Graduate Texts in Mathematics

Paul Malliavin

with Hélène Airault, Leslie Kay, Gérard Letac

Integration and Probability

Springer

Graduate Texts in Mathematics 157

Springer
New York
Berlin
Heidelberg
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.

continued after index

Paul Malliavin
In Cooperation with Hélène Airault,
Leslie Kay, Gérard Letac

Integration and Probability

Springer

Paul Malliavin
10 rue Saint Louis
en L'Isle
F-75004 Paris
France

Hélène Airault
Mathématiques--INSSET
Université de Picardie Jules-Verne
48 rue Raspail
F-02100 Saint-Quentin (Aisne), France

Leslie Kay
Department of Mathematics
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24061, USA

Gérard Letac
Laboratoire de Statistique
Université Paul Sabatier
118 Route de Narbonne
F-31062 Toulouse, France

Second French Edition: *intégration, analyse de Fourier, probabilités, analyse gaussienne*

Mathematics Subject Classification (1991): 28-01, 43A25, 60H07

Library of Congress Cataloging-in-Publication Data
Malliavin, Paul, 1925-
[Intégration et probabilités . English]
Integration and probability / Paul Malliavin in cooperation with
Hélène Airault, Leslie Kay, and Gérard Letac.
p. cm. – (Graduate texts in mathematics ; 157)
Includes bibliographical references and index.
ISBN-13:978-1-4612-8694-3 e-ISBN-13:978-1-4612-4202-4
DOI: 10.1007/978-1-4612-4202-4
1. Calculus, Integral. 2. Spectral theory (Mathematics)
3. Fourier analysis. I. Title. II. Series.
QA308.M2713 1995
515'.4–dc20 94-38936

Printed on acid-free paper.

Production managed by Frank Ganz; manufacturing supervised by Joe Quatela.
Photocomposed pages prepared from the translator's LaTeX file.

9 8 7 6 5 4 3 2

Foreword

It is a distinct pleasure to have the opportunity to introduce Professor Malliavin's book to the English-speaking mathematical world.

In recent years there has been a noticeable retreat from the level of abstraction at which graduate-level courses in analysis were previously taught in the United States and elsewhere. In contrast to the practices used in the 1950s and 1960s, when great emphasis was placed on the most general context for integration and operator theory, we have recently witnessed an increased emphasis on detailed discussion of integration over Euclidean space and related problems in probability theory, harmonic analysis, and partial differential equations.

Professor Malliavin is uniquely qualified to introduce the student to analysis with the proper mix of abstract theories and concrete problems. His mathematical career includes many notable contributions to harmonic analysis, complex analysis, and related problems in probability theory and partial differential equations. Rather than developed as a thing-in-itself, the abstract approach serves as a context into which special models can be couched. For example, the general theory of integration is developed at an abstract level, and only then specialized to discuss the Lebesgue measure and integral on the real line. Another important area is the entire theory of probability, where we prefer to have the abstract model in mind, with no other specialization than total unit mass. Generally, we learn to work at an abstract level so that we can specialize when appropriate.

A cursory examination of the contents reveals that this book covers most of the topics that are familiar in the first graduate course on analysis. It also treats topics that are not available elsewhere in textbook form. A notable example is Chapter V, which deals with Malliavin's stochastic calculus of variations developed in the context of Gaussian measure spaces. Originally inspired by the desire to obtain a probabilistic proof of Hörmander's theorem on the smoothness of the solutions of second-order hypoelliptic differential equations, the subject has found a life of its own. This is partly due to Malliavin and his followers' development of a suitable notion of "differentiable function" on a Gaussian measure space. The novice should be warned that this notion of differentiability is not easily related to the more conventional notion of differentiability in courses on manifolds. Here we have a family of Sobolev spaces of "differentiable functions" over the measure space, where the definition is global, in terms of the Sobolev norms. The finite-dimensional Sobolev spaces are introduced through translation operators, and immediately generalizes to the infinite-dimensional case. The main theorem of the subject states that if a differentiable vector-valued function has enough "variation", then it induces a smooth measure on Euclidean space.

Such relations illustrate the interplay between the "upstairs" and the "downstairs" of analysis. We find the natural proof of a theorem in real analysis (smoothness of a measure) by going up to the infinite-dimensional Gaussian measure space where the measure is naturally defined. This interplay of ideas can also be found in more traditional forms of finite-dimensional real analysis, where we can better understand and prove formulas and theorems on special functions on the real line by going up to the higher-dimensional geometric problems from which they came by "projection"; Bessel and Legendre functions provide some elementary examples of such phenomena.

The mathematical public owes an enormous debt of gratitude to Leslie Kay, whose superlative efforts in editing and translating this text have been accomplished with great speed and accuracy.

Mark Pinsky
Department of Mathematics
Northwestern University
Evanston, IL 60208, USA

Preface

We plan to survey various extensions of Lebesgue theory in contemporary analysis: the abstract integral, Radon measures, Fourier analysis, Hilbert spectral analysis, Sobolev spaces, pseudo-differential operators, probability, martingales, the theory of differentiation, and stochastic calculus of variations.

In order to give complete proofs within the limits of this book, we have chosen an axiomatic method of exposition; the interest of the concepts introduced will become clear only after the reader has encountered examples later in the text. For instance, the first chapter deals with the abstract integral, but the reader does not see a nontrivial example of the abstract theory until the Lebesgue integral is introduced in Chapter II. This axiomatic approach is now familiar in topology; it should not cause difficulties in the theory of integration.

In addition, we have tried as much as possible to base each theory on the results of the theories presented earlier. This structure permits an economy of means, furnishes interesting examples of applications of general theorems, and above all illustrates the unity of the subject. For example, the Radon-Nikodym theorem, which could have appeared at the end of Chapter I, is treated at the end of Chapter IV as an example of the theory of martingales; we then obtain the stronger result of convergence almost everywhere. Similarly, conditional probabilities are treated using (i) the theory of Radon measures and (ii) a general isomorphism theorem showing that there exists only one model of a nonatomic separable measure space, namely $\mathbf{R}$ equipped with Lebesgue measure. Furthermore, the spectral theory of unitary operators on an abstract Hilbert space is derived from

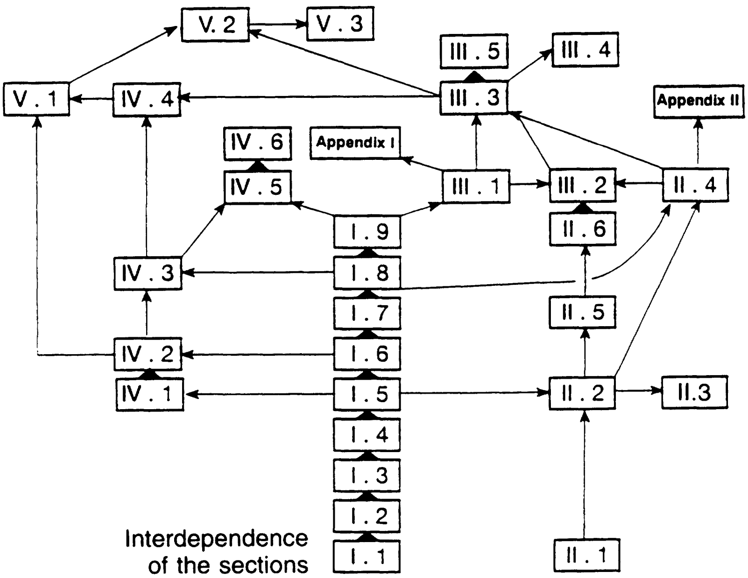
Bochner's theorem characterizing Fourier series of measures. The treatment in Chapter V of Sobolev spaces over a probability space parallels that in Chapter III of Sobolev spaces over $\mathbf{R}^n$.

In the detailed table of contents, the reader can see how the book is organized. It is easy to read only selected parts of the book, depending on the results one hopes to reach; at the beginning of the book, as a reader's guide, there is a diagram showing the interdependence of the different sections. There is also an index of terms at the end of the work. Certain parts of the text, which can be skipped on a first reading, are printed in smaller type.

Readers interested in probability theory can focus essentially on Chapters I, IV, and V; those interested in Fourier analysis, essentially on Chapters I and III. Chapter III can be read in different ways, depending on whether one is interested in partial differential equations or in spectral analysis.

The book includes a variety of exercises by Gérard Letac. Detailed solutions can be found in *Exercises and Solutions Manual for Integration and Probability* by Gérard Letac, Springer-Verlag, 1995. The upcoming book *Stochastic Analysis* by Paul Malliavin, Grundlehren der Mathematischen Wissenschaften, volume 313, Springer-Verlag, 1995, is meant for second-year graduate students who are planning to continue their studies in probability theory.

March 1995 P. M.



Interdependence of the sections

Contents

Index of Notation

Set-theoretic notation:
A^c denotes the complement of A. $A - B = A \cap B^c$.

The sign $\square$ indicates the end of a proof.

Prologue

We recall briefly the definition and properties of the usual integral of continuous functions on $\mathbf{R}$.

The concepts involved are elementary and well known. However, since this integral will be used to construct the Lebesgue integral, we sketch a few facts for convenience.

Given the segment $[0,1] \subset \mathbf{R}$, a *partition* of $[0,1]$ is a finite subset π of $[0,1]$ containing 0 and 1. The partition π' is said to be finer than π if $\pi' \supset \pi$. Let $0 = t_1 < t_2 < \ldots < t_{r-1} < t_r = 1$ ($r = \operatorname{card}(\pi)$) be an enumeration of the points of π. With every function f continuous on $[0,1]$, we associate the sum

$$s_\pi(f) = \sum_{k=1}^{r-1} (t_{k+1} - t_k) f(t_k).$$

This is a positive linear functional:

$$s_\pi(f_1 + f_2) = s_\pi(f_1) + s_\pi(f_2) \quad \text{and} \quad s_\pi(f) \geq 0 \quad \text{if} \quad f \geq 0.$$

The number $\delta(\pi) = \sup(t_{k+1} - t_k)$ is called the *diameter* of the partition π. We have the following statement.

Given a continuous function f, for every $\epsilon > 0$ there exists η such that

$$|s_\pi(f) - s_{\pi'}(f)| < \epsilon$$

for any partitions π and π' satisfying $\delta(\pi) < \eta$ and $\delta(\pi') < \eta$.

Indeed, since f is continuous on the compact set $[0,1]$, it is uniformly continuous. Hence we can find η such that $|f(x) - f(x')| \leq \frac{\epsilon}{2}$ if $|x - x'| < \eta$.

Let $\pi'' = \pi \cup \pi'$. Then, writing $\pi''_k = \pi' \cup [t_{k+1} - t_k]$, where $t_1, \ldots, t_r$ denote the points of the subdivision of π,

$$\pi'' = \cup \pi''_k \quad \text{and} \quad s_{\pi''}(f) = \sum_{k=1}^{r-1} s_{\pi''_k}(f).$$

Moreover,

$$|s_{\pi''_k}(f) - (t_{k+1} - t_k) f(t_k)| < \frac{\epsilon}{2}(t_{k+1} - t_k),$$

whence

$$|s_\pi(f) - s_{\pi''}(f)| < \frac{\epsilon}{2} \sum (t_{k+1} - t_k) = \frac{\epsilon}{2}$$

and

$$|s_\pi(f) - s_{\pi'}(f)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Choosing a sequence π_k of partitions such that $\delta(\pi_k) \to 0$, we find that $s_{\pi_k}(f)$ is a Cauchy sequence whose limit is independent of the choice π_k. Set

$$\int_0^1 f(x)\, dx = \lim s_{\pi_k}(f).$$

Then the integral is a positive linear functional. In particular,

$$\left| \int_0^1 f(x)\, dx \right| \leq \int_0^1 |f(x)|\, dx \leq \max |f(x)|.$$

The change of variable $x = a + t(b - a)$ reduces the integral over $[a, b]$ to the preceding case:

$$\int_a^b f(x)\, dx = \frac{1}{b-a} \int_0^1 f(a + t(b-a))\, dt.$$

Differentiation. Let f be continuous. Set

$$F(x) = \int_0^x f(t)\, dt.$$

Then F is differentiable and $F'(x) = f(x)$. Evaluating integrals of continuous functions is reduced to finding primitives.
Improper integrals. Integrals will be evaluated either on all of $\mathbf{R}$ or on $[0, 1]$. The functions we integrate on $\mathbf{R}$ will be continuous; those we integrate on $[0, 1]$ will be continuous on $(0, 1)$. The elementary procedure consists of passing to the limit:

$$\int = \lim_{n \to +\infty} \int_{-n}^{n}, \qquad \int_0^1 = \lim_{n \to +\infty} \int_{\frac{1}{n}}^{1 - \frac{1}{n}}.$$

We have the concepts of *convergence* and of *absolute convergence*. The Lebesgue theory will be developed in the second setting: every Lebesgue integrable function will have Lebesgue-integrable absolute value. For this reason, we consider here only absolutely convergent improper integrals. The following results can easily be proved by calculating primitives.

If f is continuous and positive on $\mathbf{R}$ and if $f(x) \sim |x|^{-\alpha}$ as $|x| \to +\infty$, then the integral of f on $\mathbf{R}$ exists if and only if $\alpha > 1$.

If f is continuous and positive on $(0, 1]$ and if $f(x) \sim |x|^{-\beta}$ as $x \to 0$, then $\int_0^1 f$ exists if and only if $\beta < 1$.

These results generalize to $\mathbf{R}^n$ by passing to polar coordinates. We find in the first case that $\alpha > n$, and in the second that $\beta < n$. (In the second case, we integrate a function continuous on $\mathbf{R}^n$ and zero outside a compact set.)

I

Measurable Spaces and Integrable Functions

Introduction

In this chapter, we follow an axiomatic method of exposition. The interest of the concepts introduced will not appear until Chapter II. We introduce the notion of a *measure space*, a space endowed with a family of *measurable subsets* satisfying the axioms of a σ-algebra. This approach parallels that of the theory of topological spaces, where a topological space is a space endowed with a family of open subsets. As we will see in Chapter IV, a peculiarity of the concept of a σ-algebra is that it is adapted to the propositional calculus (Boolean algebra). Since *negation* is an operation of this calculus, this leads to the axiom that the *complement* of a measurable set is measurable. The fact that σ-algebras are closed under taking complements is an essential difference between the family of open sets of a topological space and the family of measurable sets of a measure space. In order to be able to take limits of sequences, we impose another axiom: *A countable union of measurable sets is measurable.*

Having defined the concept of a measurable space, we introduce a class of morphisms adapted to it: the *measurable mappings*. We introduce a natural measurable structure on a topological space: the *Borel* structure. Continuous mappings are thus special cases of measurable mappings. A remarkable result is that the limit of a pointwise convergent sequence of measurable mappings is itself measurable. Thus all the functions appearing in practice in mathematical analysis are measurable functions. A *measure space* is a measurable space which is given a "mass distribution". The concept of *negligible sets*, or sets of measure zero, is introduced; two measurable mappings are considered *equivalent* if they differ on a negligible set.

We introduce the concept of *convergence in measure*, which gives a complete metric space structure to the space M of equivalence classes of measurable mappings from a measure space to a complete metric space. When we consider *functions* on a measure space, i.e. mappings with values in $\mathbf{R}$, we introduce *simple* functions, those that assume finitely many values. The integral, defined trivially on certain simple functions, extends to an appropriate completion, which defines the space L^1 of integrable functions. The theorems on passage to the limit under the integral sign are then an easy consequence of the fact that L^1 is a complete space. The chapter concludes with Fubini's theorem and the duality between L^p spaces.

1 σ-algebras

Let X be an abstract set. A *σ-algebra* on X is a family $\mathcal{A}$ of subsets of X satisfying the following three axioms:

1.0.1 *The set X belongs to $\mathcal{A}$.*

1.0.2 *If $A \in \mathcal{A}$, its complement $A^c \in \mathcal{A}$.*

1.0.3 *Every countable union of sets in $\mathcal{A}$ belongs to $\mathcal{A}$; i.e., if $A_n \in \mathcal{A}$ $\forall n \in \mathbf{N}$, then $(\cup_{n\in\mathbf{N}} A_n) \in \mathcal{A}$.*

A *Boolean algebra* on X is a family $\mathcal{B}$ of subsets of X satisfying 1.0.1, 1.0.2, and

1.0.4 *Every finite union of sets in the algebra $\mathcal{B}$ is in $\mathcal{B}$.*

Every σ-algebra is thus a Boolean algebra. By using Axiom 1.0.2 and passing to the complement, we find that 1.0.3 implies

1.0.5 *If $A_n \in \mathcal{A}$, then $\left(\cap_{n\in\mathbf{N}} A_n\right) \in \mathcal{A}$.*

An analogous statement is obtained for Boolean algebras by restricting to *finite* intersections. In what follows, we will not pursue the parallels between Boolean algebras and σ-algebras, but the reader should note that most theorems involving passage to the limit are false for Boolean algebras.

1.1 Sub-σ-algebras. Intersection of σ-algebras

Given two σ-algebras $\mathcal{A}$ and $\mathcal{A}'$ on the abstract set X, we say that $\mathcal{A}'$ is a sub-σ-algebra of $\mathcal{A}$ if $A \in \mathcal{A}'$ implies $A \in \mathcal{A}$. More formally, let $\mathcal{P}(X)$ denote the set of subsets of X. We may view a σ-algebra $\mathcal{A}$ on X as a subset of $\mathcal{P}(X)$. The "order relation" between σ-algebras corresponds to the relation of inclusion between the subsets of $\mathcal{P}(X)$.

1.1.1 More generally, if $\mathcal{G}$ is an arbitrary family of subsets of X and $\mathcal{A}$ is a σ-algebra on X, we say that $\mathcal{A} \supset \mathcal{G}$ if $A \in \mathcal{G}$ implies $A \in \mathcal{A}$.

1.1.2 Intersection of σ-algebras

Definition. Let $\{\mathcal{A}_\alpha, \alpha \in I\}$, be a family of σ-algebras on X. We denote by $\mathcal{A}' = \cap_{\alpha \in I} \mathcal{A}_\alpha$ the family of subsets of X defined by $A \in \mathcal{A}'$ if and only if $A \in \mathcal{A}_\alpha$ for all $\alpha \in I$. $\mathcal{A}'$ is a σ-algebra called the *intersection* of the $\mathcal{A}_\alpha$.

We verify only 1.0.3, the other axioms being even more obvious. Let $A_n \in \mathcal{A}'$, set

$$Z = \cup_{n \in \mathbf{N}} A_n,$$

and fix α_0. Since $A_n \in \mathcal{A}_{\alpha_0}$ and $\mathcal{A}_{\alpha_0}$ satisfies 1.0.3, it follows that $Z \in \mathcal{A}_{\alpha_0}$. As this is true for all α_0, we conclude that $Z \in \mathcal{A}'$.

1.2 σ-algebra generated by a family of sets

1.2.1 Theorem. *Let $\mathcal{G}$ be a family of subsets of X. Then there exists on X a smallest σ-algebra containing $\mathcal{G}$.*

PROOF. Consider the σ-algebras $\mathcal{B}$ on X such that

$$(P) \qquad \mathcal{B} \supset \mathcal{G}.$$

Let I denote the family of σ-algebras $\mathcal{B}$ satisfying (P), and set $\mathcal{A}_0 = \cap_{\mathcal{B} \in I} \mathcal{B}$. Then $\mathcal{A}_0$ is a σ-algebra by 1.1.2, and it is the smallest σ-algebra of the family I. □

1.2.2 **Definition.** $\mathcal{A}_0$ is called the *σ-algebra generated* by $\mathcal{G}$. We say that $\mathcal{G}$ is a system of *generators* of $\mathcal{A}_0$.

1.2.3 Fundamental example: Borel algebras

Let X be a topological space and let $\mathcal{O}_X$ be the family of open subsets of X. The σ-algebra generated by $\mathcal{O}_X$ is called a Borel algebra, and written $\mathcal{B}_X$.

An element of $\mathcal{B}_X$ is called a *Borel set*. Open sets are Borel sets, as are closed sets (as complements of open sets). The family of closed sets could equally well be taken as a system of generators of $\mathcal{B}_X$.

1.3 Limit of a monotone sequence of sets

1.3.1 **Definition.** Let A_n be an *increasing* sequence of subsets of X. We call the union of the A_n the *limit* of the sequence A_n, and we set

$$A_\infty = \lim \uparrow A_n = \bigcup_n A_n, \quad \text{where} \quad A_n \subset A_{n+1}.$$

Similarly, given a *decreasing* sequence B_n of subsets of X, we call the intersection of the B_n its *limit*:

$$B_\infty = \lim \downarrow B_n = \bigcap_n B_n, \quad \text{where} \quad B_n \supset B_{n+1}.$$

A sequence of subsets of X is *monotone* if it is either increasing or decreasing.

1.3.2 A *monotone class* is a family $\mathcal{M}$ of subsets of X such that if $\{A_n\}$ is a monotone sequence for which $A_n \in \mathcal{M}$ for each n, then its limit is in $\mathcal{M}$.

1.3.3 Proposition. *A σ-algebra is a monotone class.*

PROOF. Let $\{A_n\}$ be an increasing sequence of sets in the σ-algebra $\mathcal{A}$. Then by 1.0.3

$$\lim \uparrow A_n = \bigcup A_n \in \mathcal{A}.$$

Similarly, 1.0.5 proves the statement for decreasing sequences.

1.3.4 *An arbitrary intersection of monotone classes is a monotone class.*

Thus, given a family $\mathcal{Z}$ of subsets of X, there exists a smallest monotone class $\mathcal{M}_0$ containing $\mathcal{Z}$. $\mathcal{M}_0$ is called the monotone class *generated* by $\mathcal{Z}$.

1.4 Theorem. *Let $\mathcal{B}_b$ be a Boolean algebra of subsets of X, $\mathcal{M}$ the monotone class generated by $\mathcal{B}_b$, and $\mathcal{B}$ the σ-algebra generated by $\mathcal{B}_b$. Then $\mathcal{B} = \mathcal{M}$.*

PROOF. By 1.3.3, $\mathcal{B}$ is a monotone class. Since $\mathcal{B}$ contains $\mathcal{B}_b$, it contains the smallest monotone class containing $\mathcal{B}_b$; thus $\mathcal{B} \supset \mathcal{M}$.

