} f_i\right) = \prod_{i=1}^{p} \mathrm{Supp}(f_i)$.

We shall denote by $\bigotimes_{i=1}^{p} \mathscr{K}(X_i)$ the linear subspace generated by $\left\{\bigotimes_{i=1}^{p} f_i \middle| f_i \in \mathscr{K}(X_i), i = 1, \ldots, p\right\}$. In fact, we have that $\bigotimes_{i=1}^{p} \mathscr{K}(X_i)$ is dense in $\mathscr{K}\left(\prod_{i=1}^{p} X_i\right)$ with respect to the sup-norm (see Choquet [1969, Lemma 13.8]).

Now, for every $i = 1, \ldots, p$ fix $\mu_i \in \mathscr{M}(X_i)$. Then there exists a uniquely determined Radon measure υ on $\prod_{i=1}^{p} X_i$ such that for every $(f_i)_{1 \le i \le p} \in \prod_{i=1}^{p} \mathscr{K}(X_i)$

$$\upsilon\left(\bigotimes_{i=1}^{p} f_i\right) = \prod_{i=1}^{p} \mu_i(f_i). \tag{1.2.53}$$

Such a measure is called the *tensor product* of the family $(\mu_i)_{1 \le i \le p}$ and is denoted by $\bigotimes_{i=1}^{p} \mu_i$ or $\mu_1 \otimes \cdots \otimes \mu_p$. Thus, if $f_i \in \mathscr{K}(X_i)$, $1 \le i \le p$, then

$$\left(\bigotimes_{i=1}^{p} \mu_i\right)\left(\bigotimes_{i=1}^{p} f_i\right) = \prod_{i=1}^{p} \mu_i(f_i). \tag{1.2.54}$$

Note that, if $f \in \mathscr{K}\left(\prod_{i=1}^{p} X_i\right)$, then for every $j = 1, \ldots, p-1$ and $(x_1, \ldots, x_j) \in \prod_{i=1}^{j} X_i$ the function

$$x_j \mapsto \int \ldots \left(\int f(x_1, \ldots, x_j, x_{j+1}, \ldots, x_p) \, d\mu_p(x_p) \right) \ldots d\mu_{j+1}(x_{j+1})$$

from X_j into $\mathbb{R}$ is continuous and has compact support. As a matter of fact it is possible to show that

$$\left(\bigotimes_{i=1}^{p} \mu_i\right)(f) = \int \left(\ldots \left(\int f(x_1, \ldots, x_p) \, d\mu_p(x_p) \right) \ldots \right) d\mu_1(x_1). \tag{1.2.55}$$

We shall also denote the right-hand side of (1.2.55) by

$$\int \ldots \int f(x_1, \ldots, x_p) \, d\mu_1(x_1) \ldots d\mu_p(x_p). \tag{1.2.56}$$

In fact, from the above formula (1.2.55) one can also deduce the *Fubini's theorem* for functions $f\colon \prod_{i=1}^{p} X_i \to \mathbb{R}$ which are $\left(\bigotimes_{i=1}^{p} \mu_i\right)$-integrable, namely

$$\int f \, d\left(\bigotimes_{i=1}^{p} \mu_i\right) = \int \ldots \int f(x_1, \ldots, x_p) \, d\mu_1(x_1) \ldots d\mu_p(x_p). \tag{1.2.57}$$

Note also that, if every μ_i is positive, then $\bigotimes_{i=1}^{p} \mu_i$ is positive. Furthermore, if each X_i is compact and $\mu_i(1) = 1$, then, because of (1.2.51) and (1.2.54), for each $j = 1, \ldots, p$ we obtain

$$pr_j\left(\bigotimes_{i=1}^{p} \mu_i\right) = \mu_j. \tag{1.2.58}$$

Some of the main properties of tensor products of measures are listed below (for a proof see Bourbaki [1969, Chapter III, Section 4] or Choquet [1969, Section 13]).

1.2.5 Proposition. *Let $(X_i)_{1\le i\le p}$ be a finite family of locally compact Hausdorff spaces and for every $i = 1,\dots,p$ fix $\mu_i \in \mathcal{M}(X_i)$.*

Then the following statement hold:

(1) $\operatorname{Supp}\left(\bigotimes_{i=1}^{p} \mu_i\right) = \prod_{i=1}^{p} \operatorname{Supp}(\mu_i)$.

(2) *If each μ_i is bounded then $\bigotimes_{i=1}^{p} \mu_i$ is bounded and* $\left\|\bigotimes_{i=1}^{p} \mu_i\right\| = \prod_{i=1}^{p} \|\mu_i\|$.

(3) *(Commutativity property) If $\sigma\colon \{1,\dots,p\} \to \{1,\dots,p\}$ is a permutation, then* $\bigotimes_{i=1}^{p} \mu_{\sigma(i)} = \bigotimes_{i=1}^{p} \mu_i$.

(4) *(Associativity property) If $(I_k)_{1\le k\le n}$ is a partition of $\{1,\dots,p\}$, then* $\bigotimes_{k=1}^{n}\left(\bigotimes_{i\in I_k} \mu_i\right) = \bigotimes_{i=1}^{p} \mu_i$.

Note that, in general, the mapping $(\mu_i)_{1\le i\le p} \mapsto \bigotimes_{i=1}^{p} \mu_i$ from $\prod_{i=1}^{p} \mathcal{M}(X_i)$ into $\mathcal{M}\left(\prod_{i=1}^{p} X_i\right)$ is *not continuous* with respect to the product topology (of the vague topologies) and the vague topology on $\mathcal{M}\left(\prod_{i=1}^{p} X_i\right)$.

However, we have the following useful result (see Choquet [1969, Proposition 13.12]).

1.2.6 Proposition. *Given a finite family $(X_i)_{1\le i\le p}$ of locally compact Hausdorff spaces, then the mapping $(\mu_i)_{1\le i\le p} \mapsto \bigotimes_{i=1}^{p} \mu_i$ from $\prod_{i=1}^{p} \mathcal{M}^+(X_i)$ into $\mathcal{M}^+\left(\prod_{i=1}^{p} X_i\right)$ is continuous with respect to the product topology (of the vague topologies) and the vague topology on $\mathcal{M}^+\left(\prod_{i=1}^{p} X_i\right)$.*

By using tensor products of measures we can also construct positive linear operators on spaces of continuous functions on the product space.

First, note that to every positive linear operator defined on a suitable function space, it is possible to associate a family of positive Radon measures.

More precisely, given two locally compact Hausdorff spaces X and Y, consider two function spaces E and F on X and Y, respectively (i.e., E and F are vector subspaces of continuous functions on X and Y, respectively). Furthermore suppose that $\mathcal{K}(X) \subset E$.