Conversely, for all $A \in \mathcal{P}(X)$, let

1.4.1 $$\Phi(A) = \{B \in \mathcal{P}(X) : A \cup B, A - B, B - A \in \mathcal{M}\}.$$

Then the assertions $B \in \Phi(A)$ and $A \in \Phi(B)$ are equivalent.

Fixing A, we show that $\Phi(A)$ is a monotone class. Indeed, if B_n is an increasing sequence of elements of $\Phi(A)$, then

$$\begin{cases} A \cup B_n & \text{is an increasing sequence of elements of } \mathcal{M}, \\ B_n - A & \text{is an increasing sequence of elements of } \mathcal{M}, \\ A - B_n & \text{is a decreasing sequence of elements of } \mathcal{M}, \end{cases}$$

and their limits are elements of $\mathcal{M}$. Furthermore,

$$\lim \uparrow (A \cup B_n) = A \cup \lim \uparrow B_n,$$

whence $\lim \uparrow B_n \in \Phi(A)$.

Let $A_0 \in \mathcal{B}_b$; then $B_0 \in \Phi(A_0)$ for all $B_0 \in \mathcal{B}_b$. Hence $\Phi(A_0)$ is a monotone class containing $\mathcal{B}_b$. Thus $\Phi(A_0) \supset \mathcal{M}$, or $B \in \Phi(A_0)$ for any $A_0 \in \mathcal{B}_b$, $B \in \mathcal{M}$.

Conversely, $A_0 \in \Phi(B)$; i.e., $\Phi(B) \supset \mathcal{B}_b$ for any fixed $B \in \mathcal{M}$.

Since $\Phi(B)$ is a monotone class, it follows that $\Phi(B) \supset \mathcal{M}$.

We have proved that

1.4.2 $B - B'$, $B' - B$, $B \cup B' \in \mathcal{M}$ whenever B, $B' \in \mathcal{M}$.

Taking $B' = X$ shows that $B^c \in \mathcal{M}$ if $B \in \mathcal{M}$, and thus

1.4.3 $\mathcal{M}$ is a Boolean algebra.

The following lemma, 1.4.4, implies that $\mathcal{M}$ is a σ-algebra. Since $\mathcal{M} \supset \mathcal{B}_b$, $\mathcal{M}$ contains the σ-algebra generated by $\mathcal{B}_b$; hence $\mathcal{B} \subset \mathcal{M}$. □

1.4.4 Lemma. *Let $\mathcal{Z}$ be a Boolean algebra which is closed under increasing limits. (That is, if Z_n is an increasing sequence of elements of $\mathcal{Z}$, then* $\lim \uparrow Z_n \in \mathcal{Z}$.*) Then $\mathcal{Z}$ is a σ-algebra.*

PROOF. Let $A_n \in \mathcal{Z}$ and set $Z_n = \cup_{1\le p\le n} A_p$; then

$$\cup_n A_n = \cup_n Z_n = \lim \uparrow Z_n \in \mathcal{Z},$$

and Axiom 1.0.3 is satisfied.

1.5 Product σ-algebras

Definition. Let X_1, X_2 be abstract sets equipped with σ-algebras $\mathcal{A}_1$, $\mathcal{A}_2$, and let the Cartesian product $X_1 \times X_2$ be denoted by X.

1.5.1 A *rectangle* R is a subset of X of the form

$$R = A_1 \times A_2 \quad \text{with} \quad A_i \in \mathcal{A}_i \quad (i = 1, 2).$$

The set of all rectangles is denoted by $\mathcal{R}$.

1.5.2 The σ-algebra generated by $\mathcal{R}$ is called the *product σ-algebra* and denoted by $\mathcal{A}_1 \otimes \mathcal{A}_2$.

1.5.3 The union of a *finite number* of *disjoint* rectangles is called an *elementary set.* The family of elementary sets is denoted by $\mathcal{E}$.

1.5.4 Proposition. *The elementary sets form a Boolean algebra.*

PROOF. Note first that the union of a finite number of *disjoint* elementary sets is an elementary set.

Let $R = A_1 \times A_2$, $R' = A_1' \times A_2'$ be two rectangles; then

$$(R)^c = (A_1^c \times X_2) \cap (X_1 \times A_2^c).$$

Hence

$$R' - R = R_1 \cup R_2 \cup R_3,$$

where $R_1 = (A_1^c \cap A_1') \times (A_2 \cap A_2')$, $R_2 = (A_1 \cap A_1') \times (A_2^c \cap A_2')$, and $R_3 = (A_1^c \cap A_1') \times (A_2^c \cap A_2')$. Thus

(i) $$R' - R \quad \text{is an elementary set.}$$

Let $E = R \cup R_4$ be an elementary set that is the union of *two* disjoint rectangles. (We restrict to two in order to simplify notation.)

$$R' - E = (R' - R) - R_4 = (R_1 \cup R_2 \cup R_3) - R_4 = (R_1 - R_4) \cup (R_2 - R_4) \cup (R_3 - R_4).$$

Applying (i), we obtain

(ii) $$R' - E \quad \text{is an elementary set if} \quad E \in \mathcal{E}, \quad R' \in \mathcal{R}.$$

If $E' \in \mathcal{E}$ then $E' = \cup R_i$ (R_i disjoint) and $E' - E = \cup(R_i - E)$, whence

(iii) $$(E' - E) \in \mathcal{E} \quad \text{for any} \quad E, E' \in \mathcal{E}.$$

Taking $E' = X_1 \times X_2$, we obtain 1.0.2. Furthermore,

$$(A_1 \times A_2) \cap (A'_1 \times A'_2) = (A_1 \cap A'_1) \times (A_2 \cap A'_2).$$

Hence the intersection of two rectangles is a rectangle and, more generally,

$$(iv) \qquad E \cap E' \in \mathcal{E} \quad \text{if} \quad E, E' \in \mathcal{E}.$$

Indeed, if $E = \cup R_j$ and $E' = \cup R'_q$, then $E \cap E' = \cup_{j,q}(R_j \cap R'_q)$. (Note that the sets $R_j \cap R'_q$ are disjoint.)

Finally,

$$(v) \qquad E \cup E' = (E - E') \cup (E' - E) \cup (E \cap E').$$

The three quantities in parentheses on the right-hand side are elementary sets by (ii) and (iv); since they are disjoint, $E \cup E' \in \mathcal{E}$ and 1.0.4 is satisfied.

1.5.5 Corollary. *The σ-algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$ is the monotone class generated by the elementary sets.*

PROOF. 1.5.4 and 1.4.

2 Measurable Spaces

2.1 Inverse image of a σ-algebra

Let X, X' be abstract sets and let f be a mapping from X to X'. Let $\mathcal{G}'$ be a family of subsets of X'. We write

$$f^{-1}(\mathcal{G}') = \{A \in \mathcal{P}(X) : A = f^{-1}(A') \quad \text{with} \quad A' \in \mathcal{G}'\}.$$

2.1.1 Proposition. *Let $\mathcal{A}'$ be a σ-algebra on X'; then $f^{-1}(\mathcal{A}')$ is a σ-algebra on X. It is called the inverse image of $\mathcal{A}'$ under f and denoted by $\mathcal{A} = f^{-1}(\mathcal{A}')$.*

PROOF. The inverse image of X' is X. In addition,

$$\begin{array}{rcll} \cup_s f^{-1}(A'_s) & = & f^{-1}(\cup_s A'_s) & \quad \text{(Axiom 1.0.3 is satisfied);} \\ \left[f^{-1}(A')\right]^c & = & f^{-1}(A'^c) & \quad \text{(Axiom 1.0.2 is satisfied).} \end{array}$$

2.1.2 *Taking the inverse image preserves inclusion between σ-algebras: $f^{-1}(\mathcal{A}'_1) \supset f^{-1}(\mathcal{A}'_2)$ whenever $\mathcal{A}'_1 \supset \mathcal{A}'_2$.*

2.1.3 EXAMPLE. Let Y be a subset of the set X', let i be the canonical injection of Y into X', and let $\mathcal{A}'$ be a σ-algebra on X'. Then

$$\begin{array}{rcl} i^{-1}(\mathcal{A}') & = & \{B \in \mathcal{P}(Y) : i^{-1}(B) \in \mathcal{A}'\} \\ & = & \{B \in \mathcal{P}(Y) : \exists A' \in \mathcal{A}' \text{ such that } A' \cap Y = B\}. \end{array}$$

In this special case, $i^{-1}(\mathcal{A}')$ is called the *trace σ-algebra* of the σ-algebra $\mathcal{A}'$ on the subset Y.

Since Y is a subset of X', every subset of Y can be identified with a subset of X'. It is easy to verify that

$$(i) \qquad i^{-1}(\mathcal{A}') \subset \mathcal{A}' \quad \Leftrightarrow \quad Y \in \mathcal{A}'.$$

2.1.4 Transitivity of inverse images

Suppose that X, X', and X'' are three abstract sets, f and h are mappings such that $X \overset{f}{\mapsto} X' \overset{h}{\mapsto} X''$, and $\mathcal{G}''$ is a family of subsets of X''. Then

$$f^{-1}(h^{-1}(\mathcal{G}'')) = (h \circ f)^{-1}(\mathcal{G}'').$$

2.2 Closure under inverse images of the generated σ-algebra

2.2.1 Theorem. *Suppose that X and X' are abstract sets, f is a mapping from X to X', $\mathcal{G}'$ is a family of subsets of X', and $\mathcal{A}'$ is the σ-algebra generated by $\mathcal{G}'$. Then $f^{-1}(\mathcal{A}')$ is the σ-algebra generated by $f^{-1}(\mathcal{G}')$.*

PROOF. Let $\mathcal{B}$ denote the σ-algebra generated by $f^{-1}(\mathcal{G}')$.
$\mathcal{B} \subset f^{-1}(\mathcal{A}')$ since $f^{-1}(\mathcal{G}') \subset f^{-1}(\mathcal{A}')$.
To prove that $\mathcal{B} \supset f^{-1}(\mathcal{A}')$, we let

$$\mathcal{B}' = \{B' \subset X' : f^{-1}(B') \in \mathcal{B}\}$$

and prove that $\mathcal{B}'$ *is a σ-algebra.*

(i) $f^{-1}(X') = X \in \mathcal{B}$; hence $X' \in \mathcal{B}'$.
(ii) Let $B' \in \mathcal{B}'$; then $f^{-1}(X' - B') = X - f^{-1}(B') \in \mathcal{B}$ since $\mathcal{B}$ is a σ-algebra .
(iii) Let $B'_n \in \mathcal{B}'$; then $f^{-1}(\cup_n B'_n) = \cup_n f^{-1}(B'_n) \in \mathcal{B}$.

$\mathcal{B}' \supset \mathcal{G}'$; hence $\mathcal{B}'$ contains $\mathcal{A}'$, the σ-algebra generated by $\mathcal{G}'$. Let $A' \in \mathcal{A}'$. Then $A' \in \mathcal{B}'$ since $\mathcal{B}' \supset \mathcal{A}'$. Hence $f^{-1}(A') \in \mathcal{B}$. □

2.3 Measurable spaces and measurable mappings

2.3.1 **Definition.** The pair $(X, \mathcal{A})$ consisting of a set X together with a σ-algebra $\mathcal{A}$ of subsets of X is called a *measurable space.*

2.3.2 **Definition.** Given two measurable spaces $(X, \mathcal{A})$ and $(X', \mathcal{A}')$, a *mapping* f of X to X' is called measurable if $f^{-1}(\mathcal{A}') \subset \mathcal{A}$.

$\mathcal{M}((X, \mathcal{A}); (X', \mathcal{A}'))$ will denote the set of measurable mappings of $(X, \mathcal{A})$ into $(X', \mathcal{A}')$.

2.3.3 Proposition. *The composition of measurable mappings is measurable.*

PROOF. Let $f_1 \in \mathcal{M}((X, \mathcal{A}); (X', \mathcal{A}'))$, $f_2 \in \mathcal{M}((X', \mathcal{A}'); (X'', \mathcal{A}''))$. Then by 2.1.4 $f = f_2 \circ f_1$ satisfies $(f_2 \circ f_1)^{-1}(\mathcal{A}'') = f_1^{-1}(f_2^{-1}(\mathcal{A}'')) \subset f_1^{-1}(\mathcal{A}') \subset \mathcal{A}$, and hence $f_2 \circ f_1$ is measurable. □

2.3.4 Proposition (Measurability criterion). *Let $(X, \mathcal{A})$ and $(X', \mathcal{A}')$ be measurable spaces, let $\mathcal{A}'$ be the σ-algebra generated by $\mathcal{G}$, and let $\mathcal{G}' \subset \mathcal{A}'$. Then the following are equivalent:*

(i) $f \in \mathcal{M}((X, \mathcal{A}); (X', \mathcal{A}'))$
(ii) $f^{-1}(\mathcal{G}') \subset \mathcal{A}$.

PROOF. Let $\mathcal{A}_1$ be the σ-algebra generated by $f^{-1}(\mathcal{G}')$. Then (ii) is equivalent to $\mathcal{A}_1 \subset \mathcal{A}$. Furthermore, $\mathcal{A}_1 = f^{-1}(\mathcal{A}')$ by Theorem 2.2.1; hence (ii) is equivalent to (i). □

2.3.5 Measurable mappings into a product

Let $(X, \mathcal{A})$, $(Y_1, \mathcal{B}_1)$, and $(Y_2, \mathcal{B}_2)$ be measurable spaces. Let $Y_1 \times Y_2$ be given the product σ-algebra $\mathcal{B}_1 \otimes \mathcal{B}_2$, defined in 1.5.2, and let π_i (i = 1,2) be the natural projection of $Y_1 \times Y_2$ onto Y_i.

Lemma. $\pi_1 \in \mathcal{M}((Y_1 \times Y_2, \mathcal{B}_1 \otimes \mathcal{B}_2); (Y_1, \mathcal{B}_1))$.

PROOF. We must consider $\pi_1^{-1}(B_1)$, where $B_1 \in \mathcal{B}_1$. But $\pi_1^{-1}(B_1) = B_1 \times Y_2$ is a rectangle, and hence an element of $\mathcal{B}_1 \otimes \mathcal{B}_2$. □

Proposition (Measurability criterion for a mapping into a product). *Let f be a mapping of X into $Y_1 \times Y_2$. Then f is measurable if and only if its components $f_i = \pi_i \circ f$ (i=1,2) are measurable.*

PROOF. Suppose that f is measurable. Then, by the preceding lemma, $\pi_1 \circ f$ is a composition of measurable mappings and hence measurable. Conversely, suppose that f_1 and f_2 are measurable and let $R = B_1 \times B_2$ be a rectangle. Then $f^{-1}(R) = f_1^{-1}(B_1) \cap f_2^{-1}(B_2)$. Each $f_i^{-1}(B_i)$ is in $\mathcal{A}$, hence so is their intersection, and the measurability criterion 2.3.4 then shows that f is measurable. □

2.4 Borel algebras. Measurability and continuity. Operations on measurable functions

2.4.1 Separability and measurability

Separability of topological spaces

Let Y be a Hausdorff space.

(i) Y satisfies the *first separability axiom* if there exists a subset D of Y which is countable and dense in Y (closure of $D = Y$).

(ii) Y satisfies the *second separability axiom* if there exists a countable family of open subsets H_i such that every open set in Y may be written as a union of the H_i that it contains. The family H_i is called a *basis* of open sets for Y.

(iii) EXAMPLE. Let $Y = \mathbf{R}$ and let $\mathbf{Q}$ be the set of rational numbers. Setting $H_{q_1, q_2} = (q_1, q_2)$, we obtain a countable family of intervals. Then every interval

(x_1, x_2) can be written as a union of the H_i that it contains. The same holds for any open set.

(iv) Proposition. *Let Y be a metric space satisfying the first separability axiom. Then it satisfies the second.*

PROOF. Let $\{y_i\}$ be a dense sequence in Y. We denote by d the distance on Y and set $H_{i,m} = \{y \in Y : d(y, y_i) < m^{-1}\}$, where $m \in \mathbf{N}$. For each open set O in Y, let O' be the union of the H_{im} contained in O. Then O' is an open subset of O. Let $z \in O$. Then there exists m_0 such that the ball with center z and radius m_0^{-1} is contained in O. Let j be such that $d(y_j, z) < (2m_0)^{-1}$. Then $z \in H_{j,2m_0}$, and hence $O \subset O'$.

(v) The space $\mathbf{R}^n$ satisfies the first separability axiom and hence the second.

(vi) The second separability axiom implies the first. It suffices to choose a point y in each H_i to obtain a dense sequence.

Because of (vi) and (iv), we refer to a metric space which has a dense sequence as a *separable metric space.*

(vii) Let Y, Y' be two separable metric spaces. Then their product Y'' is separable. Set $y''_{j,k} = (y_j, y'_k)$; then the $\{y''_{j,k}\}$ form a countable dense subset of Y''.

(viii) Proposition (Measurability criterion). *Suppose that $(X, \mathcal{A})$ is a measurable space, Y is a topological space satisfying the second separability axiom, and H_i is a basis of open sets of Y. Then a mapping $f : X \to Y$ is measurable if and only if*

$$f^{-1}(H_i) \in \mathcal{A}, \quad i \in \mathbf{N}.$$

PROOF. This follows immediately from the measurability criterion 2.3.4. It must be shown that, for every open set O, $f^{-1}(O) \in \mathcal{A}$. Let $O = \cup_s H_{i_s}$; then $f^{-1}(O) = \cup_s f^{-1}(H_{i_s}) \in \mathcal{A}$. □

REMARK. (viii) provides an explicit criterion for the measurability of a function.

2.4.2 Product of Borel algebras

Proposition. *Consider two separable metric spaces X_1 and X_2 and their product $Y = X_1 \times X_2$. Let Y be equipped with the product topology. Denote by $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_Y$ the associated Borel algebras. Then $\mathcal{B}_Y = \mathcal{B}_1 \otimes \mathcal{B}_2$.*

PROOF. Y is separable by 2.4.1. The family of open sets of the product topology is generated by the countable unions of open rectangles: $R_0 = O_1 \times O_2$, where $O_i \in \mathcal{O}_{X_i}$. Hence $R_0 \in \mathcal{B}_1 \otimes \mathcal{B}_2$; that is, $\mathcal{O}_Y \subset \mathcal{B}_1 \otimes \mathcal{B}_2$. It follows that $\mathcal{B}_Y \subset \mathcal{B}_1 \otimes \mathcal{B}_2$.

Let π_1 be the projection of Y onto X_1. Then $\pi_1 \in \mathcal{M}((Y, \mathcal{B}_Y); (X_1, \mathcal{B}_1))$ since $\pi_1^{-1}(\mathcal{O}_{X_1}) \subset \mathcal{B}_Y$.

It follows from 2.3.5 and the fact that $\pi_i \circ j$ $(j = 1, 2)$ is measurable that the identity mapping $j : (Y, \mathcal{B}_Y) \mapsto (Y, \mathcal{B}_1 \otimes \mathcal{B}_2)$ is measurable. Thus $j^{-1}(\mathcal{B}_1 \otimes \mathcal{B}_2) \subset \mathcal{B}_Y$, or $\mathcal{B}_1 \otimes \mathcal{B}_2 \subset \mathcal{B}_Y$. $\square$

2.4.3 Measurability and continuity

Let X and X' be topological spaces. Equipping them with their Borel algebras $\mathcal{B}_X$ and $\mathcal{B}_{X'}$, we obtain measure spaces $(X, \mathcal{B}_X)$ and $(X', \mathcal{B}_{X'})$.

Proposition. *Every continuous mapping f from X to X' is a measurable mapping from $(X, \mathcal{B}_X)$ to $(X', \mathcal{B}_{X'})$.*

PROOF. We use the measurability criterion 2.3.4. It must be shown that $f^{-1}(\mathcal{O}_{X'}) \subset \mathcal{B}_X$. But since f is continuous, the inverse image of an open set is open, whence $f^{-1}(\mathcal{O}_{X'}) \subset \mathcal{O}_X$. Since $\mathcal{O}_X \subset \mathcal{B}_X$, the conclusion follows. $\square$

2.4.4 Algebraic operations on measurable functions

Consider the field of real numbers $\mathbf{R}$ with its Borel algebra $\mathcal{B}_\mathbf{R}$. Given a measurable space $(X, \mathcal{A})$, we denote by $\mathcal{L}^0(X, \mathcal{A})$ the set of *measurable mappings* from $(X, \mathcal{A})$ to $(\mathbf{R}, \mathcal{B}_\mathbf{R})$. Elements of $\mathcal{L}^0(X, \mathcal{A})$ are called *measurable functions.* When X is a topological space with its Borel algebra $\mathcal{B}_X$, elements of $\mathcal{L}^0(X, \mathcal{B}_X)$ are often called *Borel functions.*

Proposition. *The absolute value of a measurable function f is measurable. The sum and product of two measurable functions are measurable. The multiplicative inverse of a measurable function which is everywhere nonzero is measurable.*

PROOF. Let u be the mapping from $\mathbf{R}$ to $\mathbf{R}$ defined by the absolute value: $u(\zeta) = |\zeta|$. Then u is continuous, hence measurable, and 2.3.3 implies that $|f| = u \circ f$ is measurable.

Let Φ be the continuous mapping of $\mathbf{R}^2 \to \mathbf{R}$ defined by $\Phi(\zeta_1, \zeta_2) = \zeta_1 + \zeta_2$. Similarly, let $\Psi(\zeta_1, \zeta_2) = \zeta_1 \zeta_2$.

Let f_1 and f_2 be measurable functions on X, and let $F(x) = (f_1(x), f_2(x))$. Then $F : X \to \mathbf{R}^2$ and, by 2.3.5,

$$F \in \mathcal{M}((X, \mathcal{A}); (\mathbf{R}^2, \mathcal{B}_\mathbf{R} \otimes \mathcal{B}_\mathbf{R})).$$

By 2.4.2, $\mathcal{B}_\mathbf{R} \otimes \mathcal{B}_\mathbf{R} = \mathcal{B}_{\mathbf{R}^2}$; hence $F \in \mathcal{M}((X, \mathcal{A}), (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}))$. Since Φ is continuous, $\Phi \in \mathcal{M}((\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}); (\mathbf{R}, \mathcal{B}_\mathbf{R}))$. Thus, by 2.3.3,

$$\Phi \circ F \in \mathcal{M}((X, \mathcal{A}); (\mathbf{R}, \mathcal{B}_\mathbf{R})) = \mathcal{L}^0(X, \mathcal{A}).$$

But $(\Phi \circ F)(x) = f_1(x) + f_2(x)$.