Given a positive linear operator $T: E \to F$, for every $y \in Y$ we consider the linear form $\mu_y^T: \mathscr{K}(X) \to \mathbb{R}$ defined by

$$\mu_y^T(f) := Tf(y) \quad \text{for every } f \in \mathscr{K}(X). \tag{1.2.59}$$

Then μ_y^T is positive and, hence, it is a Radon measure on X.

More generally, if $\mu: F \to \mathbb{R}$ is a positive linear form, we shall denote by $T(\mu)$ the positive Radon measure on X defined by

$$T(\mu)(f) := \mu(T(f)) \quad \text{for every } f \in \mathscr{K}(X). \tag{1.2.60}$$

Thus $\mu_y^T = T(\varepsilon_{y|F})$.

Now, let $(X_i)_{1 \le i \le p}$ and $(Y_i)_{1 \le i \le p}$ be two finite families of locally compact Hausdorff spaces. For every $i = 1, \ldots, p$ let us consider a positive linear operator $T_i: \mathscr{K}(X_i) \to \mathscr{C}(Y_i)$.

Then we define a linear operator $T: \mathscr{K}\left(\prod_{i=1}^p X_i\right) \to \mathscr{C}\left(\prod_{i=1}^p Y_i\right)$ by

$$Tf(y) := \left(\bigotimes_{i=1}^p \mu_{y_i}^{T_i}\right)(f) = \int \cdots \int f(x_1, \ldots, x_p)\, d\mu_{y_1}^{T_1}(x_1) \ldots d\mu_{y_p}^{T_p}(x_p) \tag{1.2.61}$$

for every $f \in \mathscr{K}\left(\prod_{i=1}^p X_i\right)$ and $y = (y_i)_{1 \le i \le p} \in \prod_{i=1}^p Y_i$, where the measures $\mu_{y_i}^{T_i}$ are defined as in (1.2.59). (Note that $T(f)$ is continuous by virtue of Proposition 1.2.6).

The operator T is positive and is denoted by $\bigotimes_{i=1}^p T_i$. It is also called the *tensor product of the family* $(T_i)_{1 \le i \le p}$.

Thus $\bigotimes_{i=1}^p T_i: \mathscr{K}\left(\prod_{i=1}^p X_i\right) \to \mathscr{C}\left(\prod_{i=1}^p Y_i\right)$ and for every $f \in \mathscr{K}\left(\prod_{i=1}^p X_i\right)$ and $y = (y_i)_{1 \le i \le p} \in \prod_{i=1}^p Y_i$,

$$\left(\bigotimes_{i=1}^p T_i\right)(f)(y) = \left(\bigotimes_{i=1}^p \mu_{y_i}^{T_i}\right)(f) = \int \cdots \int f(x_1, \ldots, x_p)\, d\mu_{y_1}^{T_1}(x_1) \ldots d\mu_{y_p}^{T_p}(x_p). \tag{1.2.62}$$

In particular, taking (1.2.59) and (1.2.54) into account, for every $(f_i)_{1 \le i \le p} \in \prod_{i=1}^p \mathscr{K}(X_i)$ we have

$$\left(\bigotimes_{i=1}^p T_i\right)\left(\bigotimes_{i=1}^p f_i\right) = \bigotimes_{i=1}^p T_i(f_i). \tag{1.2.63}$$

*1.3 Some basic principles of probability theory

Here we survey some classical material on probability theory that will be used mainly in Section 5.2. For more details and proofs see for instance Bauer [1981, part II] and Feller [1957], [1966].

Random variables

Consider a *probability space* $(\Omega, \mathscr{F}, P)$, i.e., $\mathscr{F}$ is a σ-algebra in the set Ω and P is a measure on $\mathscr{F}$ such that $P(\Omega) = 1$, and let $(\Omega', \mathscr{F}')$ be a measurable space. A *random variable* from Ω into Ω' is a mapping $Z: \Omega \to \Omega'$ which is *measurable* with respect to $\mathscr{F}$ and $\mathscr{F}'$ (i.e., $Z^{-1}(B) \in \mathscr{F}$ for every $B \in \mathscr{F}'$).

When $\Omega' = \mathbb{R}$ and $\mathscr{F}' = \mathfrak{B}(\mathbb{R})$ we shall speak of real random variables on Ω. The set of all real random variables will be denoted by $M(\Omega)$.

If $Z: \Omega \to \Omega'$ is a random variable from Ω into Ω', the image measure $Z(P)$ is called the *distribution* of Z (with respect to P) or the *probability law* of Z and is denoted by P_Z.

Thus P_Z is a probability measure on $\mathscr{F}'$ and for every $B \in \mathscr{F}'$

$$P_Z(B) := P\{Z^{-1}(B)\}. \tag{1.3.1}$$

The subset $Z^{-1}(B)$ and the number $P\{Z^{-1}(B)\}$ are often denoted by $\{Z \in B\}$ and $P\{Z \in B\}$, respectively.

If μ is a probability measure on $\mathscr{F}'$ and $P_Z = \mu$, we also say that Z is *distributed according* to μ.

If $Z \in M(\Omega)$, then $P_Z \in \mathscr{M}_1^+(\mathbb{R})$. More generally, if Ω' is a locally compact Hausdorff space which is countable at infinity and $\mathscr{F}' = \mathfrak{B}_0(\Omega')$, then $P_Z \in \mathscr{M}_1^+(\Omega')$ for every random variable $Z: \Omega \to \Omega'$.

In this case, if Ω' has a countable base, then $\operatorname{Supp}(P_Z) \subset \overline{X(\Omega)}$ (we recall that if $\mu \in \mathscr{M}^+(\Omega')$, the *support* $\operatorname{Supp}(\mu)$ of μ is the complement of the largest open subset of Ω' of measure zero with respect to μ).

If $Z: \Omega \to \Omega'$ is a random variable and if $f: \Omega' \to \tilde{\mathbb{R}}$ is positive and $\mathscr{F}'$-measurable, then

$$\int_{\Omega'} f \, dP_Z = \int_{\Omega} f \circ Z \, dP. \tag{1.3.2}$$

Furthermore a $\mathscr{F}'$-measurable function $f: \Omega' \to \tilde{\mathbb{R}}$ is P_Z-integrable if and only if $f \circ Z$ is P-integrable. In this case (1.3.2) holds as well.

If a real random variable $Z\colon \Omega \to \mathbb{R}$ is positive or P-integrable, then we set

$$E(Z) := \int_{\Omega} Z\,dP = \int_{-\infty}^{+\infty} x\,dP_Z(x) \tag{1.3.3}$$

and we call $E(Z)$ the *expected value* of Z.

If the random variable $Z\colon \Omega \to \mathbb{R}$ is P-integrable, we call

$$\operatorname{Var}(Z) := E((Z - E(Z))^2) \in \tilde{\mathbb{R}}_+, \tag{1.3.4}$$

and

$$\sigma(Z) := \sqrt{\operatorname{Var}(Z)}, \tag{1.3.5}$$

the *variance* and the *standard deviation* (or *dispersion*) of Z, respectively.

A real random variable Z is P-integrable and has finite variance if and only if it is square-integrable. If this is the case, then

$$\operatorname{Var}(Z) = E(Z^2) - E(Z)^2 = \int_{-\infty}^{+\infty} x^2\,dP_Z(x) - \left(\int_{-\infty}^{+\infty} x\,dP_Z(x)\right)^2. \tag{1.3.6}$$

If $Z\colon \Omega \to \mathbb{R}^p$ is a random variable with components $Z_1, \ldots, Z_p$ (i.e., $Z(\omega) = (Z_1(\omega), \ldots, Z_p(\omega))$ for every $\omega \in \Omega$), we set

$$E(Z) := (E(Z_1), \ldots, E(Z_p)) \in \mathbb{R}^p, \tag{1.3.7}$$

provided every Z_i is integrable. Furthermore, if every Z_i is square integrable, we also set

$$\operatorname{Var}(Z) := \sum_{i=1}^{p} \operatorname{Var}(Z_i) = E(\|Z\|^2) - \|E(Z)\|^2. \tag{1.3.8}$$

Let $Z\colon \Omega \to \mathbb{R}$ be a real random variable such that $P\{Z \in \mathbb{N}_0\} = 1$. For every $n \in \mathbb{N}_0$ we set $\alpha_n := P\{Z = n\}$ $\left(\text{hence } \sum_{n=0}^{\infty} \alpha_n = 1\right)$.

The *probability generating function* of Z is the function $g_Z\colon [-1, 1] \to \mathbb{R}$ defined by

$$g_Z(t) := \sum_{n=0}^{\infty} \alpha_n t^n = E(t^Z) = \int_{-\infty}^{+\infty} t^x\,dP_Z(x) \quad \text{for every } t \in [-1, 1]. \tag{1.3.9}$$

In this case we have

$$E(Z) = \sum_{n=1}^{\infty} n\alpha_n = g_Z'(1) \tag{1.3.10}$$

and

$$\operatorname{Var}(Z) = g_Z''(1) + g_Z'(1) - g_Z'(1)^2. \tag{1.3.11}$$

If $\operatorname{Var}(Z) < +\infty$, then

$$\operatorname{Var}(Z) = \sum_{n=1}^{\infty} n^2\alpha_n - \left(\sum_{n=1}^{\infty} n\alpha_n\right)^2. \tag{1.3.12}$$

More generally, if for every $p \in \mathbb{N}$, we denote by $m_p(Z)$ the *p-th factorial moment of* Z, i.e.,

$$m_p(Z) := E(Z(Z-1)\dots(Z-p+1)) = \int_{-\infty}^{+\infty} x(x-1)\dots(x-p+1)\,dP_Z(x), \tag{1.3.13}$$

then

$$m_p(Z) = g_Z^{(p)}(1). \tag{1.3.14}$$

A random variable $Z\colon \Omega \to \mathbb{R}^p$ is said to be *discretely distributed* if there exists a sequence $(a_n)_{n\in\mathbb{N}_0}$ in $\mathbb{R}^p$ and a sequence $(\alpha_n)_{n\in\mathbb{N}_0}$ in $\mathbb{R}_+$ satisfying $\sum_{n=0}^{\infty} \alpha_n = 1$, such that

$$P_Z = \sum_{n=0}^{\infty} \alpha_n \varepsilon_{a_n}, \tag{1.3.15}$$

where, for every $n \in \mathbb{N}_0$, $\varepsilon_{a_n} \in \mathscr{M}_1^+(\mathbb{R}^p)$ denotes the *unit mass* at a_n, i.e., for every $B \in \mathfrak{B}(\mathbb{R}^p)$

$$\varepsilon_{a_n}(B) := \begin{cases} 1, & \text{if } a_n \in B, \\ 0, & \text{if } a_n \notin B. \end{cases} \tag{1.3.16}$$

In particular we have that $P\{Z = a_n\} = \alpha_n$ for every $n \in \mathbb{N}_0$. Furthermore a measurable function $f\colon \mathbb{R}^p \to \bar{\mathbb{R}}$ is P_Z-integrable if and only if $\sum_{n=0}^{\infty} \alpha_n |f(a_n)| < +\infty$. In this case

$$\int_{\mathbb{R}^p} f\,dP_Z = \sum_{n=0}^{\infty} \alpha_n f(a_n). \tag{1.3.17}$$

The same formula holds if f is positive (not necessarily integrable).

In the case $p = 1$ we have that Z is integrable (square-integrable, respectively) if and only if $\sum_{n=0}^{\infty} \alpha_n |a_n| < +\infty$ $\left(\sum_{n=0}^{\infty} \alpha_n a_n^2 < +\infty, \text{ respectively}\right)$.

Furthermore

$$E(Z) = \sum_{n=0}^{\infty} \alpha_n a_n \tag{1.3.18}$$

and

$$\operatorname{Var}(Z) = \sum_{n=0}^{\infty} \alpha_n a_n^2 - \left(\sum_{n=0}^{\infty} \alpha_n a_n\right)^2. \tag{1.3.19}$$

Important examples of discretely distributed real random variables are the *binomial* or *Bernoulli* random variables. A real random variable Z is said to be a Bernoulli random variable with parameters n and p ($n \geq 1$, $0 \leq p \leq 1$) if it is distributed according to the *binomial* or *Bernoulli* distribution

$$\beta_{n,p} := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \varepsilon_k. \tag{1.3.20}$$

In this case

$$E(Z) = np \quad \text{and} \quad \operatorname{Var}(Z) = np(1-p). \tag{1.3.21}$$

Furthermore for every $t \in [-1, 1]$

$$g_Z(t) = (pt + (1-p))^n. \tag{1.3.22}$$

Another class of important discretely distributed real random variables are the *Poisson random variables*. A Poisson random variable Z with parameter $\alpha > 0$ is a real random variable which is distributed according to the *Poisson distribution* with parameter α

$$\pi_\alpha := \sum_{k=0}^{\infty} \exp(-\alpha) \frac{\alpha^k}{k!} \varepsilon_k. \tag{1.3.23}$$

We have

$$E(Z) = \operatorname{Var}(Z) = \alpha \tag{1.3.24}$$

and

$$g_Z(t) = \exp(\alpha(t-1)) \quad \text{for every } t \in [-1, 1]. \tag{1.3.25}$$

As an example of discretely distributed random variables on $\mathbb{R}^p$ with $p \geq 2$, consider finitely many positive real numbers $r_1, \ldots, r_p$ satisfying $\sum_{i=1}^{p} r_i \leq 1$ and fix $n \in \mathbb{N}$.