Similarly, $\Psi \circ F \in \mathcal{L}^0((X, \mathcal{A}))$ and $(\Psi \circ F)(x) = f_1(x) f_2(x)$.

We denote $\mathbf{R} - \{0\}$ by $\mathbf{R}'$. Let η be the continuous mapping of $\mathbf{R}' \to \mathbf{R}'$ defined by $\eta(\zeta) = \frac{1}{\zeta}$ and let $f \in \mathcal{L}^0(X, \mathcal{A})$, $f(x) \neq 0$ for all $x \in X$. If O is an open set in $\mathbf{R}$, then $O' = O \cap \mathbf{R}'$ is an open set in $\mathbf{R}'$. Set $g(x) = \frac{1}{f(x)}$. Then $g^{-1}(O) = g^{-1}(O') = f^{-1}(\eta^{-1}(O'))$. Since $\eta^{-1}(O')$ is an open set in $\mathbf{R}'$ and $\mathbf{R}'$ is open in $\mathbf{R}$, $\eta^{-1}(O')$ is open in $\mathbf{R}$. Since f is measurable, $f^{-1}(\eta^{-1}(O')) \in \mathcal{A}$. □

2.5 Pointwise convergence of measurable mappings

In this section, $(X, \mathcal{A})$ denotes a measurable space, Y a metric space, and $\mathcal{B}_Y$ the Borel algebra of Y. We say that a sequence of mappings $f_n : X \mapsto Y$ *converges pointwise* to f_0 if $\lim f_n(x) = f_0(x)$ for every $x \in X$.

2.5.1 Theorem. *Let f_n be a sequence of measurable mappings which converge pointwise to f_0. Then f_0 is measurable.*

REMARK. It is well known that the pointwise limit of a sequence of *continuous* functions is not necessarily continuous. This theorem shows the great *stability* of the property of measurability.

PROOF. Let $f_n \in \mathcal{M}((X, \mathcal{A}); (Y, \mathcal{B}))$. Let d denote the distance in Y and let O be an open set in Y. For every $k > 0$, let

$$O_k = \left\{ x \in O : d(x, O^c) > \frac{1}{k} \right\}.$$

Then O_k is an increasing sequence of open sets in O and $O = \cup_{k \in \mathbf{N}} O_k$. Moreover, denoting by $\overline{O_k}$ the closure of O_k, we have $\overline{O_k} \subset O_{k+1}$.

Since $d(f_0(x), f_m(x)) \to 0$, it follows that

$$f_0(x) \in O_k \Rightarrow f_q(x) \in O_k \quad \text{if } q \text{ is large enough, say } q \geq m_0.$$

Set $H^k_{m_0} = \cap_{q \geq m_0} f_q^{-1}(O_k)$. Since f_q is measurable, each $f_q^{-1}(O_k) \in \mathcal{A}$, whence $H^k_{m_0} \in \mathcal{A}$. Let $G^k = \cup_{m_0} H^k_{m_0}$; then $G^k \in \mathcal{A}$.

We have thus shown that $f_0(x) \in O_k \Rightarrow x \in G^{k+1}$ or, taking the union over k, $f_0(x) \in O \Rightarrow x \in \bigcup_{r \in \mathbf{N}} G^r$, which may be written as

$$(i) \qquad f_0^{-1}(O) \subset W, \quad \text{where} \quad W = \cup_{r \in \mathbf{N}} G^r \in \mathcal{A}.$$

We now prove the reverse inclusion. Let $x_1 \in G^r$. Then there exists m_1 such that $x_1 \in H^r_{m_1}$, or $x_1 \in f_q^{-1}(O_r)$ if $q > m_1$. Thus $\lim f_q(x_1) \in \overline{O_r} \subset O_{r+1} \subset O$ and therefore

$$(ii) \qquad f_0^{-1}(O) \supset W.$$

From (i) and (ii) it follows that $f_0^{-1}(O) = W$, or $W \in \mathcal{A}$, whence f_0 is measurable. □

For emphasis, we restate (i) and (ii) in the following form.

2.5.2 Fundamental lemma. *Let $\{f_n\}$ be a sequence of mappings from X to the metric space Y that converges pointwise to f_0. Then for every open set O in Y,*

$$f_0^{-1}(O) = \bigcup_{r,m} \left[\bigcap_{q \geq m} f_q^{-1}(O_r) \right], \quad \textit{where} \quad O_r = \left\{ x \in O : d(x, O^c) > \frac{1}{r} \right\}.$$

2.6 Supremum of a sequence of measurable functions

For convenience of notation, we introduce in this section the set $\overline{\mathbf{R}}$ of real numbers completed by adjoining the two elements $+\infty$ and $-\infty$.

Addition and multiplication in $\overline{\mathbf{R}}$ are defined in the elementary way, except for the "indefinite forms" $+\infty - -\infty$ and $0 \cdot \infty$.

$\overline{\mathbf{R}}$ is given the obvious order relation, with $+\infty$ the largest and $-\infty$ the smallest element. A distance is defined on $\overline{\mathbf{R}}$ by setting

$$d(x, x') = |\text{Arctan}\, x - \text{Arctan}\, x'|.$$

Every subset of $\overline{\mathbf{R}}$ has a supremum, or least upper bound. The empty set is assigned the supremum $-\infty$.

2.6.1 Proposition. *Let $\{f_n\}$ be a sequence in $\mathcal{M}((X, \mathcal{A}); (\overline{\mathbf{R}}, \mathcal{B}_{\overline{\mathbf{R}}}))$ and let $\varphi = \sup f_n$. Then $\varphi \in \mathcal{M}((X, \mathcal{A}); (\overline{\mathbf{R}}, \mathcal{B}_{\overline{\mathbf{R}}}))$.*

PROOF. Since $\{+\infty\}$ is a closed subset of $\overline{\mathbf{R}}$, $f_n^{-1}(\{+\infty\}) \in \mathcal{A}$. Set $G = \cup_n f_n^{-1}(\{+\infty\})$. Then $G \in \mathcal{A}$ and $\varphi(x) = +\infty$ if $x \in G$.

Let $X' = G^c$, equip X' with the trace $\mathcal{A}'$ of the σ-algebra $\mathcal{A}$, and denote by f_n' the restriction of f_n to X. Then

$$f_n' \in \mathcal{M}((X', \mathcal{A}'); (\mathbf{R}, \mathcal{B}_{\mathbf{R}})) = \mathcal{L}^0(X', \mathcal{A}').$$

Moreover, by 2.4.4, $\sup(f_1', f_2') \in \mathcal{L}^0(X', \mathcal{A}')$.

More generally, let the sequence $\{g_n\}$ be defined by recursion: $g_1 = f_1'$ and $g_k = \sup(f_k', g_{k-1})$ if $k > 1$.

An induction argument shows that $g_{k+1} \in \mathcal{L}^0(X', \mathcal{A}')$. Moreover, $g_k \leq g_{k+1}$. Thus $\{g_k\}$ is an *increasing sequence*, hence convergent in $\overline{\mathbf{R}}$. Set $\varphi_1(x') = \lim g_k(x')$, $x' \in X'$. Then, by 2.5.1, $\varphi_1 \in \mathcal{M}((X', \mathcal{A}'); (\mathbf{R}, \mathcal{B}_{\mathbf{R}}))$. Furthermore, $\varphi(x) = \varphi_1(x)$ if $x \in X'$ and $\varphi(x) = +\infty$ if $x \notin X'$.

Let K be a closed subset of $\mathbf{R}$. Then

$$\begin{aligned} \varphi^{-1}(K) &= \varphi_1^{-1}(K) & &\text{if} \quad +\infty \notin K \\ \varphi^{-1}(K) &= \varphi_1^{-1}(K) \cup G & &\text{if} \quad +\infty \in K. \end{aligned}$$

Since $\varphi_1^{-1}(K) = X' \cap A$ with $A \in \mathcal{A}$ and $X' \in \mathcal{A}$, it follows that $\varphi_1^{-1}(K) \in \mathcal{A}$. $\square$

2.6.2 Corollary. *Let $f_n \in \mathcal{M}((X, \mathcal{A}); (\overline{\mathbf{R}}, \mathcal{B}_{\overline{\mathbf{R}}}))$. Then* $(\limsup f_n) \in \mathcal{M}((X, \mathcal{A}); (\overline{\mathbf{R}}, \mathcal{B}_{\overline{\mathbf{R}}}))$.

PROOF. Let $\varphi_n = \sup_{p \geq n} f_p$. Then φ_n is measurable. The sequence $\{\varphi_n(x)\}$ is decreasing, hence convergent in $\overline{\mathbf{R}}$, and 2.5.1 gives the result. □

3 Measures and Measure Spaces

Definition. Let $\overline{\mathbf{R}}^+ = \{\zeta \in \mathbf{R} : \zeta \geq 0\} \cup \{+\infty\}$. Given a measurable space $(X, \mathcal{A})$, a *measure* on $(X, \mathcal{A})$ is a mapping $\mu : \mathcal{A} \to \overline{\mathbf{R}}^+$ satisfying the following two axioms:

Countable additivity (σ-additivity) axiom

3.0.1. *Let $A_k \in \mathcal{A}$, $k \in I$, be a finite or countable family of measurable sets that are pairwise disjoint; that is, $A_k \cap A_l = \emptyset$ if $k \neq l$. Then*

(i)
$$\mu\left(\bigcup_{k \in I} A_k\right) = \sum_{k \in I} \mu(A_k).$$

In particular,

(ii) $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$ *if* $A_1 \cap A_2 = \emptyset$ *(finite additivity).*

σ-finiteness axiom

There exist $A_n \in \mathcal{A}$ such that

3.0.2
$$X = \cup_n A_n \quad \textit{and} \quad \mu(A_n) < +\infty \quad \forall n.$$

The sequence $\{A_n\}$ is called an *exhaustion sequence* for X. If $\mu(X) < +\infty$, X is said to have *finite* measure (or finite total mass) and μ itself is called a finite measure. It is possible to develop part of the theory without using 3.0.2, the σ-finiteness axiom. However, the axiom will always be satisfied for the applications we have in mind, and we take this point of view for ease of exposition.

Definition. A measurable space $(X, \mathcal{A})$ equipped with a measure μ defined on $\mathcal{A}$ is called a *measure space* and is denoted by $(X, \mathcal{A}, \mu)$.

EXAMPLE. Let $\{x_i\}$ be a countable sequence of points of X and let $\{\alpha_i\}$ be a sequence of positive real numbers. For $\mathcal{A} = \mathcal{P}(X)$ and $A \in \mathcal{A}$, set

$$\mu(A) = \sum_{x_i \in A} \alpha_i.$$

Then $(X, \mathcal{A}, \mu)$ is a measure space. If $\alpha_i = 1$, $i \in \mathbf{N}$, this measure μ is called the *counting measure* associated with the sequence $\{x_i\}$; $\mu(A)$ equals the number of points of the sequence $\{x_i\}$ which lie in A.

This example is trivial and does not reveal the complexity of the theory. In fact, we will not obtain nontrivial examples of measure spaces until Chapter II.

3.1 Convexity inequality

Proposition. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Then*

3.1.1 *μ is increasing; that is, if A_1 and $A_2 \in \mathcal{A}$ and $A_1 \subset A_2$, then $\mu(A_1) \leq \mu(A_2)$.*

3.1.2 *μ is convex; that is, if $B_1, \ldots, B_n \in \mathcal{A}$ (not necessarily disjoint), then*

$$\mu\left(\bigcup_{i=1}^{n} B_i\right) \leq \sum_{i=1}^{n} \mu(B_i).$$

PROOF. Let $A_1 \subset A_2$ and let $B = A_1^c \cap A_2$; then $B \in \mathcal{A}$ and $A_2 = A_1 \cup B$. The finite additivity axiom gives

$$\mu(A_2) = \mu(A_1) + \mu(B).$$

Since $\mu(B) \geq 0$, we conclude that $\mu(A_2) \geq \mu(A_1)$.

Similarly, let the sequence $\widetilde{B}_1, \ldots, \widetilde{B}_q, \ldots$ be defined recursively:

$$\widetilde{B}_1 = B_1 \quad \text{and} \quad \widetilde{B}_q = B_q \cap \left(\cup_{j<q} B_j\right)^c, \ q > 1.$$

Then $\widetilde{B}_q \in \mathcal{A}$, $\cup_{j=1}^{m} B_j = \cup_{j=1}^{m} \widetilde{B}_j$, and by finite additivity

$$\mu\left(\cup_{j=1}^{m} \widetilde{B}_j\right) = \sum_{j=1}^{m} \mu(\widetilde{B}_j).$$

$\widetilde{B}_j \subset B_j$ implies $\mu(\widetilde{B}_j) \leq \mu(B_j)$, and the desired inequality follows. □

3.2 Measure of limits of monotone sequences

Theorem. *Let A_1, A_2, …, A_n, … be an increasing sequence of measurable sets. Let*

$$\lim \uparrow A_i = \bigcup_{i=1}^{+\infty} A_j.$$

Then

3.2.1 $$\mu(\lim \uparrow A_i) = \lim \mu(A_i).$$

Theorem. *Let B_1, B_2, …, B_n, … be a decreasing sequence of measurable sets. Let*

$$\lim \downarrow B_i = \bigcap_{i=1}^{+\infty} B_i.$$

3.2.2 *Suppose that there exists k_0 such that $\mu(B_{k_0}) < +\infty$. Then*

3.2.3 $$\mu(\lim \downarrow B_k) = \lim \mu(B_k).$$

Remark. The properties described by these two theorems are sometimes called *continuity on increasing sequences* and *continuity on decreasing sequences.*

Proof. Consider the measure space $(X, \mathcal{A}, \mu)$. For $A_n \in \mathcal{A}$, set $\widetilde{A}_1 = A_1$ and $\widetilde{A}_{n+1} = A_n^c \cap A_{n+1}$ if $n > 1$.

Then $\widetilde{A}_{n+1} \in \mathcal{A}$, the $\widetilde{A}_n$ are disjoint, and $A_j = \cup_{q \leq j} \widetilde{A}_q$. Hence, by finite additivity,

$$\mu(A_j) = \sum_{q \leq j} \mu(\widetilde{A}_q).$$

Moreover, $\cup_{j=1}^{\infty} A_j = \cup_{j=1}^{\infty} \widetilde{A}_j$ and, by σ-additivity,

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{+\infty} \mu(\widetilde{A}_j).$$

Hence, for increasing sequences, 3.2.1 reduces to the simple observation that the sum of a series of nonnegative terms is the limit of its partial sums; that is, $\sum_{j=1}^{+\infty} \mu(\widetilde{A}_j) = \lim_r \sum_{q=1}^{r} \mu(\widetilde{A}_r)$. This limit always exists, whether it is finite or infinite.

In order to prove 3.2.3, we set $A'_k = B_{k_0} \cap B_k^c$, $k > k_0$. Then A'_k is an increasing sequence. The relation $B_{k_0} = B_k \cup A'_k$, B_k and A'_k disjoint, implies $\mu(B_{k_0}) = \mu(B_k) + \mu(A'_k)$. Hence $\mu(A'_k) \leq \mu(B_{k_0})$ and $\mu(\lim \uparrow A'_k) = \lim \mu(A'_k) = \beta \leq \mu(B_{k_0})$. We have

$$(\lim \downarrow B_k)^c \cup (\lim \uparrow A'_k) = B_{k_0},$$

whence

$$\mu(\lim \downarrow B_k) + \mu(\lim \uparrow A'_k) = \mu(B_{k_0}),$$

or finally

$$\mu(\lim \downarrow B_k) = \mu(B_{k_0}) - \lim \mu(A'_k) = \lim_k [\mu(B_{k_0}) - \mu(A'_k)] = \lim \mu(B_k). \square$$

3.2.4 Application — Exhaustion principle

We now roughly sketch a principle that will often be used. Let (P) be a *property* that is true for all finite measures. Let $(X, \mathcal{A}, \mu)$ be a measure space with an exhaustion sequence A_n. Let $X_n = A_n$, equipped with the trace σ-algebra $\mathcal{A}_n$ of the σ-algebra $\mathcal{A}$, and let μ_n be the restriction of μ to $\mathcal{A}_n$. Then each μ_n is finite and therefore satisfies (P).

To conclude that μ satisfies (P), it suffices to show that "the limits of values of μ_n appearing in (P) are finite".

3.2.5 REMARK. Let σ be a mapping from $\mathcal{A}$ to $\mathbf{R}^+$ satisfying the finite additivity axiom 3.0.1(ii) and property 3.2.1 of continuity with respect to increasing sequences. Then σ satisfies 3.0.1(i), since

$$\sigma(\cup_1^\infty A_p) = \sigma(\lim(\cup_1^n A_p)) = \lim \sigma(\cup_1^n A_p) = \lim \sum_1^n \sigma(A_p) = \sum_1^{+\infty} \sigma(A_p),$$

where the third equality follows from finite additivity.

3.3 Countable convexity inequality

Proposition. *Let $\{A_n\}$ be a sequence of (not necessarily disjoint) elements of $\mathcal{A}$. Then*

$$\mu\left(\bigcup_{n=1}^{+\infty} A_n\right) \leq \sum_{n=1}^{+\infty} \mu(A_n).$$

PROOF. Set $B_q = \bigcup_{n=1}^q A_n$. Then B_q is an increasing sequence, and by 3.2.1 we have

$$\mu\left(\bigcup_{n=1}^{+\infty} A_n\right) = \lim \mu(B_q).$$

Furthermore, by the finite convexity property 3.1.2,

$$\mu(B_q) \leq \sum_{n=1}^{q} \mu(A_n) \leq \sum_{n=1}^{+\infty} \mu(A_n).\square$$

4 Negligible Sets and Classes of Measurable Mappings

The concept of measurable mappings is extremely easy to work with. In particular, the theorem that a *pointwise limit* of measurable mappings is measurable makes the operations of analysis very convenient. The drawback of this convenience is that the space of measurable functions is enormous, and therefore hardly usable. We will work on a quotient space.

4.1 Negligible sets

Definition. Let $(X, \mathcal{A}, \mu)$ be a measure space. A subset Z of X is called *negligible* if there exists $A \in \mathcal{A}$ such that $\mu(A) = 0$ and $A \supset Z$.

4.1.1 Proposition. *A countable union of negligible sets is negligible.*

PROOF. This follows from countable convexity:

$$\mu\left(\bigcup_i A_i\right) \leq \sum_i \mu(A_i).$$

Since every term on the right-hand side is zero, the sum of the series is zero. □

Definition. A property (P) is said to be *true μ-almost everywhere* (denoted μ-a.e.) on the measure space $(X, \mathcal{A}, \mu)$ if

$$\{x : (\text{P}) \text{ does not hold at } x\} \text{ is contained in a negligible set.}$$

4.1.2 Let (P_1) be a proposition implying the proposition (P_2). Then (P_1) true μ-a.e. $\Rightarrow$ (P_2) true μ-a.e.

4.1.3 Theorem. *Let (P_1), ..., (P_n), ... be a sequence of properties defined on $(X, \mathcal{A}, \mu)$. Suppose that each of the properties (P_i) is true μ-a.e. Then their conjunction is true μ-a.e.*

PROOF. Let A_i be a negligible set that contains $\{x : (P_i) \text{ does not hold at } x\}$. Then $A_\infty = \cup_i A_i$ is negligible. If $x \notin A_\infty$, then all the (P_i) hold at x. □

4.2 Complete measure spaces

4.2.1 **Definition.** Given the measure space $(X, \mathcal{A}, \mu)$, the σ-algebra $\mathcal{A}$ is called μ-complete if *every subset of a negligible set is measurable.*

The measure space $(X, \mathcal{A}, \mu)$ is called complete when $\mathcal{A}$ is μ-complete.

The space is complete if and only if every subset of a negligible set is negligible.

On a complete measure space, a property P is true μ-a.e. if the set $\{x : (\text{P}) \text{ does not hold at } x\}$ is negligible.

4.2.2 Completion theorem. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Then there exist a σ-algebra $\mathcal{A}' \supset \mathcal{A}$ and an extension μ' of μ to $\mathcal{A}'$ such that $(X, \mathcal{A}', \mu')$ is complete and, for all $A' \in \mathcal{A}'$, there exist A_1, $A_2 \in \mathcal{A}'$ with $A_1 \subset A' \subset A_2$, $\mu(A_2 - A_1) = 0$. This σ-algebra $\mathcal{A}'$ is unique and will be called the completion of $\mathcal{A}$.*

PROOF. Define

$$\mathcal{A}' = \{Z \in \mathcal{P}(X) : \exists A_1, A_2 \in \mathcal{A} \text{ such that } A_1 \subset Z \subset A_2 \text{ and } \mu(A_2 - A_1) = 0\}.$$

Clearly $\mathcal{A}' \supset \mathcal{A}$. We show that $\mathcal{A}'$ is a σ-algebra. If $Z \in \mathcal{A}'$, then $A_2^c \subset Z^c \subset A_1^c$ and $A_1^c - A_2^c = A_2 - A_1$, whence $Z^c \in \mathcal{A}'$. Hence Axiom 1.0.2 is satisfied.

Let $Z^n \in \mathcal{A}'$. Then there exist A_1^n and A_2^n such that $A_1^n \subset Z^n \subset A_2^n$. Set $Z^\infty = \cup Z^n$, $A_1^\infty = \cup A_1^n$, and $A_2^\infty = \cup A_2^n$. Then

$$A_1^\infty \subset Z^\infty \subset A_2^\infty \quad \text{and} \quad A_2^\infty - A_1^\infty \subset \bigcup_n (A_2^n - A_1^n).$$

The right-hand side, as a countable union of negligible sets, is negligible, whence $Z \in \mathcal{A}'$ and $\mathcal{A}'$ is a σ-algebra.

To extend μ to $\mathcal{A}'$, we first note that $\mu(A_2) = \mu(A_1) + \mu(A_2 - A_1) = \mu(A_1)$. For $Z \in \mathcal{A}'$, let $\mu'(Z)$ be defined by $\mu'(Z) = \mu(A_1)$.