Every random variable Z on $\mathbb{R}^p$ distributed according to

$$\sum_{\substack{h_1, \ldots, h_p \geq 0 \\ h_1 + \cdots + h_p \leq n}} \frac{n!}{h_1! \ldots h_p!(n - h_1 - \cdots - h_p)!} \times r_1^{h_1} \ldots r_p^{h_p}(1 - r_1 - \cdots - r_p)^{n-h_1-\cdots-h_p} \varepsilon_{(h_1, \ldots, h_p)} \tag{1.3.26}$$

is called a *multinomial random variable* of order $p + 1$ with parameters $n, r_1, \ldots, r_p$.

In this case $E(Z) = (nr_1, \ldots, nr_p)$ and $\operatorname{Var}(Z) = n \sum_{i=1}^{p} r_i(1 - r_i)$.

A random variable $Z: \Omega \to \mathbb{R}^p$ is said to be *Lebesgue-continuous* if P_Z is *λ_p-continuous*, i.e., $P\{Z \in B\} = 0$ for every $B \in \mathfrak{B}(\mathbb{R}^p)$ such that $\lambda_p(B) = 0$. Here λ_p denotes the *Lebesgue-Borel measure* in $\mathbb{R}^p$.

By the Radon-Nikodym theorem (see, e.g., Bauer [1981, Theorem 2.9.10]), if Z is Lebesgue continuous, then there exists a Borel-measurable positive function $g: \mathbb{R}^p \to \tilde{\mathbb{R}}$ satisfying $\int_{\mathbb{R}^p} g(x)\,dx = 1$ such that $P_Z = g \cdot \lambda_p$, i.e., for every $B \in \mathfrak{B}(\mathbb{R}^p)$

$$P\{Z \in B\} = \int_B g(x)\,dx. \tag{1.3.27}$$

The function g is also called the *probability density* of Z.

A measurable function $f: \mathbb{R}^p \to \tilde{\mathbb{R}}$ is P_Z-integrable if and only if fg is λ_p-integrable. In this case

$$\int_{\mathbb{R}^p} f\,dP_Z = \int_{\mathbb{R}^p} f(x)g(x)\,dx. \tag{1.3.28}$$

This formula also holds provided f is positive and measurable.

In particular, when $p = 1$, we obtain that Z is integrable (square-integrable, respectively) if and only if $\int_{-\infty}^{+\infty} |x| g(x)\,dx < +\infty$ ($\int_{-\infty}^{+\infty} x^2 g(x)\,dx < +\infty$, respectively). Then

$$E(Z) = \int_{-\infty}^{+\infty} x g(x)\,dx \tag{1.3.29}$$

and

$$\operatorname{Var}(Z) = \int_{-\infty}^{+\infty} x^2 g(x)\,dx - \left(\int_{-\infty}^{+\infty} x g(x)\,dx\right)^2. \tag{1.3.30}$$

The most important Lebesgue-continuous random variables are the normal ones. We recall that a *normal* or *Gaussian random variable* Z with parameters α and σ^2 ($\alpha \in \mathbb{R}$, $\sigma > 0$) is a real Lebesgue-continuous random variable having as probability density the function

$$g_{\alpha,\sigma^2}(t) := (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(t-\alpha)^2}{2\sigma^2}\right) \quad (t \in \mathbb{R}). \tag{1.3.31}$$

In this case

$$E(Z) = \alpha \quad \text{and} \quad \operatorname{Var}(Z) = \sigma^2. \tag{1.3.32}$$

Some properties of real random variables can be also described in terms of their distribution functions. Actually, given a real random variable Z, the *distribution function* of Z is the function $F_Z\colon \mathbb{R} \to \mathbb{R}$ defined by

$$F_Z(x) := P\{Z < x\} \quad \text{for every } x \in \mathbb{R}. \tag{1.3.33}$$

The function F_Z is increasing and *left-continuous* $\left(\text{i.e., } \lim_{x \to a^-} F_Z(x) = F_Z(a) \text{ for}\right.$ $\left.\text{every } a \in \mathbb{R}\right)$ and satisfies the conditions $\lim_{x \to -\infty} F_Z(x) = 0$ and $\lim_{x \to +\infty} F_Z(x) = 1$.

Furthermore $P\{a \le Z < b\} = F_Z(b) - F_Z(a)$ provided $a < b$, and F_Z is continuous in a point $a \in \mathbb{R}$ if and only if $P\{Z = a\} = 0$.

We also recall that, if $F\colon \mathbb{R} \to \mathbb{R}$ is an increasing left-continuous function converging to 0 as $x \to -\infty$ and to 1 as $x \to +\infty$, then there exists a (in general, non-unique) random variable Z such that $F_Z = F$.

A crucial notion in probability theory is that of independence of random variables.

Consider an arbitrary family $(Z_i)_{i \in I}$ of real random variables defined on the same probability space $(\Omega, \mathscr{F}, P)$.

The family $(Z_i)_{i \in I}$ is said to be *independent* if for every finite subset $J \subset I$ and for every finite family $(B_i)_{i \in I}$ in $\mathfrak{B}(\mathbb{R})$ we have

$$P\left(\bigcap_{i \in J} \{Z_i \in B_i\}\right) = \prod_{i \in J} P\{Z_i \in B_i\}. \tag{1.3.34}$$

If this is the case, then for every finite subset $J \subset I$ the following statements hold:

(1) If we consider the random variable $\bigotimes_{i \in J} Z_i \colon \Omega \to \mathbb{R}^J$ defined as $\left(\bigotimes_{i \in J} Z_i\right)(\omega) = (Z_i(\omega))_{i \in J}$ for every $\omega \in \Omega$ and if $\bigotimes_{i \in J} P_{Z_i}$ denotes the tensorial product of the family $(P_{Z_i})_{i \in J}$, then

$$P_{\bigotimes_{i \in J} Z_i} = \bigotimes_{i \in J} P_{Z_i}. \tag{1.3.35}$$

(2) If either all Z_i are positive or all Z_i are integrable, then

$$E\left(\prod_{i \in J} Z_i\right) = \prod_{i \in J} E(Z_i). \tag{1.3.36}$$

(3) If all Z_i are integrable, then

$$\operatorname{Var}\left(\sum_{i \in J} Z_i\right) = \sum_{i \in J} \operatorname{Var}(Z_i). \tag{1.3.37}$$

(4) If $\mathop{\ast}_{i \in J} P_{Z_i}$ denotes the convolution product of the family $(P_{Z_i})_{i \in J}$, then

$$P_{\sum_{i \in J} Z_i} = \mathop{\ast}_{i \in J} P_{Z_i}. \tag{1.3.38}$$

(5) If we consider the probability generating function $g_{\sum_{i \in J} Z_i} \colon [-1, 1] \to \mathbb{R}$ of $\sum_{i \in J} Z_i$, then

$$g_{\sum_{i \in J} Z_i} = \prod_{i \in J} g_{Z_i}. \tag{1.3.39}$$

From (1.3.38) it easily follows that if Z_1 and Z_2 are two independent binomial (Poisson, normal, respectively) random variables with parameters n, p and m, p (α and β, α, σ^2 and β, τ^2, respectively), then $Z_1 + Z_2$ is a binomial (Poisson, normal, respectively) random variable with parameters $n + m$, p ($\alpha + \beta$, $\alpha + \beta$ and $\sigma^2 + \tau^2$, respectively).

Similarly, one can prove that, if $Z_1, \ldots, Z_n$ are independent normal random variables with the same parameters 0 and σ^2, then the random variable $\chi^2_{n,\sigma^2} := \sum_{i=1}^{n} Z_i^2$ is Lebesgue-continuous and its probability density is the function

$$g_{n,\sigma^2}(t) := \begin{cases} \dfrac{t^{n/2-1} \exp(-t/2\sigma^2)}{(2\sigma^2)^{n/2}\Gamma(n/2)}, & \text{if } t > 0, \\[2ex] 0, & \text{if } t \leq 0, \end{cases} \tag{1.3.40}$$

where $\Gamma(t) := \int_0^{+\infty} x^{t-1} \exp(-x)\, dx$ $(t > 0)$ denotes the *gamma function*.

The random variable χ^2_{n,σ^2} is also called a *chi-squared random variable with n degrees of freedom and parameter* σ^2.

It is well known that

$$E(\chi^2_{n,\sigma^2}) = n\sigma^2 \quad \text{and} \quad \mathrm{Var}(\chi^2_{n,\sigma^2}) = 2n\sigma^4. \tag{1.3.41}$$

Finally, by using the Kolmogorov's theorem about the existence of the infinite tensorial product of probability spaces, it is possible to show that, if $((\Omega_i, \mathscr{F}_i, P_i))_{i \in I}$ is an arbitrary family of probability spaces, then there exist a probability space $(\Omega, \mathscr{F}, P)$ and an independent family $(Z_i)_{i \in I}$ of random variables, $Z_i\colon \Omega \to \Omega_i$ $(i \in I)$, such that $P_{Z_i} = P_i$ for every $i \in I$ (for a proof see Bauer [1981, Corollary 5.4.5]).

Convergence of random variables

Consider again a probability space $(\Omega, \mathscr{F}, P)$. On the space $M(\Omega)$ of all real random variables on Ω we shall consider three different concepts of convergence that can be derived from suitable topologies.

A sequence $(Z_n)_{n \in \mathbb{N}}$ of real random variables is said to be *P-almost surely convergent* to a random variable Z if there exists a P-negligible set $N \subset \Omega$ (i.e., $N \in \mathscr{F}$ and $P(N) = 0$) such that

$$\lim_{n \to \infty} Z_n(\omega) = Z(\omega) \quad \text{for every } \omega \in \Omega \backslash N. \tag{1.3.42}$$

We say that $(Z_n)_{n \in \mathbb{N}}$ converges *P-stochastically* (or *stochastically*) to Z if

$$\lim_{n \to \infty} P\{|Z_n - Z| \geq \varepsilon\} = 0 \quad \text{for every } \varepsilon > 0. \tag{1.3.43}$$

Finally we say that $(Z_n)_{n \in \mathbb{N}}$ *converges in distribution* to Z if

$$\lim_{n \to \infty} P_{Z_n} = P_Z \quad \text{weakly.} \tag{1.3.44}$$

The logical relations between these concepts of convergence are as follows:

$$\begin{aligned}(P\text{-almost surely convergence}) &\Rightarrow (P\text{-stochastic convergence}) \Rightarrow \\ &\Rightarrow (\text{convergence in distribution}).\end{aligned}$$

In addition, if Z is P-almost surely constant and if $(Z_n)_{n \in \mathbb{N}}$ converges in distribution to Z, then $(Z_n)_{n \in \mathbb{N}}$ converges P-stochastically to Z.

Note also that $(Z_n)_{n \in \mathbb{N}}$ converges in distribution to Z if and only if $\lim_{n \to \infty} F_{Z_n}(x) = F_Z(x)$ for every point $x \in \mathbb{R}$ at which F_Z is continuous, or if the sequence of the Fourier transforms $(\hat{P}_{Z_n})_{n \in \mathbb{N}}$ converges pointwise to $\hat{P}_Z$ (*continuity theorem of* P. Lévy).

By Scheffé's theorem (see Section 1.2), if every Z_n and Z are Lebesgue-continuous with probability densities g_n and g respectively and if $g_n \to g$ P-almost everywhere, then $Z_n \to Z$ in distribution.

For the same reasons, if every Z_n and Z are discretely distributed and have a common support $\{a_k | k \in \mathbb{N}_0\} \subset \mathbb{R}$, i.e.,

$$P_{Z_n} = \sum_{k=0}^{\infty} \alpha_{n,k}\varepsilon_{a_k} \; (n \in \mathbb{N}) \quad \text{and} \quad P_Z = \sum_{k=0}^{\infty} \alpha_k \varepsilon_{a_k},$$

then $Z_n \to Z$ in distribution provided $\lim_{n\to\infty} \alpha_{n,k} = \alpha_k$ for each $k \in \mathbb{N}_0$.

A sequence $(Z_n)_{n \in \mathbb{N}}$ of integrable real random variables is said to obey the *strong law of large numbers* (the *weak law of large numbers*, respectively) if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (Z_i - E(Z_i)) = 0 \quad P\text{-almost surely} \tag{1.3.45}$$

(respectively,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (Z_i - E(Z_i)) = 0 \quad \text{stochastically}). \tag{1.3.46}$$

Two celebrated *theorems of Kolmogorov* show that an independent sequence $(Z_n)_{n \in \mathbb{N}}$ of real integrable random variables obeys the strong law of large numbers if

$$\sum_{n=1}^{\infty} \frac{\operatorname{Var}(Z_n)}{n^2} < +\infty, \tag{1.3.47}$$

or alternatively, if the Z_n's are *identically distributed* (i.e., $P_{Z_n} = P_{Z_m}$ for every n, $m \in \mathbb{N}$).

On the other hand, a sequence $(Z_n)_{n \in \mathbb{N}}$ of real integrable *pairwise uncorrelated* random variables (i.e., $E(Z_n Z_m) = E(Z_n)E(Z_m)$ for every $n \neq m$) obeys the weak law of large numbers if

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \operatorname{Var}(Z_i) = 0, \tag{1.3.48}$$

(*theorem of Markov*) or alternatively, if the Z_n's are identically distributed (*theorem of Khinchin*).