We now show that this is *independent* of the choice of $A_1 \subset Z \subset A_2$. Let $\tilde{A}_1 \subset Z \subset \tilde{A}_2$; $\tilde{A}_1, \tilde{A}_2 \in \mathcal{A}$, $\mu(\tilde{A}_2 - \tilde{A}_1) = 0$. Then $\tilde{A}_2 \supset Z \supset A_1$, whence $\mu(\tilde{A}_2) \geq \mu(A_1) = \mu(A_2)$. Similarly $\mu(A_2) \geq \mu(\tilde{A}_2)$, whence $\mu(\tilde{A}_2) = \mu(A_2)$. Moreover, if Z^n is a sequence of disjoint sets, then so is A_1^n; hence $\mu(\cup A_1^n) = \sum \mu(A_1^n)$, and we have shown that μ' is countably additive.

Finally, μ' is complete: letting $Z \in \mathcal{A}'$ with $\mu'(Z) = 0$, there exists $A_2 \in \mathcal{A}$ satisfying $Z \subset A_2$ and $\mu(A_2) = 0$. Let $Z_1 \subset Z$. Then $\emptyset \subset Z_1 \subset A_2$, where $\emptyset$, $A_2 \in \mathcal{A}$ and $\mu(A_2 - \emptyset) = 0$. Therefore $Z_1 \in \mathcal{A}'$. □

4.3 The space $M_\mu((X, \mathcal{A}); (X', \mathcal{A}'))$

(i) On $\mathcal{M}((X, \mathcal{A}); (X', \mathcal{A}'))$, let the equivalence relation be defined by

$$f \sim f' \quad \text{if} \quad f(x) = f'(x) \quad \mu\text{-a.e.}$$

The equivalence class of f is denoted by $\overline{f}$.

(ii) The transitivity of this relation follows from 4.1.3.

4.3.1 **Definition.** The quotient of $\mathcal{M}$ by this equivalence relation is denoted by $M_\mu((X, \mathcal{A}); (X', \mathcal{A}'))$.

An element $\overline{f} \in M_\mu$ is a mapping $f : X \to X'$, defined "up to a set of μ-measure zero".

4.3.2 Let $\mathcal{E}$ be a negligible set and let $\varphi : X - \mathcal{E} \to X'$.

Suppose that φ is a measurable mapping when $X - \mathcal{E}$ is given the trace σ-algebra induced by $\mathcal{A}$. Define $f : X \to X'$ by setting

$$\begin{aligned} f(x) &= \varphi(x) \qquad && \text{if} \qquad x \in X - \mathcal{E} \\ f(x) &= x'_0 && \text{if} \qquad x \in \mathcal{E}, \end{aligned}$$

where x'_0 is an arbitrarily chosen element in X'.

Then $f \in \mathcal{M}((X, \mathcal{A}); (X', \mathcal{A}'))$, and φ determines the equivalence class of f in $M_\mu((X, \mathcal{A}); (X', \mathcal{A}'))$.

4.3.3 Remark. When $X' = \mathbf{R}$ and $\mathcal{A} = \mathcal{B}_{\mathbf{R}}$, the operations defined on measurable functions (sum, product, sup) are compatible with the equivalence relation. The quotient of $\mathcal{L}^0(X, \mathcal{A}) = \mathcal{M}((X, \mathcal{A}); (\mathbf{R}, \mathcal{B}_{\mathbf{R}}))$ is denoted by $L^0_\mu(X, \mathcal{A})$.

Thus the operations sum, product, and sup are defined on $L^0_\mu(X, \mathcal{A})$. Moreover, any element of $L^0_\mu(X, \mathcal{A})$ with a representative that is nonzero almost everywhere has a well-defined inverse. $L^0_\mu(X, \mathcal{A})$ is called the space of equivalence classes of measurable functions.

5 Convergence in $M_\mu((X,\mathcal{A});(Y,\mathcal{B}_Y))$

Throughout this section $(X,\mathcal{A},\mu)$ denotes a measure space, Y a separable metric space, and $\mathcal{B}_Y$ the Borel algebra of Y.

5.1 Convergence almost everywhere

5.1.1 **Definition.** Let $\overline{f}_n \in M_\mu((X,\mathcal{A});(Y,\mathcal{B}_Y))$. $\{\overline{f}_n\}$ is said to *converge almost everywhere* if, when representatives f_n of $\overline{f}_n$ are chosen, $\{f_n(x)\}$ is convergent μ-a.e.

We first show that this definition is *independent* of the choice of representatives. Let $g_n = f_n$ μ-a.e. Denote by (P_n) and (F) the following propositions:

$$(P_n) \qquad g_n(x) = f_n(x)$$
$$(F) \qquad \lim f_n(x) \text{ exists.}$$

Let (G) be the *conjunction* of (F) and the (P_n). Then, by 4.1.3, (G) is true μ-a.e. Since (G) implies the convergence of the g_n, 4.1.2 gives the result.

5.1.2 Proposition. *Let $\overline{f}_n \in M_\mu((X,\mathcal{A});(Y,\mathcal{B}_Y))$. Suppose that $\{\overline{f}_n\}$ converges almost everywhere. Then*

$$\lim f_n(x)$$

defines an element $\overline{g}_0 \in M_\mu((X,\mathcal{A});(Y,\mathcal{B}_Y))$.

PROOF. Choose an arbitrary $y_0 \in Y$, let (F) be defined as in 5.1.1, and let K be a negligible set such that K $\supset \{x : \text{(F) is not satisfied at } x\}$. Let

$$\begin{aligned} g_n(x) &= f_n(x) && x \in K \\ g_n(x) &= y_0 && x \in K^c. \end{aligned}$$

Then, by 4.3.2, $g_n \in \mathcal{M}((X,\mathcal{A}),(Y,\mathcal{B}_Y))$ and $\overline{g}_n = \overline{f}_n$.

Moreover, if $x \in K$ then $\{g_n(x)\}$ converges by 5.1.2; if $x \notin K$, then $g_n(x) = y_0$ and hence the sequence converges.

Thus $\{g_n(x)\}$ converges for all $x \in X$, and Theorem 2.5.1 shows that $g_0 = \lim g_n$ satisfies

$$g_0 \in \mathcal{M}((X,\mathcal{A});(Y,\mathcal{B}_Y)).$$

Hence

$$\lim \overline{f}_n = \overline{g_0} \in M_\mu((X,\mathcal{A});(Y,\mathcal{B}_Y)).\square$$

5.1.3 Lemma. *Given $f,g \in \mathcal{M}((X,\mathcal{A});(Y,\mathcal{B}_Y))$, let $q_{f,g}$ be defined by $q_{f,g}(x) = d(f(x),g(x))$. Then $q_{f,g}$ is a measurable function and $\forall\eta \in \mathbf{R}^+$ $\{x : q_{f,g}(x) > \eta\}$ is measurable.*

Proof. Let $Y^2 = Y \times Y$ and let ψ be the mapping from Y^2 into $\mathbf{R}^+$ defined by the distance: $\psi(y_1, y_2) = d(y_1, y_2)$.

Let $H : X \to Y^2$ be defined by $x \mapsto (f(x), g(x))$; then, by 2.3.5, $H \in \mathcal{M}((X, \mathcal{A}); (Y^2, \mathcal{B}_{Y^2}))$.

Since ψ is continuous, $\psi \circ H \in \mathcal{M}((X, \mathcal{A}); (\mathbf{R}, \mathcal{B}_\mathbf{R}))$.

Moreover, since $(\eta, +\infty)$ is an open set in $\mathbf{R}$, $q_{f,g}^{-1}((\eta, +\infty)) \in \mathcal{A}$. □

5.1.4 Egoroff's theorem. *Let $(X, \mathcal{A}, \mu)$ be a measure space and suppose in addition that $\mu(X) < +\infty$.*

Then $\overline{f}_n \in M_\mu((X, \mathcal{A}), (Y, \mathcal{B}_Y))$ converges μ-a.e. to $\overline{f}_0$ if and only if, choosing representatives f_n, f_0 of the classes $\overline{f}_n$, $\overline{f}_0$,

$$\forall \epsilon > 0 \; \exists K_\epsilon \in \mathcal{A} \quad \textit{such that} \quad \mu(K_\epsilon^c) < \epsilon$$

and $f_n(x)$ converges uniformly on K_ϵ to f_0.

Proof. Necessity is clear. Set $\epsilon = m^{-1}$, m a positive integer. Then f_n converges to f_0 on $\cup K_{m^{-1}} = G$. Since $\mu(G^c) \leq \mu(K_{m^{-1}}^c)$ for every m, $\mu(G^c) = 0$.

We now prove sufficiency. Set

$$A_{n,q} = \left\{ x : \; d(f_n(x), f_0(x)) > \frac{1}{q} \right\}.$$

Then $A_{n,q} \in \mathcal{A}$ by 5.1.3.

Let $B_{m,q} = \cup_{n \geq m} A_{n,q}$. Since $B_{m,q}$ is a decreasing sequence for fixed q, the hypothesis of convergence μ-a.e. together with the limit theorem 3.2 imply that $\mu(B_{m,q}) \to 0$ for every fixed q as $m \to +\infty$.

Fix an increasing sequence m_k such that $\mu(B_{m_k, k^{-1}}) < \epsilon \; 2^{-k}$. Set $K_\epsilon = \bigcup_{k=1}^\infty B_{m_k, k^{-1}}$. Then

$$\mu(K_\epsilon^c) < \epsilon \quad \text{and} \quad d(f_{m_j}(x), f_0(x)) < \frac{1}{k} \quad \text{if} \quad j \geq m_k, \; x \in K_\epsilon. \square$$

5.2 Convergence in measure

Convergence almost everywhere allowed us to introduce a notion of convergence of sequences in M_μ. We now define a *metric* on the space M_μ, and thus a new notion of convergence.

Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(Y, \mathcal{B}_Y)$ be a metric space equipped with its Borel measure. We denote by d the distance on Y.

5.2.1 Construction of an extended distance on $M_\mu((X, \mathcal{A}), (Y, \mathcal{B}_Y))$

Let $f, g \in \mathcal{M}$ and let $q_{f,g}$ be as defined in 5.1.3. With the pair of functions (f, g) we associate the subset of $(\mathbf{R}^+)^2$ defined by

$$K(f, g) = \{(\epsilon, \eta) \in (\mathbf{R}^+)^2 : \mu(q_{f,g}^{-1}(\eta, +\infty)) \leq \epsilon\}.$$

Set

$$e(f,g) = \inf(\epsilon + \eta) \quad \text{where} \quad (\epsilon,\eta) \in K(f,g).$$

If $K(f,g)$ is empty, we set $e(f,g) = +\infty$.

5.2.2 An equivalent extended distance

Set

$$\widetilde{e}(f,g) = 2\inf(\lambda) \quad \text{where} \quad (\lambda,\lambda) \in K(f,g).$$

Then we have

$$e(f,g) \leq \widetilde{e}(f,g) \leq 2e(f,g).$$

The first inequality is proved by writing $\widetilde{e}(f,g) = \inf(\lambda+\lambda)$, with $(\lambda,\lambda) \in K(f,g)$. Note, moreover, that if $(\epsilon,\eta) \in K(f,g)$, then $(\epsilon+\alpha,\eta+\beta) \in K(f,g)$ for any $\alpha,\beta \geq 0$. If $\epsilon > \eta$, we take $\alpha = 0$ and $\beta = \epsilon - \eta$ to obtain $\widetilde{e}(f,g) \leq 2\epsilon \leq 2e(f,g)$. The case $\epsilon < \eta$ is treated in the same way.

5.2.3 Lemma. *Let $f,g,h \in \mathcal{M}((X,\mathcal{A});(Y,\mathcal{B}_Y))$. Then*

$$\begin{array}{rcl} e(f,g) &=& e(g,f), \\ e(f,g) &=& 0 \textit{ is equivalent to } f(x) = g(x)\ \mu\textit{-a.e., and} \\ e(f,h) &\leq& e(f,g) + e(g,h). \end{array}$$

PROOF. The first statement is clear, and we prove the second. If $e(f,g) = 0$, then there exists a pair

$$(\epsilon_n,\eta_n) \in K(f,g), \quad \epsilon_n \to 0,\ \eta_n \to 0.$$

We may assume that η_n is a decreasing sequence. Then $q^{-1}((\eta_n,+\infty))$ is an increasing sequence and, by the limit theorem 3.2.1,

$$\mu(\lim\uparrow q_{f,g}^{-1}((\eta_n,+\infty))) = \lim \mu(q_{f,g}^{-1}((\eta_n,+\infty))) \leq \lim \epsilon_n = 0,$$

whence

$$\mu(\{x : d(f(x),g(x)) > 0\}) = 0, \quad \text{i.e.} \quad f(x) = g(x)\ \mu\text{-a.e.}$$

Conversely, if $f(x) = g(x)$ μ-a.e., then

$$\mu(q_{f,g}^{-1}((\eta,+\infty))) = 0 \quad \forall \eta > 0.$$

It remains to show that the triangle inequality holds. By *the triangle inequality on Y*,

$$q_{f,h}(x) \leq q_{f,g}(x) + q_{g,h}(x).$$

Let $(\epsilon_1,\eta_1) \in K(f,g)$ and $(\epsilon_2,\eta_2) \in K(g,h)$. Then $q_{f,h}(x) > \eta_1 + \eta_2$ implies that $q_{f,g}(x) > \eta_1$ or $q_{g,h}(x) > \eta_2$. Hence

$$q_{f,h}^{-1}((\eta_1+\eta_2,+\infty)) \subset q_{f,g}^{-1}((\eta_1,+\infty)) \cup q_{g,h}^{-1}((\eta_2,+\infty))$$

and, by the convexity inequality,

$$\mu(q_{f,h}^{-1}((\eta_1+\eta_2,+\infty))) \leq \epsilon_1 + \epsilon_2.$$

We have thus shown that $(\epsilon_1, \eta_1) \in K(f,g)$ and $(\epsilon_2, \eta_2) \in K(g,h)$ imply that $(\epsilon_1 + \epsilon_2, \eta_1 + \eta_2) \in K(f,h)$.

Set

$$\begin{aligned} G &= K(f,g) + K(g,h) \\ &= \{(\epsilon, \eta) \in (\mathbf{R}^+)^2 : \ \epsilon = \epsilon_1 + \epsilon_2, \ \eta = \eta_1 + \eta_2, \\ &\qquad \text{with} \quad (\epsilon_1, \eta_1) \in K(f,g), \ (\epsilon_2, \eta_2) \in K(g,h)\}. \end{aligned}$$

Then $G \subset K(f,h)$, and we obtain

$$e(f,h) = \inf_{(\epsilon,\eta)\in K} (\epsilon + \eta) \leq \inf_{(\epsilon,\eta)\in G} (\epsilon + \eta) = \inf(\epsilon_1 + \epsilon_2 + \eta_1 + \eta_2)$$

with

$$(\epsilon_1, \eta_1) \in K(f,g) \quad \text{and} \quad (\epsilon_2, \eta_2) \in K(g,h).$$

Thus

$$e(f,h) \leq \inf(\epsilon_1 + \eta_1) + \inf(\epsilon_2 + \eta_2) = e(f,g) + e(g,h).\square$$

5.2.4 Corollary. *If $f = f'$ and $g = g'$ μ-a.e., then $e(f,g) = e(f',g')$.*

Proof. Since $e(f,g) \leq e(f,f') + e(f',g') + e(g',g)$ and the hypotheses imply that the first and third terms on the right-hand side are zero, it follows that $e(f,g) \leq e(f',g')$.

The opposite inequality is proved in the same way. $\square$

Remark. $e(f,g)$ depends only on the equivalence classes $\overline{f}$ and $\overline{g}$.

Abusing notation, we set $e(\overline{f}, \overline{g}) = e(f,g)$, where f and g are chosen in the classes of $\overline{f}$ and $\overline{g}$.

5.2.5 Proposition. *Suppose that $(X, \mathcal{A}, \mu)$ is a measure space and Y is a metric space. Let $M_\mu((X,\mathcal{A}); (Y, \mathcal{B}_Y))$ be the space of equivalence classes of measurable mappings from X to Y and let e be as defined in 5.2.2. Set*

$$d_\mu(\overline{f}, \overline{g}) = \frac{e(\overline{f}, \overline{g})}{1 + e(\overline{f}, \overline{g})}.$$

Then d_μ is a distance on M_μ.

Proof. Lemma 5.2.3 shows that e satisfies the axioms for a distance, except that e may assume the value $+\infty$. We use a construction common in topology; let

$$k(t) = \frac{t}{1+t}, \quad t \in \mathbf{R}^+, \quad k(+\infty) = 1.$$

It is elementary to verify that the function $t \mapsto k(t)$ satisfies

$$k(t_1 + t_2) \leq k(t_1) + k(t_2), \quad t_1, \ t_2 \geq 0.$$

It follows that d_μ satisfies the triangle inequality and thus defines a distance on M_μ. $\square$

Remark (i). If $\mu(X) \leq C$, then it is always true that $(C, 0) \in K(f,g)$ and hence that $e(f,g) \leq C$. In this case it is unnecessary to use d_μ; e may be taken as a distance.

REMARK (ii). A sequence $f_n \in M_\mu$ is a Cauchy sequence with respect to the distance d_μ if and only if $e(f_m, f_n) \to 0$ whenever m and $n \to +\infty$.

5.2.6 **Definition.** A sequence f_n is said to converge to f_0 *in measure* if $e(f_n, f_0) \to 0$.

Proposition. *The sequence f_n converges to f_0 in measure if and only if, for every fixed $\eta > 0$,*

$$\mu(\{x:\ d(f_n(x), f_0(x)) > \eta\}) \to 0 \quad as \quad n \to +\infty.$$

PROOF. ($\Leftarrow$) Let n_0 be such that

$$\mu(\{x:\ d(f_n(x), f_0(x)) > \eta\}) < \eta \quad \text{if} \quad n \geq n_0.$$

Then $(\eta, \eta) \subset K(f_n, f_0)$, whence

$$e(f_n, f_0) < 2\eta \quad \text{if} \quad n \geq n_0.$$

($\Rightarrow$) Let $\eta_1 < \eta$ be given. Using 5.2.2, we can find n_1 such that $\widetilde{e}(f_n, f_0) < 2\eta_1$ if $n > n_1$; i.e., $(\eta_1, \eta_1) \in K(f_n, f_0)$. Hence

$$\mu(\{x:\ d(f_n, f_0) > \eta\}) < \eta_1.$$

Since $\{x : d(f_n, f_0) > \eta_1\} \subset \{x : d(f_n, f_0) > \eta\}$, it follows a fortiori that

$$\mu(\{x : d(f_n, f_0) > \eta\}) < \eta_1 \text{ if } n > n_1. \square$$

5.2.7 Theorem (Comparison of convergence in measure and convergence almost everywhere). *Suppose that $(X, \mathcal{A}, \mu)$ is a complete measure space, Y is a metric space, $f_0 \in M_\mu((X,\mathcal{A}),(Y,\mathcal{B}_Y))$, and $\{f_n\}$ is a sequence in $M_\mu((X,\mathcal{A}),(Y,\mathcal{B}_Y))$.*

(i) If $d_\mu(f_n, f_0) \to 0$, then there exists a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that $f_{n_k} \to f_0$ μ-a.e.
(ii) Suppose in addition that μ is a finite measure. Then the μ-a.e. convergence of f_n to f_0 implies that $d_\mu(f_n, f_0) \to 0$.

The proof depends on the following important lemma:

5.2.8 Lemma (Borel-Cantelli). *Let $\{A_n\}$ be a collection of elements of $\mathcal{A}$ such that*

$$\sum \mu(A_n) < +\infty.$$

Then μ-almost every x lies in at most a finite number of A_n.

PROOF. Set $B_m = \cup_{n\geq m} A_n$. Then $\mu(B_m) \leq \sum_{n\geq m} \mu(A_n)$ by the convexity inequality; hence $\lim_{m\to\infty} \mu(B_m) = 0$. But since B_n is a decreasing sequence, it follows from the continuity theorem (3.2.3) that $\mu(\cap_n B_n) = \lim \mu(B_n) = 0$. Note finally that $x \notin \cap_n B_n \Leftrightarrow x$ is in only finitely many A_n. □

PROOF OF THE THEOREM (PART (i)). Let a_k be the general term of a convergent series (for example, $a_k = 2^{-k}$). Fix an increasing sequence $\{n_k\}$ such that $\widetilde{e}(f_m, f_0) < 2a_k$ if $m \geq n_k$. Set

$$A_k = \{x : \ d(f_{n_k}(x), f_0(x)) > a_k\}; \quad \text{then} \quad \mu(A_k) < a_k.$$

The Borel-Cantelli lemma implies that, μ-almost everywhere, x belongs to only finitely many A_k. Thus

for μ-almost every x, there exists an integer $s(x)$ such that $d(f_{n_k}(x), f_0(x)) \leq a_k$ if $k \geq s(x)$.

Hence f_{n_k} converges μ-a.e. to f_0. □

PART (ii). Fix $\epsilon > 0$. Set

$$G_n = \left\{x : \ \sup_{q\geq n} d(f_q(x), f_0(x)) > \epsilon\right\}.$$

Then $\{G_n\}$ is a decreasing sequence and, by 5.1.3 and 2.6.1, $G_n \in \mathcal{A}$. Since $G_n \subset X$ implies that $\mu(G_n) < +\infty$, we can use the limit theorem 3.2.3 to conclude that $\mu\left(\bigcap_n G_n\right) = \lim \mu(G_n)$.