For more details see Bauer [1981, Chapter 6 and Sections 7.7 and 8.2].

1.4 Selected topics on locally convex spaces

In this section we survey some classical results on locally convex vector spaces such as various Hahn-Banach extension and separation theorems, the Krein-Milman theorem and Milman's converse theorem.

As a reference for these topics see for instance Choquet [1969] and Horváth [1966].

Let E be a topological vector space over the field $\mathbb{K}$ of real or complex numbers. A point $x \in E$ is said to be a *convex combination* of n given points $x_1, \ldots, x_n \in E$ if $x = \sum_{i=1}^{n} \lambda_i x_i$ for some $\lambda_1, \ldots, \lambda_n \geq 0$ such that $\sum_{i=1}^{n} \lambda_i = 1$.

A subset X of E is said to be *convex* if $\lambda x + (1 - \lambda)y \in X$ for every $x, y \in X$ and $\lambda \in [0, 1]$ or, equivalently, if $\sum_{i=1}^{n} \lambda_i x_i \in X$ for every finite family $(x_i)_{1 \leq i \leq n}$ of elements of X and for every $\lambda_1, \ldots, \lambda_n \geq 0$ such that $\sum_{i=1}^{n} \lambda_i = 1$.

The *convex hull* of a subset X of E is, by definition, the smallest convex subset of E containing X and is denoted by $\mathrm{co}(X)$. In fact, we have

$$\mathrm{co}(X) = \left\{ \sum_{i=1}^{n} \lambda_i x_i \,|\, x_i \in X, \lambda_i \geq 0, i = 1, \ldots, n \text{ and } \sum_{i=1}^{n} \lambda_i = 1 \right\}. \tag{1.4.1}$$

In general, if X is compact, $\mathrm{co}(X)$ is not compact. If E is Hausdorff and complete, then the closure $\overline{\mathrm{co}}(X)$ of $\mathrm{co}(X)$ is compact.

A topological vector space is said to be *locally convex* if the origin possesses a fundamental system of convex neighborhoods.

In fact, in a locally convex vector space the origin possesses a fundamental system of balanced, closed and convex neighborhoods (we recall that a subset X of E is said to be *balanced* if $\lambda x \in X$ for each $x \in X$ and $\lambda \in \mathbb{K}$, $|\lambda| \leq 1$).

Given a topological vector space E, we shall denote by E' the space of all continuous linear forms on E and call it the *dual space* of E.

On E' we shall often consider the *weak*-topology* $\sigma(E', E)$ which is, by definition, the coarsest topology on E' for which all the linear forms $\psi_x (x \in E)$ are continuous, where

$$\psi_x(\varphi) := \varphi(x) \quad \text{for every } x \in E \text{ and } \varphi \in E'. \tag{1.4.2}$$

Thus a net $(\varphi_i)_{i \in I}^{\leq}$ in E' converges to an element φ in $(E', \sigma(E', E))$ if and only if $\lim_{\substack{\leq \\ i \in I}} \varphi_i(x) = \varphi(x)$ for every $x \in E$.

The dual space E', endowed with the weak*-topology, is a locally convex Hausdorff space. Furthermore, if for every $x \in E$, $x \neq 0$, there exists $\varphi \in E'$ such that $\varphi(x) \neq 0$, then the dual space of $(E', \sigma(E', E))$ can be identified with E itself, that is, if a linear form $\psi: E' \to \mathbb{K}$ is continuous for the topology $\sigma(E', E)$, then there exists a unique $x \in E$ such that $\psi = \psi_x$.

A useful criterion to decide if a subset $\mathfrak{A}$ of E' is $\sigma(E', E)$-compact is furnished by the Alaoglu-Bourbaki theorem.

We recall that a subset $\mathfrak{A}$ of E' is *equicontinuous* if for every $\varepsilon > 0$ there exists a neighborhood V of the origin of E such that $|\varphi(x)| \leq \varepsilon$ for every $x \in V$ and $\varphi \in \mathfrak{A}$.

More generally, a subset $\mathfrak{A}$ of linear mappings from E into another topological vector space F is said to be *equicontinuous* if for every neighborhood W of the origin of F there exists a neighborhood V of the origin of E such that $u(x) \in W$ for every $x \in V$ and $u \in \mathfrak{A}$.

If E and F are both normed spaces, then $\mathfrak{A}$ is equicontinuous if and only if each $u \in \mathfrak{A}$ is continuous and $\sup\{\|u\| \,|\, u \in \mathfrak{A}\} < +\infty$.

If E is a Banach space and $\mathfrak{A}$ is a subset of continuous linear mappings such that $\sup\{\|u(x)\| \,|\, u \in \mathfrak{A}\} < +\infty$ for every $x \in E$, then $\mathfrak{A}$ is equicontinuous (*uniform boundedness principle*) (for a proof see Choquet [1969, Theorem 7.4]).

1.4.1 Theorem (*Alaoglu-Bourbaki*). *Given a topological vector space E, then every equicontinuous subset of E' is relatively compact for the weak*-topology.*

For a proof of the above result see Hórvath [1966, Chapter 3, Section 4, Theorem 1].

In the sequel we shall present several extension and separation theorems.

The first one concerns the extension of positive linear forms.

To this end we recall that an *ordered vector space* is a real vector space E endowed with a partial ordering $\leq$ satisfying the following properties:

$$x + z \leq y + z \quad \text{for every } x, y, z \in E, x \leq y, \tag{1.4.3}$$

and

$$\lambda x \leq \lambda y \quad \text{for every } x, y \in E, x \leq y \text{ and } \lambda \geq 0. \tag{1.4.4}$$

A *vector lattice* is an ordered vector space E such that for every $x, y \in E$ there exists $\sup\{x, y\}$ in E.

In this case we set $x \vee y := \sup\{x, y\}$, $x \wedge y := -\sup\{-x, -y\} = \inf\{x, y\}$, $|x| := \sup\{-x, x\}$, $x^+ := \sup\{x, 0\}$ and $x^- := \sup\{-x, 0\}$.

A linear form $\varphi: E \to \mathbb{R}$ is said to be positive if $\varphi(x) \geq 0$ for every $x \in E$, $x \geq 0$.

The proof of the next result can be found in Choquet [1969, Theorem 34.2].