The hypothesis of convergence μ-a.e. implies that the left-hand side is zero. Let n_0 be such that $\mu(G_{n_0}) < \epsilon$; then we have $\widetilde{e}(f_n, f_0) < 2\epsilon$ if $n > n_0$. □

5.2.9 Theorem. *Suppose that $(X, \mathcal{A}, \mu)$ is a measure space and Y is a complete metric space. Then $M_\mu((X, \mathcal{A}); (Y, \mathcal{B}_Y))$, equipped with the metric d_μ, is a complete metric space.*

PROOF. Our approach parallels that of the proof of 5.2.7(ii). Let $\{f_n\}$ be a Cauchy sequence in M_μ; using a result from topology, we need only show that the sequence $\{f_n\}$ has a subsequence that is convergent with respect to the distance d_μ. Let a_k be the general term of a convergent numerical sequence. Fix an increasing subsequence $\{n_k\}$ such that $\widetilde{e}(f_{n_k}, f_m) < 2a_k$ for all $m \geq n_k$. Set

$$A_k = \{x : \ d(f_{n_k}(x), f_{n_{k+1}}(x)) > a_k\};$$

then $\mu(A_k) < a_k$, or

$$d(f_{n_k}(x), f_{n_{k+1}}(x)) \leq a_k \quad \text{if} \quad x \notin A_k.$$

Let $\epsilon > 0$ be given, fix k_0 such that $\sum_{k \geq k_0} a_k < \epsilon$, and set $D_{k_0} = \cup_{k \geq k_0} A_k$. Then

$$(i) \qquad \mu(D_{k_0}) < \epsilon \text{ and } \sum_{k \geq k_0} d(f_{n_k}(x), f_{n_{k+1}}(x)) < \epsilon \quad \text{if} \quad x \notin D_{k_0}.$$

Hence $\{y_k\} = \{f_{n_k}(x)\}$ is a Cauchy sequence if $x \notin D_{k_0}$.

As this is true for every k_0, it follows that $\{f_{n_k}(x)\}$ converges if $x \notin \cap_k D_k$; but $\mu(\cap_k D_k) = 0$, i.e. $\{f_{n_k}\}$ converges μ-a.e. to $f_0 \in M_\mu$.

By inequality (i) and the triangle inequality,

$$(ii) \qquad d(f_{n_k}(x), f_{n_{k'}}(x)) < \epsilon \quad \text{if} \quad k, k' > k_0 \quad \text{and if} \quad x \notin D_{k_0}.$$

Fixing n_k and letting k' go to infinity, we obtain

$$d(f_{n_k}(x), f_0(x)) \leq \epsilon \quad \text{if} \quad k > k_0 \text{ and } x \notin D_{k_0},$$

whence

$$\widetilde{e}(f_{n_k}, f_0) \leq 2\epsilon \quad \text{if} \quad k > k_0,$$

or

$$d_\mu(f_{n_k}, f_0) \leq 2\epsilon \quad \text{if} \quad k > k_0. \square$$

6 The Space of Integrable Functions

In this section, we exhibit a vector subspace of $M_\mu((X, \mathcal{A}); (\mathbf{R}, \mathcal{B}_\mathbf{R})) = L^0_\mu(X, \mathcal{A})$ which will be provided with a Banach space structure. The distance defined by this norm will be an upper bound for the distance d_μ, and will thus define a finer topology than that associated with d_μ.

6.1 Simple measurable functions

Let $(X, \mathcal{A})$ be a measurable space. A *simple function* is a measurable mapping from X to $\mathbf{R}$ such that cardinal $(f(X)) < +\infty$. We denote by $\mathcal{E}^0(X, \mathcal{A})$ the set of simple functions.

Let $(X, \mathcal{A}, \mu)$ be a measure space. We denote by $E^0_\mu(X, \mathcal{A})$ the subset of $L^0_\mu(X, \mathcal{A})$ consisting of those equivalence classes of measurable functions which contain a simple function.

If $f, g \in \mathcal{E}^0(X, \mathcal{A})$, then

$$\mathrm{card}((f+g)(X)) \leq \mathrm{card}(f(X))\mathrm{card}((g(X))$$

and

$$\mathrm{card}((fg)(X)) \leq \mathrm{card}(f(X))\mathrm{card}(g(X)),$$

so that $\mathcal{E}^0(X, \mathcal{A})$ is a vector space equipped with a product. The same holds for $E^0_\mu(X, \mathcal{A})$. Moreover, if $f \in \mathcal{E}^0(X, \mathcal{A})$ so is $|f|$; hence the operation sup is defined on $\mathcal{E}^0$ and E^0_μ.

6.2 Finite σ-algebras

Let Y be an abstract set and let $\mathcal{B}$ be a σ-algebra on Y. $\mathcal{B}$ is called finite if it has only a finite number of elements. Note that the finite Boolean algebras coincide with the finite σ-algebras: the countable union property reduces to the finite union property in this case. Sets $B \in \mathcal{B}$ such that

$$B' \subset B,\ B' \in \mathcal{B} \quad \text{implies either} \quad B' = B \quad \text{or} \quad B' = \emptyset$$

are called *atoms*.

Atoms are the minimal elements with respect to the inclusion relation in a σ-algebra. If B and $\widetilde{B}$ are distinct atoms, then $B \cap \widetilde{B} = \emptyset$.

6.2.1 Proposition. *Let $\mathcal{B}$ be a finite σ-algebra. Then every nonempty set in $\mathcal{B}$ is the union of the atoms it contains.*

PROOF. Let $A \in \mathcal{B}$. Either A is an atom or $\exists A_1 \subset A$, $A_1 \neq A$, $A_1 \in \mathcal{B}$. Repeat the argument, starting from A_1: either A_1 is an atom, or $\exists A_2 \subset A_1$, $A_2 \neq A_1$, $A_2 \in \mathcal{B}$. This produces a sequence of subsets of Y, each strictly contained in the preceding one. Since $\mathcal{B}$ is finite, the process must terminate after finitely many steps, yielding an atom. We have thus shown that

every nonempty set $A \in \mathcal{B}$ contains at least one atom of $\mathcal{B}$.

Let $H_1, \ldots, H_q$ be the atoms of $\mathcal{B}$ contained in A and let $\widetilde{A} = \cup_j H_j$. Then $\widetilde{A} \subset A$. Moreover, $\widetilde{A}^c \cap A \in \mathcal{B}$. If $\widetilde{A}^c \cap A$ were nonempty, $\widetilde{A}^c \cap A$ would contain an atom; but all the atoms contained in A are contained in $\widetilde{A}$, whence $\widetilde{A} = A$. □

6.2.2 Corollary. *Let $\mathcal{B}$ be a finite σ-algebra of subsets of Y. Then there exist a finite set $S_{\mathcal{B}}$ and a bijection between $\mathcal{B}$ and $\mathcal{P}(S_{\mathcal{B}})$, the set of all subsets of $S_{\mathcal{B}}$, such that the bijection respects the Boolean algebra structure (the operations of union and intersection).*

PROOF. We take for $S_{\mathcal{B}}$ the set of atoms of $\mathcal{B}$. The bijection between $\mathcal{B}$ and $S_{\mathcal{B}}$ is obtained by associating with each set $B \in \mathcal{B}$ the atoms it contains. □

6.2.3 Partitions

Definition. A *partition* of X is a finite family of pairwise disjoint subsets of X, say $K_1, \ldots, K_n$, whose union is X. The σ-algebra $\mathcal{B}$ generated by the K_i, $1 \leq i \leq n$, consists of sets B of the form $B = \cup_s K_{j_s}$.

The atoms of $\mathcal{B}$ are precisely the K_i. Conversely, given a finite σ-algebra $\mathcal{B}$ on X, its atoms form a partition of X.

6.2.4 Finite σ-algebras and simple mappings

Let $(X,\mathcal{A})$ be a measurable space and let Y be a metric space. A function $f \in \mathcal{M}((X,\mathcal{A});(Y,\mathcal{B}_Y))$ is called *simple* if $\operatorname{card}(f(X))$ is finite.

Proposition. *A mapping f is simple if and only if $f^{-1}(\mathcal{B}_Y)$ is a finite σ-algebra .*

PROOF. ($\Rightarrow$) Let $x'_1, \ldots\ x'_q$ be an enumeration of the image $f(X)$. Then $f^{-1}(\{x'_k\})$ are the atoms of $f^{-1}(\mathcal{B}_Y)$.

($\Leftarrow$) Let U be an atom of $f^{-1}(\mathcal{B}_Y)$. Suppose that f assumes two distinct values on U, say y_1 and y_2. Let O_1 and O_2 be disjoint open sets in Y, $y_i \in O_i$ $(i = 1,2)$.

Set $U \cap f^{-1}(O_i) = U_i$ $(i = 1,2)$. Then $U_i \in f^{-1}(\mathcal{B}_Y)$, $U_1 \neq \emptyset$, $U_1 \subset U$, and $U_1 \neq U$, contradicting the hypothesis that U is an atom. □

6.3 Simple functions and indicator functions

Given a subset A of X, the *indicator function* of A, written $\mathbf{1}_A$, is the function equal to 1 on A and zero on A^c: $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise.

The next proposition is easily verified.

6.3.1 Proposition. *$\mathbf{1}_A\, \mathbf{1}_B = \mathbf{1}_{A\cap B}$ and $\mathbf{1}_A + \mathbf{1}_C = \mathbf{1}_{A\cup C} + \mathbf{1}_{A\cap C}$. Moreover, A is measurable if and only if $\mathbf{1}_A \in \mathcal{E}^0(X,\mathcal{A})$.*

6.3.2 Proposition. *Suppose that f assumes only finitely many values. Let $\mathcal{B}$ be a finite σ-algebra such that $\mathcal{B} \supset f^{-1}(\mathcal{B}_{\mathbf{R}})$. Then f can be written uniquely in the form*

$$f = \sum \alpha_i \mathbf{1}_{H_i} \quad \textit{with} \quad \alpha_i \in \mathbf{R}, \quad \textit{where the } H_i \textit{ range over the atoms of } \mathcal{B}.$$

PROOF. Let $H_0, \ldots, H_q$ be the atoms of $\mathcal{B}$. Let $\xi \in f(X)$; then the hypothesis $f^{-1}(\xi) \in \mathcal{B}$ implies that $f^{-1}(\xi)$ can be written as a union of atoms. Hence f has constant value, say α_i, on H_i. The two sides of the identity coincide on H_i for every i, and since $\cup_i H_i = X$ the identity holds everywhere. □

6.3.3 Corollary. *The measurable indicator functions generate the vector space of simple functions.*

PROOF. Let f be a simple function and let $\mathcal{B} = f^{-1}(\mathcal{B}_{\mathbf{R}}) \subset \mathcal{A}$. Then $\mathcal{B}$ is a finite σ-algebra by 6.2.4. □

6.4 Approximation by simple functions

6.4.1 Proposition. *Let $f \in \mathcal{L}^0(X,\mathcal{A})$ be bounded. Then there exists a sequence of simple functions g_n converging uniformly to f.*

PROOF. Consider the half-open interval

$$J_k = [kn^{-1}, (k+1)n^{-1}).$$

We may write it as a countable union of closed sets in the following way:

$$J_k = \bigcup_q \left[kn^{-1}, \left(k + \left(1 - \frac{1}{q}\right)\right) n^{-1}\right].$$

Hence J_k is a *Borel subset* of $\mathbf{R}$, and $\mathbf{1}_{J_k} \in \mathcal{L}^0(X, \mathcal{B}_{\mathbf{R}})$.

Let C and k_0 be such that $|f(x)| \leq C$ and

$$\bigcup_{k=-k_0}^{+k_0} J_k \supset [-C, +C].$$

Set

$$G_n = \sum_{-k_0 \leq k \leq +k_0} kn^{-1}\mathbf{1}_{J_k}.$$

Since the $J_k \cap [-C, +C]$ form a partition of $[-C, +C]$, we have $t - n^{-1} < G_n(t) \leq t$ if $|t| \leq C$.

Moreover, G_n takes only finitely many values and

$$G_n \in \mathcal{M}((\mathbf{R}, \mathcal{B}_{\mathbf{R}}), (\mathbf{R}, \mathcal{B}_{\mathbf{R}})).$$

Set $g_n = G_n \circ f$; then $g_n \in \mathcal{E}^0(X, \mathcal{A})$ and

$$|g_n(x) - f(x)| \leq n^{-1}.\square$$

6.4.2 Corollary. *Let $f \in \mathcal{L}^0(X, \mathcal{A})$. Then there exists a sequence $\{\varphi_n\}$ of simple functions converging pointwise to f.*

PROOF. Let $A_n = \{x : \ |f(x)| < n\}$. Then $f_n = \mathbf{1}_{A_n} f$ is a bounded measurable function. Let φ_n be a simple function, constructed (as in 6.4.1) so that

$$|f_n(x) - \varphi_n(x)| \leq n^{-1} \quad \text{for all } x.$$

Then

$$\lim \varphi_n(x) = f(x) \quad \forall x \in X.\square$$

6.4.3 Corollary. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(X, \mathcal{A}', \mu')$ be its completion (in the sense of 4.2.3). Let $f' \in \mathcal{L}^0(X, \mathcal{A}')$ be given. Then there exists $f \in \mathcal{L}^0(X, \mathcal{A})$ such that $f(x) = f'(x)$ μ-a.e.*

In particular, $L^0_\mu(X, \mathcal{A})$ can be identified with $L^0_{\mu'}(X, \mathcal{A}')$.

PROOF. Consider first the indicator function of a set $A' \in \mathcal{A}'$. There exist $B, C \in \mathcal{A}$ such that $B \subset A' \subset C$ and $\mu(C - B) = 0$. In particular, $\mathbf{1}_B = \mathbf{1}_{A'}$ μ'-a.e. Hence the corollary is true for $\mathcal{A}'$-measurable simple functions.

Now let $f' \in \mathcal{L}^0(X, \mathcal{A}')$ be given. By 6.4.2 there exist $\varphi'_n \in \mathcal{E}^0(X, \mathcal{A}')$ such that $\lim \varphi'_n(x) = f(x)\ \forall_{x \in X}$. By the argument above, there exist $\widehat{\varphi}_n \in \mathcal{E}^0(X, \mathcal{A})$ such that $A'_n = \{x : \widehat{\varphi}_n(x) \neq \varphi'_n(x)\}$ satisfies $\mu'(A'_n) = 0$.

Let $A_\infty = \cup_n A'_n$; note that $A_\infty \in \mathcal{A}$ and $\mu(A_\infty) = 0$. Define $\varphi_n(x) = \widehat{\varphi}_n(x)$ if $x \notin A_\infty$ and $\varphi_n(x) = 0$ otherwise. Then $\varphi_n \in \mathcal{L}^0(X, \mathcal{A})$ and $\{\varphi_n(x)\}$ converges for every x. Moreover, setting $\lim \varphi_n(x) = f(x)$, we see by 2.5.1 that $f \in \mathcal{L}^0(X, \mathcal{A})$ and $f'(x) = f(x)$ μ-a.e. □

6.5 *Integrable simple functions*

6.5.1 **Definition.** Simple functions f such that $\mu(\{x : f(x) \neq 0\}) < +\infty$ are called *integrable* simple functions. We denote by $\mathcal{E}^1_\mu(X, \mathcal{A})$ the integrable simple functions and by $E^1_\mu(X, \mathcal{A})$ the equivalence classes in $\mathcal{E}^0(X, \mathcal{A})$ generated by the integrable simple functions.

$\mathcal{E}^1_\mu(X, \mathcal{A})$ is a vector subspace of $\mathcal{E}^0(X, \mathcal{A})$ which is closed under multiplication and absolute value.

6.5.2 Definition of the integral on $\mathcal{E}^1_\mu(X, \mathcal{A})$

Let $f \in \mathcal{E}^1_\mu(X, \mathcal{A})$ be written in the unique form associated with the σ-algebra $f^{-1}(\mathcal{B}_{\mathbf{R}})$, as in 6.3.2:

$$f = \sum_i \alpha_i \mathbf{1}_{H_i} \quad \text{(where } \alpha_i \neq 0\ \forall i\text{)}.$$

The integral of f is defined by the formula

$$I(f) = \sum_i \alpha_i \mu(H_i).$$

If $f_1 \in \mathcal{E}^1_\mu(X, \mathcal{A})$, $f_1 = f$ a.e., then it is easily verified that $I(f) = I(f_1)$. It follows that the function $I(\cdot)$ is defined on $E^1_\mu(X, \mathcal{A})$.

6.5.3 Lemma (Evaluating the integral on certain finite σ-algebras). *Let f be an integrable simple function and let $\mathcal{B}$ be a finite σ-algebra such that $\mathcal{B} \supset f^{-1}(\mathcal{B}_{\mathbf{R}})$. Denoting by $K_1, \ldots, K_r$ the atoms of $\mathcal{B}$, let*

$$f = \sum_q \beta_q \mathbf{1}_{K_q} \quad \textit{(where } \beta_q \neq 0\ \forall q\textit{)}$$

be the decomposition of f given by 6.3.2. Then

$$I(f) = \sum_q \beta_q \mu(K_q).$$

PROOF. Let $\{H_s\}$ be the set of all atoms of $f^{-1}(\mathcal{B}_{\mathbf{R}})$. Since each H_s is in $\mathcal{B}$, H_s can be written as a union of atoms of $\mathcal{B}$: $H_s = \bigcup_{q \in I_s} K_q$, where I_s is a finite set.

On each K_q ($q \in I_s$), $f = \alpha_s$; thus $\alpha_s = \beta_q$ if $q \in I_s$ and

$$\sum_q \beta_q \mu(K_q) = \sum_s \sum_{q \in I_s} \alpha_s \mu(K_q) = \sum_s \alpha_s \sum_{q \in I_s} \mu(K_q).$$

But $\mu(H_s) = \sum_{q \in I_s} \mu(K_q)$. □

6.5.4 Theorem (Properties of the integral on simple functions).

(i) *The integral defines a positive linear functional on* $E^1_\mu(X, \mathcal{A})$.
(ii) *Setting* $q(f) = I(|f|)$ *defines a norm on* $E^1_\mu(X, \mathcal{A})$. *Moreover,* $|I(f)| \leq q(f)$.
(iii) $\mu(\{x : |f_1(x) - f_2(x)| > \eta\}) < \frac{1}{\eta} q(f_1 - f_2)$ *(Chebyshev's inequality).*
(iv) $e(f_1, f_2) \leq 2q(f_1 - f_2)^{\frac{1}{2}}$.
(v) *Every Cauchy sequence in the normed space* $E^1_\mu(X, \mathcal{A})$ *is a Cauchy sequence with respect to the distance of convergence in measure. Convergence in norm implies convergence in measure.*

Proof of (i). Let γ be a constant. Then $I(\gamma f) = \gamma I(f)$ for every $f \in E^1_\mu(X, \mathcal{A})$.

Now let $f_1, f_2 \in \mathcal{E}^1_\mu(X, \mathcal{A})$. Let F be the mapping from X to $\mathbf{R}^2$ defined by setting $F(x) = (f_1(x), f_2(x))$.

Then F is a simple mapping and $F^{-1}(\mathcal{B}_{\mathbf{R}^2}) = \mathcal{B}$ is a finite sub-σ-algebra of $\mathcal{A}$ containing $f_i^{-1}(\mathcal{B}_{\mathbf{R}})$ ($i = 1, 2$). The decomposition of f_i on the σ-alge-$\mathcal{B}$ gives

$$f_1 = \sum \beta_s \mathbf{1}_{K_s} \quad \text{and} \quad f_2 = \sum \delta_s \mathbf{1}_{K_s},$$

where the K_s range over the atoms of $\mathcal{B}$. Then $f_1 + f_2$ can be decomposed in the σ-algebra $\mathcal{B}$ as $f_1 + f_2 = \sum(\beta_s + \delta_s)\mathbf{1}_{K_s}$, whence

$$I(f_1+f_2) = \sum(\beta_s+\delta_s)\mu(K_s) = \sum \beta_s \mu(K_s) + \sum \delta_s \mu(K_s) = I(f_1)+I(f_2).$$

If $f(x) \geq 0$ μ-a.e., the only coefficients appearing in the sum are the non-negative β_s. Thus

(vi) $f(x) \geq 0$ μ-a.e. implies $I(f) \geq 0$ (positivity of the integral).

Proof of (ii). By the positivity of the integral, the inequality $|f + h| \leq |f| + |h|$ implies that $q(f + h) \leq q(f) + q(h)$.

That $q(\alpha f) = |\alpha| q(f)$ is trivial. It remains to show that $q(f) \geq 0$ and that $q(f) = 0$ implies $f = 0$ μ-a.e.

The first inequality follows from the positivity of the integral. Moreover, in a σ-algebra adapted to f,

$$I(|f|) = \sum_{\alpha_i \neq 0} |\alpha_i| \mu(K_i),$$

and this sum of nonnegative terms can be zero only if all the terms are zero.

Finally, $-|f| \leq f \leq |f|$ implies the same inequality for the integrals: $-I(|f|) \leq I(f) \leq I(|f|)$.

PROOF OF (iii). We use the same finite σ-algebra $\mathcal{B}$ as in the proof of (i) and the same decompositions of f_1 and f_2 on the atoms of $\mathcal{B}$. Then $f_1 - f_2 = \sum(\beta_s - \delta_s)\mathbf{1}_{K_s}$ and $q(f_1 - f_2) = \sum_s \gamma_s \mu(K_s)$, where $\gamma_s = |\beta_s - \delta_s|$.

$$\mu(\{x : |f_1(x) - f_2(x)| > \eta\}) = \sum_{s \in J} \mu(K_s), \quad \text{where} \quad J = \{s : \gamma_s > \eta\},$$

and

$$q(f_1 - f_2) \geq \sum_{s \in J} \gamma_s \mu(K_s) > \eta \sum_{s \in J} \mu(K_s).$$

It follows that

$$q(f_1 - f_2) > \eta\mu(\{x : |f_1(x) - f_2(x)| > \eta\}).$$

PROOF OF (iv). Consider the subset $K(f_1, f_2)$ of $(\mathbf{R}^+)^2$, which was used to define $e(f_1, f_2)$:

$$K(f_1, f_2) = \{(\epsilon, \eta) : \mu(|f_1 - f_2| > \eta) \leq \epsilon\}.$$

Then, by (iii),

$$(\eta^{-1} q(f_1 - f_2), \eta) \in K(f_1, f_2) \quad \text{for all} \quad \eta > 0.$$

Hence

$$e(f_1, f_2) = \inf(\epsilon + \eta) \leq \inf_{\eta}(\eta + \eta^{-1} q(f_1 - f_2)).$$

Taking $\eta = [q(f_1 - f_2)]^{\frac{1}{2}}$ shows that $e(f_1, f_2) \leq 2[q(f_1 - f_2)]^{\frac{1}{2}}$. □

PROOF OF (v). It follows immediately from (iv) that a Cauchy sequence in the normed space E^1_μ is a Cauchy sequence with respect to the distance of convergence in measure. Similarly, a sequence that converges to f_0 in norm also converges in measure. □

6.6 *Some spaces of bounded measurable functions*

6.6.0 Definitions

$$\begin{aligned} \mathcal{L}^\infty(X, \mathcal{A}) &= \{f \in \mathcal{L}^0(X, \mathcal{A})) : \exists M < \infty \text{ such that } |f(x)| \leq M\}. \\ \mathcal{L}^{\infty,1}_\mu(X, \mathcal{A}) &= \{f \in \mathcal{L}^\infty(X, \mathcal{A}) : \mu(\{x : f(x) \neq 0)\}) < \infty\}. \end{aligned}$$

6.6.1 Proposition. *For every $f \in \mathcal{L}_\mu^{\infty,1}(X,\mathcal{A})$, there exist $\varphi_n \in \mathcal{E}_\mu^1(X,\mathcal{A})$ such that*

(i) $\{\varphi_n\}$ *converges uniformly to* f, *and*
(ii) $\{x : \varphi_n(x) \neq 0\} = \{x : f(x) \neq 0)\}$.

PROOF. Cf. Proposition 6.4.1. □

6.6.2 Proposition. *If $\{\varphi_n\}$ satisfies 6.6.1, then $\{I(\varphi_n)\}$ is a Cauchy sequence.*

PROOF. Let $K = \{x : f(x) \neq 0)\}$. Then $\varphi_n = \varphi_n \mathbf{1}_K$ and

$$I(\varphi_n - \varphi_m) = I((\varphi_n - \varphi_m)\mathbf{1}_K) \leq \mu(K) \sup|\varphi_n - \varphi_m| \to 0$$

by the uniform convergence of $\{\varphi_n\}$. □

6.6.3 **Definition.** $\widetilde{I}(f) = \lim I(\varphi_n)\ \forall f \in \mathcal{L}_\mu^{\infty,1}$, where $\{\varphi_n\}$ is the sequence of Proposition 6.6.1.

This is independent of the choice of sequence. Let $\{\varphi'_n\}$ be another sequence satisfying 6.6.1(i). Set

$$\begin{aligned} \varphi''_m &= \varphi_{m/2} && \text{if } m \text{ is even, and} \\ \varphi''_m &= \varphi'_{(m-1)/2} && \text{if } m \text{ is odd.} \end{aligned}$$

Then φ''_m satisfies 6.6.1(i) and hence $\lim I(\varphi''_m)$ exists. But this implies that $\lim I(\varphi_n) = \lim I(\varphi'_n)$.

6.6.4 Proposition. *Let $f \in \mathcal{L}_\mu^{\infty,1}$. Then the following statements are true:*

(i) $\widetilde{I}(f_1 + f_2) = \widetilde{I}(f_1) + \widetilde{I}(f_2)$.
(ii) $f_1 \geq f_2 \Rightarrow \widetilde{I}(f_1) \geq \widetilde{I}(f_2)$.
(iii) $f_1 = f_2$ *a.e.* $\Rightarrow \widetilde{I}(f_1) = \widetilde{I}(f_2)$.

6.7 The truncation operator

For a fixed positive integer n, let φ_n be the continuous function defined on $\mathbf{R}$ by

$$\begin{aligned} \varphi_n(t) &= t && \text{if } -n \leq t \leq +n \\ \varphi_n(t) &= n && \text{if } t > n \\ \varphi_n(t) &= -n && \text{if } t < -n. \end{aligned}$$

Let $A_1 \subset A_2 \ldots \subset A_n \ldots$ be an exhaustion of X, i.e. $\mu(A_k) < +\infty\ \forall k$ and $X = \cup_k A_k$.

We define T_n, the truncation operator of order n on $\mathcal{L}^0(X,\mathcal{A})$, as follows:

6.7.1 $$T_n(f) = f_n \mathbf{1}_{A_n}, \quad \text{where} \quad f_n = \varphi_n \circ f.$$

f_n is bounded and (since φ_n is continuous) measurable. Furthermore, since the set $\{x : (T_n f)(x) \neq 0\} \subset A_n$, it has finite measure. Hence, by the definition of $\mathcal{L}_\mu^{\infty,1}$,

6.7.2 $$T_n(f) \in \mathcal{L}_\mu^{\infty,1}(X,\mathcal{A}) \quad \text{for any} \quad f \in \mathcal{L}^0(X,\mathcal{A}).$$

6.8 Construction of L^1

6.8.1 Definition of $L^1_\mu(X,\mathcal{A})$

(i) **Definition.** $\mathcal{L}^1_\mu(X,\mathcal{A}) = \{f \in \mathcal{L}^0(X,\mathcal{A}) : \lim_{n\to\infty} \widetilde{I}(|T_n(f)|) < +\infty\}$.

Proposition. *If $f_1 \in \mathcal{L}^1_\mu$ and $f_2 = f_1$ a.e., then $f_2 \in \mathcal{L}^1_\mu(X,\mathcal{A})$. This justifies the notation*

$$L^1_\mu(X,\mathcal{A}) = \{\textit{equivalence classes of } \mathcal{L}^1_\mu(X,\mathcal{A})\}.$$

(ii) $\|f\|_{L^1} = \lim \widetilde{I}(|T_n(f)|)$.
(iii) If $f \in L^0_\mu$ and $|f| \leq |h|$, where $h \in L^1_\mu$, then $f \in L^1_\mu$.
(iv) If $f \in \mathcal{L}_\mu^{\infty,1}(X,\mathcal{A})$, then $f \in \mathcal{L}^1_\mu(X,\mathcal{A})$.

6.8.2 Proposition. *If $f \in L^1_\mu$, then $\lim_{n\to\infty} \tilde{I}(T_n(f))$ exists.*

PROOF. Let $f^+ = \sup(f,0)$ and let $f^- = \sup(-f,0)$. Although T_n is not a linear operator, it is elementary to verify that, for all $x \in X$,

$$T_n(f)(x) = T_n(f^+)(x) - T_n(f^-)(x)$$

and

$$|T_n(f)| = T_n(f^+) + T_n(f^-),$$

whence

$$\widetilde{I}(T_n(f^+)) \leq \widetilde{I}(|T_n(f)|) \leq \|f\|_{L^1}.$$

$\{\widetilde{I}(T_n(f^+))\}$ is thus an increasing sequence which is bounded above, and therefore converges. □

Definition. For $f \in L^1_\mu$, the integral of f is defined by $\displaystyle\int f = \lim \tilde{I}(T_n(f))$.

6.8.3 Proposition. *L^1_μ is a vector space with the following properties:*

(i) $\int (f_1 + f_2) = \int f_1 + \int f_2$.
(ii) If $f \geq 0$, then $\int f \geq 0$.

Set $\|f\|_{L^1} = \int |f|$. Then

(iii) $|\int f| \leq \|f\|_{L^1}$.
(iv) $\mu(\{x : f(x) > c\}) \leq \frac{1}{c}\|f\|_{L^1}$.
(v) $\|f\|_{L^1}$ is a norm.

PROOF. The statements clearly hold for $\mathcal{L}_\mu^{\infty,1}$ and pass to L^1_μ. □

7 Theorems on Passage to the Limit under the Integral Sign

7.1 Fatou-Beppo Levi theorem. *Let $\{f_n\}$ be an increasing sequence of integrable functions such that $\int f_n \leq C$, where C is a constant independent of n. Then*

(i) $\lim f_n = f_\infty$ *exists and is finite μ-a.e.,*
(ii) $f_\infty \in L^1_\mu$, *and*
(iii) $\|f_n - f_\infty\|_{L^1} \to 0$.

PROOF. By setting $\widetilde{f_n} = f_n - f_1$, we may assume that $f_n \geq 0$. Then $\int f_n = \|f_n\|_{L^1}$, $T_q(f_\infty) = \lim_n T_q(f_n)$, and $\int T_q(f_n) \leq \int f_n \leq C$. It follows that $\|T_q(f_\infty)\|_{L^1} \leq C$, whence $\mu(\{x : (T_q(f_\infty))(x) > n\}) < Cn^{-1}$. Furthermore,

$$\{x : f_\infty(x) > n\} = \lim \uparrow \{n : (T_q(f_\infty))(x) > n\}.$$

Thus

$$\mu(\{x : f_\infty(x) > n\}) \leq \frac{C}{n} \quad \text{and} \quad \|T_q(f_\infty)\|_{L^1} \leq C.$$

Hence $f_\infty \in L^1$.

We now show that $\|f_\infty - f_n\|_{L^1} \to 0$. Let $u_n = f_\infty - f_n$. Then

$$T_q(u_1) - T_{q_0}(u_1) \geq T_q(u_n) - T_{q_0}(u_n), \quad \text{where} \quad q_0 < q.$$

Let q_0 be chosen so that $\int T_q(u_1) - \int T_{q_0}(u_1) < \frac{\epsilon}{2}$. Then

$$\|u_n\| < \frac{\epsilon}{2} + \|T_{q_0}(u_n)\|_{L^1}.$$

Let $v_n = T_{q_0}(u_n)$. Then $0 \leq v_n \leq q_0$, $v_n(x) = 0$ if $x \in A^c_{q_0}$, and $v_n \to 0$ a.e. Recall, from 6.7, that $\mu(A_{q_0}) < +\infty$.

By Egoroff's theorem, there exists K such that $\mu(K^c) < \dfrac{\epsilon}{4q_0}$ and v_n converges uniformly to zero on K. Hence

$$\|u_n\| < \frac{\epsilon}{2} + \frac{\epsilon}{4} + \mu(A_{q_0}) \sup_{x \in K}(v_n(x)) \to 0 \quad \text{as} \quad n \to \infty. \square$$

7.2 Lebesgue's theorem on series. *Let $\{u_n\}_{n=1}^\infty$ be a sequence of elements of L^1 such that $\sum \|u_n\|_{L^1} < \infty$. Then $\sum_1^\infty u_n$ converges absolutely a.e. Let $s_n = u_1 + \ldots + u_n$ and let $s_\infty = \lim_n s_n$. Then $s_\infty \in L^1$, $\int s_\infty = \lim \int s_n$, and $\|s_\infty - s_n\|_{L^1} \to 0$.*

PROOF. Set $f_n(x) = \sum_{k=1}^n |u_k(x)|$. Then $\{f_n\}$ is an increasing sequence and $\int f_n \leq \sum_{k=1}^{+\infty} \|u_k\|_{L^1} < +\infty$. By the theorem of Fatou-Beppo Levi, this implies that $\lim f_n = f_\infty$ exists, $f_\infty \in L^1$, and $f_\infty < +\infty$ a.e. Thus $s_\infty \in L^1$ since $|s_\infty| \leq f_\infty$, and $\|s_\infty - s_n\|_{L^1} \leq \|f_\infty - f_n\|_{L^1}$, which approaches zero by Fatou-Beppo Levi.

7.3 Proposition. *The truncation operator is a contraction on $L^1_\mu(X,\mathcal{A})$; that is,*

$$\|T_n(f) - T_n(\widetilde{f})\|_{L^1} \leq \|f - \widetilde{f}\|_{L^1}, \quad \forall f, \widetilde{f} \in L^1_\mu(X,\mathcal{A}).$$

PROOF. Assume first that f and $\widetilde{f}$ are simple functions. Let $\mathcal{B}$ be the σ-algebra generated by $f^{-1}(\mathcal{B}_{\mathbf{R}})$, $\widetilde{f}^{-1}(\mathcal{B}_{\mathbf{R}})$, and $\{A_n\}$, and let $\mathcal{S}$ denote the atoms of $\mathcal{B}$. Then

$$f = \sum \alpha_k \mathbf{1}_{H_k} \quad \text{and} \quad \widetilde{f} = \sum \widetilde{\alpha}_k \mathbf{1}_{H_k}, \quad \text{where} \quad H_k \in \mathcal{S}.$$

Let $I = \{H \in \mathcal{S} : H \cap A_n \neq 0\}$. Then

$$\begin{aligned} T_n(f) &= \sum_{k\in I} \varphi_n(\alpha_k)\mathbf{1}_{H_k}, \\ T_n(\widetilde{f}) &= \sum_{k\in I} \varphi_n(\widetilde{\alpha}_k)\mathbf{1}_{H_k}, \\ |T_n(f) - T_n(\widetilde{f})| &= \sum_{k\in I} |\varphi_n(\alpha_k) - \varphi_n(\widetilde{\alpha}_k)|\mathbf{1}_{H_k}. \end{aligned}$$

Using the elementary inequality $|\varphi_n(t) - \varphi_n(t')| \leq |t - t'|, \quad \forall t, t' \in \mathbf{R}$,

$$\|T_n(f) - T_n(\widetilde{f})\|_{L^1} \leq \sum_{k\in I} \mu(H_k)|\alpha_k - \widetilde{\alpha}_k| \leq \sum_{k\in\mathcal{S}} \mu(H_k)|\alpha_k - \widetilde{\alpha}_k| = \|f - \widetilde{f}\|_{L^1}.$$

Now let f and $\widetilde{f} \in L^1$. We can find two sequences h_q, $\widetilde{h}_q$ of simple functions converging in the L^1 norm to f and $\widetilde{f}$. Passing if necessary to a subsequence, we may suppose in addition that h_q and $\widetilde{h}_q$ converge a.e. Then $\|T_n(h_q) - T_n(\widetilde{h}_q)\|_{L^1} \leq \|h_q - \widetilde{h}_q\|_{L^1}$; hence $T_n(h_q)$ is a Cauchy sequence in the L^1 norm. Let k be its limit. Then $k = T_n(f)$ since h_q converges a.e. to f, and hence

$$\|T_n(h_q) - T_n(f)\|_{L^1} \to 0.$$

It follows that

$$\|T_n(f) - T_n(\widetilde{f})\|_{L^1} = \lim_q \|T_n(h_q) - T_n(\widetilde{h}_q)\|_{L^1} \leq \lim_q \|h_q - \widetilde{h}_q\|_{L^1} = \|f - \widetilde{f}\|_{L^1}.$$

□

7.4 Integrability criteria

7.4.1 Theorem. *Let $f \in L^0_\mu(X,\mathcal{A})$. Then $f \in L^1_\mu(X,\mathcal{A})$ if and only if there exists a constant C such that, for all n, $\|T_n(f)\|_{L^1} \leq C$.*

PROOF. ($\Leftarrow$) Applying 7.3 with $\widetilde{f} = 0$ yields $\|T_n(f)\|_{L^1} \leq \|f\|_{L^1}$.

($\Rightarrow$) We prove this first in the special case that $f \geq 0$, where $T_n(f) \leq T_{n+1}(f)$. By the Fatou-Beppo Levi theorem, there exists $g \in L^1$ such that $\lim T_n(f) = g$

a.e. Moreover, a direct calculation shows that $\lim T_n(f)(x) = f(x)$ for all $x \in X$. Hence $f = g$, and therefore $f \in L^1$.

For the general case, set $f^+ = \sup(f, 0)$ and $f^- = \sup(-f, 0)$. Then $f^+, f^- \in L^0$, f^+, f^- are positive, and $f = f^+ - f^-$.

Since $|T_n(f)| = T_n(f^+) + T_n(f^-)$ (cf. the proof of Proposition 6.8.2),

$$\|T_n(f)\|_{L^1} = \int |T_n(f)| = \int T_n(f^+) + \int T_n(f^-).$$

It follows that

$$\|T_n(f^+)\|_{L^1} \leq C \quad \text{and} \quad \|T_n(f^-)\|_{L^1} \leq C.$$

Since f^+ and f^- are nonnegative, this implies that f^+ and $f^- \in L^1$ and hence that $f \in L^1$. □

7.4.2 Corollary. *Let $f \in L^0_\mu(X, \mathcal{A})$. Then $f \in L^1_\mu(X, \mathcal{A})$ if and only if $|f| \in L^1(X, \mathcal{A})$.*

PROOF. The direct implication follows from 2.4.4 and 6.8.1(iii).

Conversely, assume that $|f| \in L^1_\mu(X, \mathcal{A})$. It is easy to see that $|T_n(f)| = T_n(|f|)$, whence $\|T_n(f)\|_{L^1} = \|T_n(|f|)\|_{L^1}$. The conclusion follows by applying Theorem 7.4.1. □

7.4.3 Corollary. *Let $f \in L^0_\mu(X, \mathcal{A})$ and suppose that there exists $u \in L^1_\mu(X, \mathcal{A})$ such that $|f| \leq u$. Then $f \in L^1_\mu(X, \mathcal{A})$.*

PROOF. $\|T_n(f)\|_{L^1} \leq \|T_n(u)\|_{L^1} \leq \|u\|_{L^1}$. □

7.5 Definition of the integral on a measurable set

Let $(X, \mathcal{A}, \mu)$ be a measure space and let Y be a fixed element of $\mathcal{A}$. We denote by $\mathcal{A}'$ the trace on Y of the σ-algebra $\mathcal{A}$ and by μ' the restriction of μ to the elements of $\mathcal{A}'$, thus obtaining a measure space $(Y, \mathcal{A}', \mu')$. Let j be the canonical injection of Y into X. The restriction operator defines a mapping $L^0_\mu(X, \mathcal{A}) \mapsto L^0_{\mu'}(Y, \mathcal{A}')$ by $f \to f \circ j$.

Let $f \in L^1_\mu(X, \mathcal{A}, \mu)$. We denote by $\int_Y f$ the integral of $f \circ j$ evaluated on the measure space $(Y, \mathcal{A}', \mu')$, and call $\int_Y f$ the integral of f on Y.

7.5.1 Proposition. *Let $f \in L^1_\mu(X, \mathcal{A})$. Then $f\mathbf{1}_Y \in L^1_\mu(X, \mathcal{A})$ and $\int f\mathbf{1}_Y = \int_Y f$.*

PROOF. Since $|f\mathbf{1}_Y| \leq |f|$, Corollary 7.4.3 implies that $f\mathbf{1}_Y \in L^1_\mu(X, \mathcal{A})$. The result follows by verifying that the integrals agree on simple functions and passing to the limit.□

7.5.2 Proposition. *Let $f \geq 0$, $f \in L^1_\mu(X, \mathcal{A})$, and set $\rho(A) = \int_A f \quad \forall A \in \mathcal{A}$. Then ρ is a measure on X and $\rho(X) < +\infty$.*

PROOF. Finite additivity follows from the fact that

$$\mathbf{1}_{A_1} + \mathbf{1}_{A_2} = \mathbf{1}_{A_1 \cup A_2} \quad \text{if} \quad A_1 \cap A_2 = \emptyset.$$

The theorem of Fatou-Beppo Levy implies that ρ is continuous on increasing sequences; this gives countable additivity. □

7.5.3 Proposition. *Let A_n be an increasing sequence of elements of $\mathcal{A}$ such that $\cup A_n = X$. Let $f \in L^0_\mu(X, \mathcal{A})$. Suppose that $\int_{A_n} |f|$ is bounded above by a constant C independent of n. Then $f \in L^1_\mu(X, \mathcal{A})$.*

PROOF. Since $\|T_n(f)\|_{L^1} \leq \int_{A_n} |f|$, the result follows from 7.4.1. □

7.6 Lebesgue's dominated convergence theorem

Theorem. *Let $f_n \in L^1_\mu(X, \mathcal{A})$. Suppose that*

(i) f_n converges to h μ-a.e.

and that

(ii) $\exists g \in L^1_\mu(X, \mathcal{A})$ such that $|f_n| \leq g$ $\forall n$ (domination hypothesis).

Then $h \in L^1$,

(iii)
$$\|f_n - h\|_{L^1} \to 0,$$

and

(iv)
$$\int f_n \to \int h.$$

PROOF. It follows from 5.1.2 that $h \in L^0_\mu(X, \mathcal{A})$. By (ii) and 7.4.3, $h \in L^1$.

As in 7.5.2, we introduce the measure ρ associated with g:

$$\rho(A) = \int_A g.$$

Let $\{A_n\}$ be an exhaustion sequence for X: $A_n \subset A_{n+1}$ and $\mu(A_n) < +\infty$. Then $\rho(A_n) \to \rho(X) < +\infty$. Fix m such that

$$\rho(A^c_m) < \frac{\epsilon}{6}.$$

For this fixed m, we will apply Egoroff's theorem (5.1.4) to A_m. We can find a sequence $\{K_q\}$ of sets in $\mathcal{A}$ such that $K_q \subset K_{q+1}$, $f_n \to f_0$ uniformly on K_q, and $\mu(K^c_q \cap A_m) < q^{-1}$.

Set $G_q = K^c_q \cap A_m$. Then $\{G_q\}$ is a decreasing sequence; setting $H = \cap G_q$, we have $\lim \rho(G_q) = \rho(H)$. But $\mu(H) = 0$, whence $g \cdot \mathbf{1}_H = 0$ μ-a.e.; i.e., $g \cdot \mathbf{1}_H = 0$ in L^1_μ and $\rho(H) = 0$. Fix q_0 such that

$$\rho(G_{q_0}) < \frac{\epsilon}{6}.$$

The identity

$$\mathbf{1}_X = \mathbf{1}_{A^c_m} + \mathbf{1}_{K_{q_0}} + \mathbf{1}_{G_{q_0}}$$

gives

$$\int |f_n - h| = \int_{A_m^c} + \int_{K_{q_0}} + \int_{G_{q_0}} |f_n - h|.$$

Using the upper bound $2g$ for the function $|f_n - h|$ in the first and last integrals, we obtain

$$\|f_n - h\|_{L^1} \le 2\rho(A_m^c) + 2\rho(G_{q_0}) + \int_{K_q} |f_n - h|.$$

Each of the first two terms is bounded above by $\epsilon/3$. Furthermore,

$$\int_{K_{q_0}} |f_n - h| \le \left(\sup_{x \in K_{q_0}} |f_n(x) - h(x)| \right) \mu(K_{q_0}).$$

The last term tends to zero as $n \to +\infty$, proving (iii). Finally, (iv) follows from the continuity of the integral with respect to the norm $\|\cdot\|_{L^1}$ (cf. 6.8.3(ii)). □

7.7 Fatou's lemma. *Let $f_n \in L^1_\mu(X, \mathcal{A})$. Suppose that*

(i) $\|f_n\|_{L^1} \le C$, where C is a constant independent of n, and
(ii) f_n converges μ-a.e. to h.

Then

(iii) $h \in L^1$ and $\|h\|_{L^1} \le C$.