1.4.2 Theorem. *Let E be an ordered vector space and F a subspace of E such that for every $x \in E$ there exists $y \in F$ satisfying $x \leq y$.*

Given a positive linear form $\varphi: F \to \mathbb{R}$, then for every $x \in E$ and $\alpha \in \mathbb{R}$ satisfying

$$\sup_{\substack{y \in F \\ y \leq x}} \varphi(y) \leq \alpha \leq \inf_{\substack{z \in F \\ x \leq z}} \varphi(z),$$

there exists a positive linear form $\tilde{\varphi}$ on E satisfying $\tilde{\varphi}(x) = \alpha$ and extending φ over E.

From this result one can derive the classical Hahn-Banach theorem. To state it we recall that, given a vector space E, a mapping $p: E \to \mathbb{R}$ is said to be *sublinear* if

$$p(\lambda x) = \lambda p(x) \quad \text{for every } x \in E \text{ and } \lambda \geq 0, \tag{1.4.5}$$

and

$$p(x + y) \leq p(x) + p(y) \quad \text{for every } x, y \in E. \tag{1.4.6}$$

A *seminorm* $p: E \to \mathbb{R}$ is a sublinear mapping such that $p(-x) = p(x)$ for each $x \in E$.

1.4.3 Theorem (*Hahn-Banach*). *Let E be a real vector space and $p: E \to \mathbb{R}$ a sublinear mapping. If F is a subspace of E and $\varphi: F \to \mathbb{R}$ is a linear form satisfying $\varphi \leq p_{|F}$, then there exists a linear form $\tilde{\varphi}: E \to \mathbb{R}$ satisfying $\tilde{\varphi} \leq p$ and extending φ over E.*

The following corollaries are direct consequences of the above result. For a proof we refer to Choquet [1969, Section 21] or Hórvath [1966, Chapter 3, Section 1].

1.4.4 Corollary. *Given a real locally convex space E and a continuous seminorm $p: E \to \mathbb{R}$, then for every $x_0 \in E$ there exists $\varphi \in E'$ such that $\varphi(x_0) = p(x_0)$.*

Then, if E is Hausdorff, for every $x_0 \in E$ there exists $\varphi \in E'$ such that $\varphi(x_0) \neq 0$.

1.4.5 Corollary. *Let E be a real normed space, F a subspace of E and $\varphi \in F'$. Then there exists $\tilde{\varphi} \in E'$ such that $\tilde{\varphi}_{|F} = \varphi$ and $\|\tilde{\varphi}\| = \|\varphi\|$.*

From Corollary 1.4.5 (that holds also for complex normed spaces) it follows in particular that, if E is a normed space and $x_0 \in E$, $x_0 \neq 0$, then there always exists $\varphi \in E'$ such that $\varphi(x_0) = \|x_0\|$ and $\|\varphi\| = 1$.

The Hahn-Banach theorem has a wide range of applications.

However, in some cases the domination is required only on a cone and the sublinear mapping may attain the value $+\infty$.

In these settings several extension theorems are available. Here we shall quote a useful one due to Anger and Lembcke [1974].

Let E be a real locally convex Hausdorff space and $P \subset E$ a *convex cone* (i.e., $x + y \in P$ and $\lambda x \in P$ for every $x, y \in P$ and $\lambda > 0$) such that $0 \in P$.

A mapping $p: P \to \mathbb{R} \cup \{+\infty\}$ which satisfies (1.4.5) and (1.4.6) for every $x, y \in P$ and $\lambda \geq 0$ is said to be a *hypolinear mapping* (in (1.4.5) the convention $0 \cdot (+\infty) = 0$ must be observed).

If $p(P) \subset \mathbb{R}$ and $p(x + y) = p(x) + p(y)$ for every $x, y \in P$, p is called *linear*.

Consider two convex cones P and C of E such that $0 \in P \cap C$. Consider a hypolinear mapping $p: P \to \mathbb{R} \cup \{+\infty\}$ and a linear mapping $\varphi: C \to \mathbb{R}$.

For every $x \in E$ we set

$$p_\varphi(x) := \inf\{p(x_1) + \varphi(y_1) - \varphi(y_2) | x_1 \in P, y_1, y_2 \in C, x_1 + y_1 - y_2 = x\} \tag{1.4.7}$$

(with the convention $\inf \varnothing = +\infty$) and

$$\hat{p}_\varphi(x) := \liminf_{y \to x} p_\varphi(y). \tag{1.4.8}$$

Then $\hat{p}_\varphi$ is the largest lower semi-continuous minorant of p_φ.

1.4.6 Theorem (*Anger-Lembcke*). *If $\hat{p}_\varphi(0) > -\infty$, then for every $x \in E$ and for every $\alpha \in]-\hat{p}_\varphi(-x), \hat{p}_\varphi(x)[$ there exists $\tilde{\varphi} \in E'$ such that $\tilde{\varphi}_{|C} = \varphi$, $\tilde{\varphi} \leq p$ on P and $\varphi(x) = \alpha$.*

Note that, if $C = \{0\}$, $\varphi = 0$ and $P = E$, then $p_\varphi = p$ so that $\hat{p}_\varphi = p$ provided p is lower semi-continuous on E. Accordingly we have the following result.

1.4.7 Corollary. *Let E be a real locally convex Hausdorff space and $p: E \to \mathbb{R} \cup \{+\infty\}$ a hypolinear lower semi-continuous mapping. Then for every $x \in E$ and $\alpha \in]-p(-x), p(x)[$ there exists $\varphi \in E'$ such that $\varphi \leq p$ and $\varphi(x) = \alpha$.*

Other important consequences of the Hahn-Banach theorem concern the separation of convex sets.

Let E be a real vector space. An *affine subspace* G of E is a translate of a subspace of E, i.e., there exist a subspace F of E and $a \in E$ such that $G = F + a := \{x + a | x \in F\}$.

A *hyperplane* in E is a subspace of E of codimension 1. A subspace G of E is a hyperplane if and only if there exists a linear form $\varphi: E \to \mathbb{R}$ such that $G = \{x \in E | \varphi(x) = 0\}$.

Note that each hyperplane is either closed or dense in E.

An *affine hyperplane* is a translate of a hyperplane. Thus an affine hyperplane is necessarily of the form $\{x \in E \mid \varphi(x) = \lambda\}$ where φ is a linear form on E and $\lambda \in \mathbb{R}$.

Given two subsets U and V of E, we say that a hyperplane $G = \{x \in E \mid \varphi(x) = \lambda\}$ *separates* U and V (*separates strictly* U and V, respectively) if

$$U \subset \{x \in E \mid \varphi(x) \geq \lambda\} \quad \text{and} \quad V \subset \{x \in E \mid \varphi(x) \leq \lambda\} \tag{1.4.9}$$

(respectively,

$$U \subset \{x \in E \mid \varphi(x) > \lambda\} \quad \text{and} \quad V \subset \{x \in E \mid \varphi(x) < \lambda\}). \tag{1.4.10}$$

In the next result we collect some important separation theorems (see Choquet [1969, Section 21]).

1.4.8 Theorem. *Let E be a real topological vector space. Then the following statements hold*:

(1) *If U is a non-empty open convex subset of E and G an affine subspace of E disjoint from U, then there exists a closed affine hyperplane disjoint from U which contains G.*

(2) *If U and V are non-empty disjoint convex subsets of E with U open, then there exists a closed affine hyperplane which separates U and V.*

If V is also open, the separation is strict.

(3) *If E is locally convex and Hausdorff and if U and V are disjoint closed convex subsets of E with U compact, then there exists a closed affine hyperplane which separates strictly U and V.*

The above separation theorems can be fruitfully used in the approximation of lower semi-continuous convex functions on convex compact sets.

Let E be a locally convex Hausdorff space and K a convex compact subset of E.

A function $u\colon K \to \mathbb{R}$ is said to be *affine* (*convex*, *concave*, respectively) if for every $x, y \in K$ and $\lambda \in [0,1]$

$$u(\lambda x + (1-\lambda)y) = \lambda u(x) + (1-\lambda)u(y), \tag{1.4.11}$$

$(u(\lambda x + (1-\lambda)y) \leq \lambda u(x) + (1-\lambda)u(y)$, $u(\lambda x + (1-\lambda)y) \geq \lambda u(x) + (1-\lambda)u(y)$, respectively).

In this case, for every $x_1, \ldots, x_n \in K$, $n \geq 2$, and $\lambda_1, \ldots, \lambda_n \geq 0$ satisfying $\sum_{i=1}^{n} \lambda_i = 1$, one has

$$u\left(\sum_{i=1}^{n} \lambda_i x_i\right) = \sum_{i=1}^{n} \lambda_i u(x_i) \tag{1.4.12}$$

$\left(u\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i u(x_i), u\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq \sum_{i=1}^{n} \lambda_i u(x_i)\right.$, respectively$\Big)$.

We shall denote by $A(K)$ the closed subspace of $\mathscr{C}(K)$ consisting of all continuous affine functions on K.

Moreover we shall denote by $A(K, E)$ the subspace of all restrictions to K of continuous affine functions on E. Note that

$$A(K, E) = \{\varphi_{|K} + \lambda | \varphi \in E', \lambda \in \mathbb{R}\}. \tag{1.4.13}$$

1.4.9 Theorem. *Let E be a real locally convex Hausdorff space and K a convex compact subset of E. Then the following statements hold:*

(1) *If $u: K \to \mathbb{R}$ is a convex lower semi-continuous function, then for every $x \in K$*

$$u(x) = \sup\{a(x) | a \in A(K, E), a < u\}.$$

(2) *(Mokobodzki) If $u: K \to \mathbb{R}$ is an upper semi-continuous function, then for every $x \in K$*

$$u(x) = \inf\{v(x) | v \text{ convex continuous function}, u < v\}.$$

Moreover, if u is also convex, then there exists a net $(u_i)_{i \in I}^{\leq}$ of convex continuous functions on K which converges pointwise to u.

(3) *If $u \in A(K)$, then there exists an increasing sequence in $A(K, E)$ which converges uniformly to u. Thus, $A(K, E)$ is dense in $A(K)$.*

For a proof of the above theorem see Alfsen [1971, Proposition I.1.2, Corollary I.1.5 and Proposition I.5.1].

Now we list some properties of convex compact sets including the classical Krein-Milman theorem.

Let K be a convex compact subset of a locally convex Hausdorff space. A point $x_0 \in K$ is said to be an *extreme point* of K if $K \backslash \{x_0\}$ is convex or, equivalently, if for every $x_1, x_2 \in K$ and $\lambda \in]0, 1[$ satisfying $x_0 = \lambda x_1 + (1 - \lambda) x_2$, it necessarily follows that $x_0 = x_1 = x_2$.

The set of all extreme points of K will be denoted by $\partial_e K$.

For example, if X is a compact Hausdorff space and $K = \mathscr{M}_1^+(X)$ endowed with the vague topology, then

$$\partial_e K = \{\varepsilon_x | x \in X\}, \tag{1.4.14}$$

while, if $K = \{\mu \in \mathscr{M}(X) | \|\mu\| \leq 1\}$, then

$$\partial_e K = \{\varepsilon_x | x \in X\} \cup \{-\varepsilon_x | x \in X\}. \tag{1.4.15}$$

An important topological property of $\partial_e K$ which was proved by Choquet (see Choquet [1969, Theorem 27.9]) is that $\partial_e K$ is a Baire space in the relative topology (i.e., the intersection of every sequence of dense open subsets of $\partial_e K$ in dense in $\partial_e K$).

Moreover, if K is metrizable, $\partial_e K$ is a countable intersection of open subsets of K and hence it is a complete metric space.

We also recall that a *ray* of a locally convex Hausdorff space E is a subset δ of E of the form

$$\delta = \delta_{x_0} := \{\lambda x_0 | \lambda > 0\}, \tag{1.4.16}$$

where $x_0 \in E$, $x_0 \neq 0$.

Clearly every affine subspace which does not contain the origin and which has non-empty intersection with a ray δ intersects the ray in a unique point.

Given a convex cone C of E, a ray $\delta \subset C$ is said to be an *extreme ray* of C if $C \backslash \delta$ is convex.

This also means that every affine subspace G of E not containing the origin and having non-empty intersection with δ intersects δ in an extreme point of $G \cap C$.

If C is *proper*, i.e., $C \cap (-C) = \{0\}$, then a ray δ of C is an extreme ray if and only if for every $a \in \delta$ and $b \in C$ such that $a - b \in C$, there exists $\lambda \geq 0$ such that $b = \lambda a$ (and hence $b \in \delta$).

If the convex cone C has a compact *base* K (i.e., K is the intersection of C with an affine hyperplane and for every $y \in C$ there exist $x \in K$ and $\lambda > 0$ such that $y = \lambda x$), then a ray δ is an extreme ray of C if and only if $\delta = \delta_{x_0}$ for some $x_0 \in \partial_e K$ (Choquet [1953]).

Note that in general a convex cone may have no extreme rays, while every convex compact subset has extreme points. This will follow from the Bauer's maximum principle for convex compact sets (see Choquet [1969, Theorem 25.9]).

1.4.10 Theorem (*Bauer's maximum principle*). *Let K be a convex compact subset of a real locally convex Hausdorff space and u: $K \to \mathbb{R}$ a convex upper semi-continuous function. Then there exists $x_0 \in \partial_e K$ such that $u(x_0) = \max\{u(x) | x \in K\}$.*

In particular, $\partial_e K$ is non-empty.

In addition to the above result, the next comparison principle is useful in comparing semi-continuous convex and concave functions. For a proof see Bauer [1963].