PROOF. We prove this first with the additional hypothesis

(iv) $\mu(X) < +\infty$.

In this case, convergence a.e. implies by Egoroff's theorem that, for every integer $q > 0$, there exists $K_q \subset X$ such that f_n converges uniformly on K_q to h and $\mu(K_q^c) \le \frac{1}{q}$. Thus

$$\left| \int_{K_q} |f_n| - \int_{K_q} |h| \right| \le \mu(K_q) \sup_{x \in K_q} |f_n(x) - h(x)|.$$

Since $f_n(x)$ converges uniformly to $h(x)$ on K_q, the last expression tends to zero, whence

$$\int_{K_q} |h| \le C.$$

Set $h_q = |h| \cdot \mathbf{1}_{K_q}$. Then $\{h_q\}$ is an increasing sequence since $K_q \subset K_{q+1}$, and the Fatou-Beppo Levi theorem implies that

$$\lim h_q = h_0 \in L^1 \quad \text{and} \quad \|h_0\|_{L_1} \le C.$$

If X does not have finite measure, take an exhaustion sequence for X:

$$X = \cup A_r, \quad A_r \subset A_{r+1}, \quad \mu(A_r) < +\infty.$$

For each fixed r, set $f_n^r = f_n \mathbf{1}_{A_r}$; then $\|f_n^r\|_{L^1} \leq C$. Fatou's lemma for finite measures can be applied to A_r, giving

$$h^r = \lim f_n^r = h\mathbf{1}_{A_r} \in L^1 \quad \text{and} \quad \|h^r\|_{L^1} \leq C.$$

The conclusion follows by applying the Fatou-Beppo Levi theorem to the increasing sequence $K_r = |h|\mathbf{1}_{A_r}$. □

7.8 Applications of the dominated convergence theorem to integrals which depend on a parameter

7.8.1 Integral notation in which the measure μ appears

Up to now, we have dealt only with functions defined on the measure space $(X, \mathcal{A}, \mu)$. When we consider functions defined on different spaces, the integral notation used earlier can lead to confusion, and we denote

$$\int f \quad \text{by} \quad \int_X f(x)d\mu(x) \quad \text{for all} \quad f \in L^1_\mu(X, \mathcal{A}).$$

7.8.2 Integrals depending on a parameter

Let $(X, \mathcal{A}, \mu)$ be a complete measure space. Consider a metric space Y and let

$$u(y) = \int_X k(x, y)d\mu(x)$$

be an integral depending on the parameter y. Suppose that

(i) for each fixed y the function $k_y(x) = k(x, y)$ satisfies $k_y \in L^1_\mu(X, \mathcal{A})$.

Then $u(y)$ is a well-defined function for every y.

7.8.3 Proposition (Continuity of an integral depending on a parameter). *Assume condition* (i) *of 7.8.2. Let $y_0 \in Y$ and assume in addition that*

(ii) for every sequence $y_n \to y_0$,

$$k(x, y_n) \to k(x, y_0) \quad \mu\text{-a.e.}; \quad \text{and}$$

(iii) there exist $g \in L^1_\mu(X, \mathcal{A})$ and $\epsilon > 0$ such that

$$|k(x, y)| \leq g(x) \quad \text{if} \quad d(y, y_0) < \epsilon.$$

Then the function u is continuous at y_0.

PROOF. Since u is defined on a metric space, in order to show continuity at y it suffices to prove that $u(y_n) \to u(y_0)$ for every sequence $\{y_n\}$ converging to y_0. Set $f_n(x) = k(x, y_n)$. Then the dominated convergence theorem (7.6) can be applied and

$$\int f_n \to \int f_0. \square$$

7.8.4 Proposition (Differentiability of an integral depending on a parameter). *Let $Y = (y_0 - \epsilon, y_0 + \epsilon)$ be an open interval in $\mathbf{R}$, and suppose that the following three conditions hold:*

(i) 7.8.2(i) is satisfied $\forall y \in Y$.
(ii) For μ-almost every x, $\frac{\partial k}{\partial y}(x, y_0)$ exists $\forall y \in Y$ and is continuous at y_0 as a function of y.
(iii) $\exists g \in L^1_\mu(X, \mathcal{A})$ such that, for μ-almost every x, $|\frac{\partial k}{\partial y}(x, y)| \leq g(x)$ for every $y \in Y$.

Then u is differentiable at y_0 and

$$(iv) \qquad u'(y_0) = \int_X \frac{\partial k}{\partial y}(x, y_0) d\mu(x).$$

PROOF. In order to show that u is differentiable, we must show that there exists l such that

$$\lim_{\epsilon \to 0} \epsilon^{-1}[u(y_0 + \epsilon) - u(y_0)] = l.$$

Since $\mathbf{R}$ is a metric space it suffices to show that there exists l such that, for *every sequence* $\{\epsilon_n\}$ tending to zero,

$$\lim \epsilon_n^{-1}[u(y_0 + \epsilon_n) - u(y_0)] = l.$$

Making this detour lets us apply Lebesgue's theorem, which was stated for sequences of functions. Fixing the sequence $\{\epsilon_n\}$, set

$$\epsilon_n^{-1}[u(y_0 + \epsilon_n) - u(y_0)] = \int_X f_n(x) d\mu(x),$$

where

$$f_n(x) = \epsilon_n^{-1}[k(x, y_0 + \epsilon) - k(x, y_0)].$$

Let K be the negligible set such that (ii) and (iii) are satisfied in K^c. Then, for $x \in K^c$, f_n can be calculated using the mean value theorem:

$$f_n(x) = \frac{\partial k}{\partial y}(x, y_0 + \theta_n(x)), \quad \text{where} \quad |\theta_n(x)| < \epsilon_n \quad \text{if} \quad x \in K^c.$$

Thus it follows from (ii) that

$$f_n(x) \to \frac{\partial k}{\partial y}(x, y_0) \quad \text{if} \quad x \in K^c.$$

Furthermore, by (iii), $|f_n(x)| \le g(x)$, $x \in K^c$; thus

$$|f_n(x)| \le g(x) \text{ a.e.} \quad \text{and} \quad \lim f_n(x) = \frac{\partial k}{\partial y}(x, y_0) \text{ a.e.}$$

Applying the dominated convergence theorem gives

$$\int_X f_n(x) d\mu(x) \to \int \frac{\partial k}{\partial y}(x, y) d\mu(x).\square$$

8 Product Measures and the Fubini-Lebesgue Theorem

8.1 Definition of the product measure

Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be measure spaces, let $X = X_1 \times X_2$ be the product space, and let $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ be the product σ-algebra (see 1.5). The *product measure* is a measure μ defined on the measurable space $(X, \mathcal{A})$ and satisfying

$$(i) \qquad \mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \text{if} \quad \mu_i(A_i) < +\infty \quad (i = 1, 2).$$

8.2 Proposition (Uniqueness). *There exists at most one product measure.*

PROOF. Let μ and $\widetilde{\mu}$ be two measures satisfying 8.1(i). Then they coincide on rectangles and hence, by finite additivity, on disjoint unions of rectangles, that is on the Boolean algebra $\mathcal{E}$ of elementary sets. Let

$$\mathcal{M} = \{Z \in \mathcal{A} : \mu(Z) = \widetilde{\mu}(Z)\}; \quad \text{then} \quad \mathcal{M} \supset \mathcal{E}.$$

Let $\{Z_n\}$ be an increasing sequence of sets in $\mathcal{M}$. Then, by 3.2.1,

$$\mu(\cup_n Z_n) = \lim \mu(Z_n) = \lim \widetilde{\mu}(Z_n) = \widetilde{\mu}(\cup Z_n).$$

Thus $\mathcal{M}$ is closed under increasing limits.

If we further assume that

$$(i) \qquad \mu_1(X_1) < +\infty \quad \text{and} \quad \mu_2(X_2) < +\infty,$$

then

$$\mu(X) = \mu_1(X_1)\mu_2(X_2) < +\infty.$$

3.2.3 can be applied to prove that $\mathcal{M}$ is closed under decreasing limits. Hence $\mathcal{M}$ is a *monotone class* that contains $\mathcal{E}$, and it follows by 1.5.5 that $\mathcal{M} = \mathcal{A}_1 \otimes \mathcal{A}_2$.

To complete the proof, it remains to lift the restriction (i). Let $\{Y_n\}$ and $\{Z_n\}$ be exhaustions of X_1 and X_2, and let μ_n and $\widetilde{\mu}_n$ denote the restrictions of μ and $\widetilde{\mu}$ to Y_n and Z_n. Then, by the result above,

$$\mu_n = \widetilde{\mu}_n.$$

Furthermore, $\mu_n(Y_n \times Z_n) < +\infty$ and $\cup_n(Y_n \times Z_n) = X$. Thus $Y_n \times Z_n$ is an exhaustion of X with respect to both μ and $\widetilde{\mu}$. By 3.2.1, for all $A \in \mathcal{A}$

$$\mu(A) = \lim_n \mu_n(A \cap (X_n \times Z_n))$$

and

$$\widetilde{\mu}(A) = \lim_n \widetilde{\mu}_n(A \cap (X_n \times Z_n)).$$

Since the two right-hand sides are equal, $\mu(A) = \widetilde{\mu}(A)$. □

Sections

For fixed x_1, let i_{x_1} denote the injection of X_2 into X defined by $x_2 \mapsto (x_1, x_2)$. For $Z \in \mathcal{P}(X)$, let $Z_{x_1} = i_{x_1}^{-1}(Z)$. Z_{x_1} is called the section of Z over x_1. Letting π_i be the projection of X onto X_i, we have $Z_{x_1} = \pi_2(\pi_1^{-1}(x_1) \cap Z)$.

8.3 Fundamental lemma. *Let $A \in \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$. Then*

(i) $A_{x_1} \in \mathcal{A}_2 \quad \forall x_1 \in X_1$.
(ii) Suppose that $\mu_2(X_2) < +\infty$ and set $k_A(x_1) = \mu_2(A_{x_1})$.

Then

(iii) $k_A \in L^0((X_1, \mathcal{A}_1)) \quad \forall A \in \mathcal{A}$.

PROOF. Since $\mathcal{A}$ is generated by the rectangles R, Theorem 2.2.1 implies that $i_{x_i}^{-1}(\mathcal{A})_{x_i}$ is generated by $\{i_{x_i}^{-1}(R)\}$. But $\{i_{x_i}^{-1}(R)\} = \mathcal{A}_2$; since $\mathcal{A}_2$ is a σ-algebra, it coincides with the σ-algebra it generates, whence (i). Let

$$\mathcal{M} = \{B \in \mathcal{A}\colon k_B(x_1) \text{ is a measurable function of } x_1\}.$$

The rectangles are in $\mathcal{M}$, as are finite unions of disjoint rectangles; thus the Boolean algebra of elementary sets is contained in $\mathcal{M}$. We now show that $\mathcal{M}$ is a monotone class.

Let B_n be an increasing sequence of elements of $\mathcal{M}$. By the limit theorem (3.2.1), $k_{A_n}(x_1) = \mu_2((B_n)_{x_1})$ satisfies $\lim k_{B_n}(x_1) = k_{B_\infty}(x_1)$, where $B_\infty = \cup B_n$. Hence $k_{A_\infty}(x_1)$ is measurable with respect to x_1 by 2.5, which implies that $A_\infty \in \mathcal{M}$.

Since $\mu_2(X_2) < +\infty$, Theorem 3.2.3 on the limits of decreasing sequences can also be applied, and it follows that $\mathcal{M}$ is a monotone class. Since $\mathcal{M}$ contains the Boolean algebra of elementary sets, $\mathcal{M} = \mathcal{A}$ by 1.5.5.□

8.4 Construction of the product measure

8.4.1 Theorem. *Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be measure spaces.*

(i) Suppose that $\mu_1(X_1) < +\infty$ and $\mu_2(X_2) < +\infty$.

For every $A \in \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$, set

(ii) $\rho(A) = \int_{X_1} k_A(x_1) d\mu_1(x_1)$ where $k_A(x_1) = \mu_2(A_{x_1})$.
($\rho(A)$ is well defined by Lemma 8.3.)

(iii) Then ρ is a measure on $\mathcal{A}$, of total mass $\mu_1(X_1)\mu_2(X_2) < +\infty$.

Moreover,

(iv) $\rho(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ if $A_i \in \mathcal{A}_i$.

Proof. Since ρ is a finite measure, it suffices to prove that the σ-additivity axiom is satisfied. We begin by proving *finite additivity.* Suppose that

$$A = A' \cup A'' \quad \text{and} \quad A' \cap A'' = \emptyset.$$

Then $A'_{x_1} \cap A''_{x_1} = \emptyset$, whence $k_{A'}(x_1) + k_{A''}(x_1) = k_A(x_1)$ and

$$\rho(A) = \rho(A') + \rho(A'').$$

Now let $A^p \subset A^{p+1} \subset \ldots$ be an increasing sequence of elements of $\mathcal{A}$. Set $A^\infty = \cup A^p$; then $\lim \uparrow (A^p_{x_1}) = (A^\infty)_{x_1}$ and, by 3.2.1, $k_{A^p}(x_1) \to k_{A^\infty}(x_1)$ for all x_1. Next, $k_{A^p} \leq k_{A^{p+1}}$. Applying Theorem 7.1, the theorem of Fatou-Beppo Levi,

$$\lim \int k_{A^p}(x_1) d\mu_1(x_1) \to \int k_{A^\infty}(x_1) d\mu_1(x_1), \quad \text{i.e.} \quad \lim \rho(A^p) = \rho(\lim A^p).$$

This property, together with finite additivity and 3.2.4, gives σ-additivity; hence ρ is a measure. It is trivial to see that (iv) is satisfied. □

8.4.2 Theorem on reversing the order of integration

Theorem. *Let $(X_1, \mathcal{A}_1, \mu_1)$, $(X_2, \mathcal{A}_2, \mu_2)$ be measure spaces. Suppose that $\mu_1(X) < +\infty$ and $\mu_2(X) < +\infty$. Then, if $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$,*

$$\int_{X_1} d\mu_1(x_1) \left[\int_{X_2} \mathbf{1}_A(x_1, x_2) d\mu_2(x_2) \right] = \int_{X_2} d\mu_2(x_2) \left[\int_{X_1} \mathbf{1}_A(x_1, x_2) d\mu_1(x_1) \right].$$

Proof. Although the hypotheses in 8.4.1 are symmetric in X_1 and X_2, the construction is not.

Set $l_A(x_2) = \mu_1(A_{x_2})$. Then $\sigma(A) = \int l_A(x_2) d\mu_2(x_2)$ exists and defines a product measure by 8.4.1. By 8.2, $\sigma(A) = \rho(A)$ $\forall A \in \mathcal{A}$. □

NOTATION. The product measure is denoted by $\mu_1 \otimes \mu_2$. By definition, for all $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$,

8.4.3
$$\begin{aligned}\int_X \mathbf{1}_A(x)d(\mu_1 \otimes \mu_2)(x) &= \int_{X_1} d\mu_1(x_1) \int_{X_2} \mathbf{1}_A(x_1,x_2)d\mu_2(x_2) \\ &= \int_{X_2} d\mu_2(x_2) \int_{X_1} \mathbf{1}_A(x_1,x_2)d\mu_1(x_1).\end{aligned}$$

8.4.4 Construction of the product measure in the general case

If μ_1 and μ_2 are not finite measures, let X_1^n and X_2^n be exhaustions of X_1 and X_2. Set $\mu_i^n = \mathbf{1}_{X_i^n}\mu_i$. Then $\mu_i^n(X_i) < +\infty$, $i = 1, 2$. We can define $\mu_1^n \otimes \mu_2^n$ and set

$$(\mu_1 \otimes \mu_2)(A) = \lim_n (\mu_1^n \otimes \mu_2^n)(A).$$

8.5 The Fubini-Lebesgue theorem

Theorem. *Suppose that $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ are measure spaces. Set $X = X_1 \times X_2$, $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$, and $\mu = \mu_1 \otimes \mu_2$, and let $(X, \mathcal{A}, \mu)$ be the product measure space. Suppose that*

(i) $$f \in \mathcal{L}^0(X, \mathcal{A}).$$

(ii) Then $f_{x_1} : x_2 \mapsto f(x_1, x_2)$ satisfies $f_{x_1} \in \mathcal{L}^0(X_2, \mathcal{A}_2)$ $\forall x_1 \in X_1$.

Now suppose that

(iii) $$f \in L^1_\mu(X, \mathcal{A}).$$

Then the following two properties are satisfied:

(iv) $$\begin{cases} f_{x_1} \in L^1_{\mu_2}(X_2, \mathcal{A}_2) \ \mu_1\text{-a.e. in } x_1, \text{ and} \\ k \in L^1_{\mu_1}(X_1, \mathcal{A}_1), \quad \text{where} \quad k(x_1) = \int_{X_2} f(x_1,x_2)d\mu_2(x_2). \end{cases}$$

(v) $$\begin{aligned}\int_X f(x_1,x_2)d\mu(x_1,x_2) &= \int_{X_1} d\mu_1(x_1) \left[\int_{X_2} f(x_1,x_2)d\mu_2(x_2)\right] \\ &= \int_{X_2} d\mu_2(x_2) \left[\int_{X_1} f(x_1,x_2)d\mu_1(x_1)\right].\end{aligned}$$

Conversely:

(vi) $$\begin{cases} \textit{Suppose that (i) holds, } f_{x_1} \in L^1_{\mu_2}(X_2, \mathcal{A}_2)\ \mu_1\textit{-a.e.,} \\ \textit{and there exists } k^* \in L^1_{\mu_1}(X_1, \mathcal{A}_1) \textit{ such that} \\ \quad \int_{X_2} |f(x_1,x_2)|d\mu_2(x_2) \le k^*(x_1). \end{cases}$$

Then (iii) is satisfied, and hence (iv) and (v).

REMARK. Denote the functions satisfying (ii) by Q and the functions satisfying (ii), (iii), (iv), and (v) by R. Then Q and R are vector spaces. Since the indicator functions of measurable sets are in Q by 8.3, so are finite linear combinations of indicator functions: $\mathcal{E}(X, \mathcal{A}) \subset Q$.

PROOF. First assume the following stronger hypothesis:

$$(i)' \qquad f \in \mathcal{L}^0(X, \mathcal{A}) \quad \text{and} \quad f \text{ is bounded.}$$

Then, by 6.4.1, f is the uniform limit of a sequence of simple functions φ^n:

$$\varphi^n = \sum_i \alpha_i^n \mathbf{1}_{A_i^n}.$$

By the remark, $\varphi^n \in Q$ for each n; that is,

$$(\varphi^n)_{x_1} \in \mathcal{L}^0(X_2, \mathcal{A}_2), \quad \forall x_1 \in X_1.$$

Since $(f)_{x_1} = \lim(\varphi_n)_{x_1}$, 2.5.1 shows that (i)$'$ $\Rightarrow$ (ii).

Similarly, using 6.8.1(iv), hypothesis (iii) can be replaced by this stronger hypothesis:

$$(iii)' \qquad f \text{ satisfies (i)}' \text{ and } \{x : f(x) \neq 0\} \subset A_1 \times A_2, \quad \text{with} \quad \mu_i(A_i) < +\infty.$$

Let $\{\varphi^n\}$ be a sequence of simple functions which converge uniformly to f and for which $\varphi^n(x) = 0$ if $x \notin A_1 \times A_2$. Then φ^n satisfies (iv) and (v).

Since $\varphi^n \to f$ uniformly, there exists a sequence $\{\epsilon_n\}$ such that $\epsilon_n \downarrow 0$ and

$$|f - \varphi^n| < \epsilon_n \mathbf{1}_{A_1 \times A_2}.$$

Thus

$$\int_X |f - \varphi^n| d\mu \leq \epsilon_n \int_{A_1 \times A_2} d\mu = \epsilon_n \mu_2(A_2).$$

Similarly,

$$\int_{X_2} |\varphi_{x_1}^n - f_{x_1}| d\mu_2 \leq \epsilon_n \int_{A_2} d\mu_2(x_2) = \epsilon_n \mu_2(A_2),$$

whence $\int \varphi_{x_1}^n d\mu_2$ converges uniformly to $\int f_{x_1} d\mu_2$. It follows from 2.5.3 that the left-hand side of the formula in (iv) is measurable. Repeating the same argument a third time for the integration in x_1 gives (iv) and (v). Summarizing, we have shown that (iii)$'$ $\Rightarrow$ (v).

Let $\{A_1^p\}$ and $\{A_2^p\}$ be exhaustion sequences for X_1 and X_2. Then $\{A^p\} = \{A_1^p \times A_2^p\}$ is an exhaustion sequence for X. Let T_p be the truncation operator defined in 6.7. Then $T_p(f)$ satisfies (iii)$'$.

Suppose now that (iii) holds and that

$$(vii) \qquad f \geq 0.$$

$\{T_p(f)\}$ is an increasing sequence of functions in L_μ^1 and $\|T_p(f)\|_{L^1} \leq \|f\|_{L^1}$. Since $T_p(f)$ satisfies (iii)$'$, (v) holds and

$$\int_{X_1} k_p(x_1) d\mu_1(x_1) = \|T_p f\|_{L^1}, \quad \text{where} \quad k_p(x_1) = \int_{X_2} (T_p(f))_{x_1} d\mu_2.$$

As the sequence $\{k_p\}$ is increasing, the Fatou-Beppo Levi theorem (7.1) applied to X_1 shows that

$$\lim k_p = k_\infty \in L^1_{\mu_i}(X_1, \mathcal{A}_1) \quad \text{and} \quad \int_{X_i} k_\infty = \lim \|T_p f\|_{L^1} = \|f\|_{L^1},$$

where the limit of the $k_p(x_1)$ is finite if $x_1 \notin B$ for some $B \in \mathcal{A}_1$, $\mu_1(B) = 0$.

Fix $x_1 \notin B$ and apply Fatou-Beppo Levi on the space X_2:

$$k_\infty(x_1) = \lim \int_{X_2} (T_p(f))_{x_1} d\mu_2 = \int_{X_2} \lim (T_p(f))_{x_1} d\mu_2 = \int_{X_2} (f)_{x_1} d\mu_2.$$

We have thus proved (iv) and (v) when f satisfies both (iii) and (vii). If (vii) is not satisfied, write $f = f^+ - f^-$; then f^+, $f^- \in R$, and by the remark $f \in R$.

It remains to prove the converse. Letting f satisfy (i), set $f^1 = |f|$. Using the truncation operator T_p, we have

$$\int (T_p f^1)_{x_1} d\mu_2 \le \int |f(x_1, x_2)| d\mu_2(x_2) \le k^*(x_2).$$

Moreover, since $T_p f^1 \in L^1_\mu$, we may use the identity (v) to obtain

$$\int_X T_p f^1 d\mu = \int d\mu_1 \int (T_p(f^1))_{x_1} d\mu_2 \le \int k^*(x_2) d\mu_2(x_2).$$

Hence the norm of $T_p(f^1)$ is bounded, with a bound independent of p, and the integrability criterion 7.4.1 implies that $f^1 \in L^1_\mu$. Since $f \in L^0(X, \mathcal{A})$, 7.4.3 implies (iii). □

9 The L^p Spaces

9.0 *Integration of complex-valued functions*

Let $f(x) = u(x) + iv(x)$ be a complex-valued function. Then f is a measurable mapping from X to $\mathbf{C}$ if and only if u and v are measurable. Furthermore, we say that f is *integrable* if u and v are integrable, and set

$$(i) \qquad \int f = \int u + i \int v.$$

The integral $f \mapsto \int f$ is a $\mathbf{C}$-linear functional on the space $L^1(X, \mathcal{A}, \mu; \mathbf{C})$ of complex-valued integrable functions. Moreover, setting

$$Z_f = \{x \in X : f(x) \neq 0)\},$$

$Z_f \in \mathcal{A}$ and the function arg $f(x)$ is well defined for $x \in Z_f$. The argument is defined to be zero on Z_f^c. Thus, if $f \in M_\mu(X, \mathcal{A}; \mathbf{C})$, we can write

$$(ii) \qquad f(x) = w(x) e^{i\theta(x)},$$

where $w \in M_\mu(X, \mathcal{A}; \mathbf{R}^+)$, $\theta \in M_\mu(X, \mathcal{A}; [0, 2\pi))$, and $|f(x)| = w(x)$.

(iii) Lemma. *Let f be a complex-valued integrable function. Then $|f|$ is integrable and*

$$\left|\int f\right| \leq \int |f|.$$

PROOF. $|f| \leq |u| + |v|$ and is thus dominated by two integrable functions, hence integrable. Set

$$\int f = r e^{i\varphi};$$

then

$$\left|\int f\right| = e^{-i\varphi}\int f = \int f e^{-i\varphi}.$$

Using the decomposition (ii),

$$\left|\int f\right| = Re \int w(x)\cos(\theta(x) - \varphi)d\mu(x).$$

Since $|\cos(\theta - \varphi)| \leq 1$, we obtain

$$\left|\int f\right| \leq \int w(x)d\mu(x). \square$$

NOTATION. The complex-valued integrable functions will be denoted by $L^1_\mu(X, \mathcal{A}; \mathbf{C})$.

9.1 **Definition.** Let $(X, \mathcal{A}, \mu)$ be a measure space. Let p be a real number, $1 \leq p < +\infty$.

Let

$$L^p_\mu(X, \mathcal{A}) = \{f \in L^0_\mu(X, \mathcal{A}) : |f|^p \in L^1_\mu(X, \mathcal{A})\}.$$

Set

$$\|f\|_{L^p} = \left(\int |f|^p\right)^{1/p}.$$

It is clear that $\|f\|_{L^p} = 0$ implies $f = 0$ and that $\|\alpha\|_{L^p} = |\alpha|\ \|f\|_{L^p}$ for every constant α.

Complex-valued functions with integrable pth power can be defined similarly:

$$L^p_\mu(X, \mathcal{A}; \mathbf{C}) = \{f \in M_\mu(X, \mathcal{A}; \mathbf{C}) : |f|^p \in L^1\}.$$

Writing $f = u + iv$ or $f = we^{i\theta}$, we obtain the equivalences

$$f \in L^p_\mu(X, \mathcal{A}; \mathbf{C}) \Leftrightarrow u \in L^p_\mu(X, \mathcal{A}) \quad \text{and} \quad v \in L^p_\mu(X, \mathcal{A})$$
$$f \in L^p_\mu(X, \mathcal{A}; \mathbf{C}) \Leftrightarrow w \in L^p_\mu(X, \mathcal{A}) \quad \text{and} \quad \theta \in L^0_\mu(X, \mathcal{A}).$$

9.2 Convexity inequalities

9.2.0 This section is devoted to proving the inequalities of Hölder and Minkowski. When $p = 2$ these inequalities become very easy. (Cf. Exercises, Cauchy-Schwarz inequality.)

9.2.1 **Definition.** A continuous function φ defined on $[a, b] \subset \mathbf{R}$ is called *convex* if $\varphi'_+(x) = \lim_{\epsilon \downarrow 0}(\varphi(x+\epsilon) - \varphi(x))\epsilon^{-1}$ exists $\forall x \in [a, b)$ and $\varphi'_+(x)$ is an increasing function. In particular, if φ is twice differentiable, then φ is convex if and only if $\varphi'' \geq 0$.

9.2.2 Lemma (Jensen's inequality). *Let φ be a convex function on $[a, b] \subset \mathbf{R}$. Let α_k $(1 \leq k \leq n)$ be positive numbers such that $\sum \alpha_k = 1$. Then*

$$\varphi\left(\sum_{k=1}^{n} \alpha_k t_k\right) \leq \sum_{k=1}^{n} \alpha_k \varphi(t_k) \quad \forall t_k \in [a, b].$$

REMARK. This inequality may be taken as a definition of convex functions.

PROOF. We prove the lemma for the case $n = 2$. Let a and b be constants and set

$$\widetilde{\varphi}(t) = \varphi(t) + at + b.$$

Then $\widetilde{\varphi}$ is convex. Choose a and b so that $\widetilde{\varphi}(t_1) = \widetilde{\varphi}(t_2) = 0$. Jensen's inequality reduces to showing that

$$\widetilde{\varphi}(t) \leq 0 \quad \text{for} \quad t_1 \leq t \leq t_2.$$

Otherwise the maximum of $\widetilde{\varphi}$ would be strictly positive and would be attained at a point $t_3 \in (t_1, t_2)$, and we would have

$$\widetilde{\varphi}'_+(t_3) = 0, \quad \widetilde{\varphi}(t_3) > 0.$$

Since $\widetilde{\varphi}'_+$ is increasing, $\widetilde{\varphi}'_+(t) \geq \widetilde{\varphi}'_+(t_3) = 0$ if $t \in [t_3, t_2)$, whence $\widetilde{\varphi}(t_2) \geq \widetilde{\varphi}(t_3)$, a contradiction. We proceed by induction on n. Assuming that the inequality holds for $n \leq p$, we prove it for $n = p + 1$.

Set $\xi = \beta^{-1}\left(\sum_{n=1}^{p} \alpha_i t_i\right)$, where $\beta = \sum_{i=1}^{p} \alpha_i$.

Then, by the result for $n = 2$, $\varphi(\beta\xi + \alpha_{p+1}t_{p+1}) \leq \beta\varphi(\xi) + \alpha_{p+1}\varphi(t_{p+1})$.

The first term on the right-hand side can be bounded above by using the induction hypothesis, which gives $\varphi(\xi) \leq \sum_{i=1}^{p} \beta^{-1}\alpha_i \varphi(t_i)$. □

9.2.3 Corollary. *Let ξ_1, $\xi_2 \geq 0$ and let α, $\beta > 0$ satisfy $\alpha + \beta = 1$. Then $\xi_1^\alpha \xi_2^\beta \leq \alpha\xi_1 + \beta\xi_2$.*

PROOF. If $\xi_1 = 0$, the left-hand side is zero and the inequality is obvious. Suppose that $\xi_i > 0$ $(i = 1, 2)$, and set $\eta_i = \log \xi_i$. The exponential function $\exp(t)$ satisfies the hypotheses of 9.2.1, whence

$$\exp(\alpha\eta_1 + \beta\eta_2) \leq \alpha\exp(\eta_1) + \beta\exp(\eta_2). \square$$

9.2.4 Lemma. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let α, $\beta > 0$ be such that $\alpha + \beta = 1$, and let f and g be nonnegative functions in $L^1_\mu(X, \mathcal{A})$. Then*

(i) $f^\alpha g^\beta \in L^1_\mu(X,\mathcal{A})$ *and*
(ii) $\int f^\alpha g^\beta \leq \left(\int f\right)^\alpha \left(\int g\right)^\beta$.

PROOF. If $f = 0$ a.e., both sides of the inequality (ii) are zero. Hence we may assume that $\|f\|_{L^1_\mu} > 0$ and $\|g\|_{L^1_\mu} > 0$. Setting

$$\widetilde{f} = \|f\|_{L^1}^{-1} f, \quad \widetilde{g} = \|g\|_{L^1}^{-1} g,$$

we reduce the proof of (ii) to showing that

$$\int \widetilde{f}^\alpha \widetilde{g}^\beta \leq 1.$$

We will use 9.2.3. For every x, $\widetilde{f}^\alpha(x)\widetilde{g}^\beta(x) \leq \alpha \widetilde{f}(x) + \beta \widetilde{g}(x)$.

The right-hand side is an integrable function; hence (i) follows from 7.4.3. Integrating both sides of this inequality gives

$$\int f^\alpha g^\beta \leq \alpha \int \widetilde{f} + \beta \int \widetilde{g}.$$

Since $\int \widetilde{f} = \int \widetilde{g} = 1$,

$$\int f^\alpha g^\beta \leq \alpha + \beta = 1. \square$$

9.2.5 Definition of conjugate exponents

Definition. Let $1 < p < +\infty$ and $1 < q < +\infty$. We say that p and q are *conjugate exponents* if

$$\frac{1}{p} + \frac{1}{q} = 1.$$

REMARKS. p is conjugate to itself if and only if $p = 2$.

If $1 < p < 2$, then $q > 2$.

9.2.6 Theorem (Hölder's inequality). *Let $(X, \mathcal{A}, \mu)$ be a measure space, let p and q be conjugate exponents, and let $f \in L^p$, $g \in L^q$. Then*

(i) $fg \in L^1$ *and*
(ii) $\left|\int fg\right| \leq \|f\|_{L^p}\|g\|_{L^q}$.

PROOF. Since the theorem is clear when $p = \infty$ or $q = \infty$, we may assume that $1 < p < \infty$. We first consider the case where f and g are nonnegative.

Set $u = f^p$, $v = g^q$, $\alpha = \frac{1}{p}$, $\beta = \frac{1}{q}$. Then $fg = u^\alpha v^\beta$, and applying 9.2.3 gives the theorem.

In the general case, set $|f| = f_1$, $|g| = g_1$. Then $f_1 g_1 \in L^1$ by the argument above; hence by 7.4.2 $fg \in L^1$ and

$$\left|\int fg\right| \leq \int f_1 g_1. \square$$

9.2.7 Theorem (Minkowski's inequality). *Let $(X, \mathcal{A}, \mu)$ be a measure space and let f, $g \in L^p$, where $1 \leq p < +\infty$. Then*

(i) $(f+g) \in L^p$ *and*
(ii) $\|f+g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}$.

PROOF. The theorem is true for $p = 1$ by Proposition 6.8.3. Note that the function $\varphi(t) = t^p$ is convex on $[0, +\infty)$. Using Jensen's inequality, we have

$$\left(\frac{\xi_1 + \xi_2}{2}\right)^p \leq \frac{1}{2}\xi_1^p + \frac{1}{2}\xi_2^p,$$

whence

$$|f(x) + g(x)|^p \leq (|f(x)| + |g(x)|)^p \leq 2^{p-1}|f(x)|^p + 2^{p-1}|g(x)|^p.$$

Hence the integrability criterion 7.4.3 implies (i). It suffices to prove (ii) in the case that f and $g \geq 0$. We then have

$$\int (f+g)^p = \int f(f+g)^{p-1} + \int g(f+g)^{p-1}.$$

Letting q be the conjugate exponent and using Hölder,

$$\int f(f+g)^{p-1} \leq \left(\int f^p\right)^{1/p} \left(\int (f+g)^{(p-1)q}\right)^{1/q};$$

but, since p and q are conjugate, $p+q = pq$, or $(p-1)q = p$. Writing the analogous integral for g, we obtain

$$\int (f+g)^p \leq \left[\left(\int f^p\right)^{1/p} + \left(\int g^p\right)^{1/p}\right] \left(\int (f+g)^p\right)^{1/q},$$

or

$$\|f+g\|_{L^p}^p \leq (\|f\|_{L^p} + \|g\|_{L^p})\|f+g\|_{L^p}^{p/q}.$$

If $\|f+g\|_{L^p} = 0$, Minkowski's inequality holds trivially. Otherwise we can divide both sides by $\|f+g\|_{L^p}^{p/q}$ to obtain

$$\|f+g\|_{L^p}^{p-p/q} \leq \|f\|_{L^p} + \|g\|_{L^p},$$

and the conjugacy relation gives $p - \frac{p}{q} = 1 - p\left[1 - \frac{1}{p}\right] = 1$. □

Remark. Writing $f(x) = w(x)e^{i\theta(x)}$ shows that the Hölder and Minkowski inequalities remain true for complex-valued functions.

9.2.8 Theorem. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $1 \le p < +\infty$. Then $L^p(X, \mathcal{A}, \mu)$ is a vector space on which a norm is defined by the function $f \mapsto \|f\|_{L^p}$.*

Proof. It follows from 9.2.7(i) that L^p is a vector space. Moreover, 9.2.7(ii) and 9.1 show that $\|\cdot\|_{L^p}$ is a norm.

9.3 Completeness theorem. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $1 \le p < +\infty$. Then $L^p_\mu(X, \mathcal{A})$ is a complete normed space.*

Remark. For $p = 1$, Lebesgue's theorem (7.2) implies that every normally convergent series in L^1 is convergent, and hence that L^1 is complete.

Proof. We proceed as in 6.5.4(v) by proving the following lemma:

9.3.1 Lemma. *Let $\{f_n\}$ be a Cauchy sequence in L^p_μ. Then $\{f_n\}$ converges in measure.*

Proof. Fixing ϵ, set

$$\{x : |f_n(x) - f_{n'}(x)| > \epsilon\} = A_{n,n'}.$$

Then

$$\int |f_n - f_{n'}|^p \ge \int_{A_{n,n'}} |f_n - f_{n'}|^p \ge \epsilon^p \mu(A_{n,n'}),$$

implying the Chebyshev-type inequality

$$\mu(A_{n,n'}) \le \epsilon^{-p} \|f_n - f_{n'}\|^p_{L^p}.$$

Fix n_0 such that $\|f_n - f_{n'}\|_{L^p} < \epsilon^{1+p^{-1}}$ if $n, n' > n_0$. It follows that $e(f_n, f_{n'}) < 2\epsilon$ if $n, n' > n_0$. □

9.3.2 Proof of the theorem. Since L^0_μ is a complete space, $\{f_n\}$ converges in measure (by 5.2.9) to f_0. By 5.2.7, we can extract a subsequence such that

(i) f_{n_k} converges to f_0 μ-a.e.

Since f_n is a Cauchy sequence in L^p, we have

(ii) $\|f_n\|_{L^p} < C$, or $\|\, |f_n|^p \,\|_{L^1} < C$.

By Fatou's lemma (7.7), $|f_0|^p \in L^1$.

Fixing k, consider the sequence $\{u_s\} = \{|f_{n_s} - f_{n_k}|^p\}$. Fatou's lemma can be applied since u_s converges a.e. to $|f_0 - f_{n_k}|^p$. We obtain

$$\|\lim u_s\|_{L^1} \le \sup \|u_s\|_{L^1}.$$

Fix m_0 such that $\|f_n - f_{n'}\|_{L^p} < \epsilon$ if $n, n' \geq m_0$. Take k such that $n_k \geq m_0$; then

$$\|f_0 - f_{n_k}\|_{L^p} \leq \epsilon \quad \text{if} \quad n > m_0$$

and

$$\|f_0 - f_n\|_{L^p} \leq \|f_0 - f_{n_k}\|_{L^p} + \|f_{n_k} - f_n\|_{L^p} < 2\epsilon. \square$$

REMARK. Writing $f = u + iv$, we see that 9.3 implies that $L^p_\mu(X, \mathcal{A}; \mathbf{C})$ is complete.

9.4 Notions of duality

Given a normed vector space E, the vector space E' of continuous linear functionals l on E is called the *dual* of E. For $l \in E'$, we set

$$\|l\| = \sup |l(x)| \quad \text{where} \quad \|x\| \leq 1, \quad x \in E.$$

It can be shown that E' is a Banach space.

9.4.1 Theorem. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Then $L^2_\mu(X, \mathcal{A})$ is a Hilbert space when the scalar product is defined by*

$$\text{(i)} \qquad \int fg = (f|g).$$

The scalar product for the complex-valued functions $L^2_\mu(X, \mathcal{A}, \mu)$ is defined by $\int f\overline{g} = (f|g)$.

PROOF. $(f|f) = \|f\|^2_{L^2}$, and Hölder's inequality becomes

$$\text{(ii)} \qquad |(f|g)| \leq \|f\|_{L^2}\|g\|_{L^2}.$$

This is just the Cauchy-Schwarz inequality, which can be proved directly. Moreover, L^2_μ is complete, and hence is a Hilbert space.

9.4.2 Corollary. *The dual of the space L^2 can be identified with L^2; the dual pairing is given by 9.4.1(i).*

PROOF. In a Hilbert space, by Riesz's theorem[1] every continuous linear functional can be expressed by a scalar product.

9.4.3 Proposition. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let p and q be conjugate exponents, $1 < p < +\infty$. Then there is an isometric injection u from L^q_μ into $(L^p_\mu)'$.*

---

[1]See, for example, W. Rudin, *Real and Complex Analysis* (New York: McGraw-Hill, 1974).

PROOF. Define a mapping $u : L^q_\mu \to (L^p_\mu)'$ by associating with $g \in L^q_\mu$ the linear functional

$$l_g(f) = \int fg.$$

Then by Hölder

$$(i) \qquad |l_g(f)| \le \|g\|_{L^q}\|f\|_{L^p} = C\|f\|_{L^p},$$

which shows both that l_q is a continuous linear functional and that u is a contraction:

$$\|u(g)\|_{(L^p)'} \le \|g\|_{L^q}.$$

In order to show that u is an isometry,we introduce $f_0 = (\text{signum}(g))|g|^{q/p}$. Then $|f_0|^p = |g|^q$, $\|f_0\|^p_{L^p} = \|g\|^q_{L^q}$, and

$$\int f_0 g = \int |g|^{q/p+1} = \int |g|^q = \|g\|^q_{L^q}.$$

Hence

$$l_g(f_0) = \|g\|^q_{L^q}.$$

Furthermore,

$$|l_g(f_0)| \le \|l_g\|_{(L^p)'}\|f_0\|_{L^p} = \|l_g\|_{(L^p)'}\|g\|^{q/p}_{L^q},$$

whence

$$\|l_g\|_{(L^p)'} \ge \|g\|^{q-q/p}_{L^q}.$$

But $q - q/p = 1$, and hence u *is an isometry.*

It follows that u is an *injective* mapping of L^q into $(L^p)'$. □

REMARK. It will be shown in Section IV.6 that u is surjective, and thus identifies $(L^p)'$ with L^q $(1 < p < +\infty)$.

9.5 The space L^∞

9.5.1 **Definition.** $\overline{f} \in L^0_\mu(X, \mathcal{A})$ is said to be *essentially bounded* if there exists a *bounded* representative f of $\overline{f}$. The space of essentially bounded measurable functions is denoted by $L^\infty_\mu(X, \mathcal{A})$. We define $A_{g,\xi} = \{x : |g(x)| > \xi\}$ and $K(g) = \{\xi \in \mathbf{R}^+ : \mu(A_{g,\xi}) = 0\}$.

If $g \in L^\infty_\mu$, then $K(g) \ne \emptyset$ and we set

$$\|g\|_{L^\infty_\mu} = \inf K(g).$$

9.5.2 Lemma. $\mu(A_{g,\xi}) > 0$ *if and only if* $\xi < \|g\|_{L^\infty_\mu}$.

PROOF. The only case that is not obvious occurs when $\xi = \|g\|_{L^\infty}$. We then apply the continuity theorem for increasing sequences of measurable sets.

Setting $\xi_0 = \|g\|_{L^\infty}$ gives $\mu\left(\cup_{n=1}^{\infty} A_{g,\xi_0+n^{-1}}\right) = 0$. But $\cup_{n=1}^{\infty} A_{g,\xi_0+n^{-1}} = A_{g,\xi_0}$. □

9.5.3 Proposition. *$L^\infty_\mu(X, \mathcal{A}, \mu)$ is a complete normed vector space.*

PROOF. We first prove the triangle inequality for $\|\cdot\|_{L^\infty}$. Let $f, g \in L^\infty_\mu$ and set $h = f+g$; then $|h(x)| \leq |f(x)|+|g(x)|$ implies that $A^c_{h,\xi+\eta} \supset A^c_{f,\xi} \cap A^c_{g,\eta}$. Taking complements, we obtain

$$\mu(A_{h,\xi+\eta}) \leq \mu(A_{f,\xi}) + \mu(A_{g,\eta}),$$

or $(\xi + \eta) \in K(h)$ if $\xi \in K(f)$ and $\eta \in K(g)$. Thus

$$\|h\|_{L^\infty_\mu} \leq \|f\|_{L^\infty_\mu} + \|g\|_{L^\infty_\mu}.$$

If $\|h\|_{L^\infty_\mu} = 0$, then $h(x) = 0$ a.e. by 9.5.2, and hence $\|\cdot\|_{L^\infty_\mu}$ is *a norm.*

Let f_n be a Cauchy sequence in the norm $\|\cdot\|_{L^\infty_\mu}$. Choose representatives f_n of the class $\overline{f}_n$ and set $u_{n,n'} = f_n - f_{n'}$. Let $A_{n,n'} = \{x : |u_{n,n'}(x)| > 3\|u_{n,n'}\|_{L^\infty_\mu}\}$; then, by the definition, $\mu(A_{n,n'}) = 0$.

Set $Z = \cup_{n,n'} A_{n,n'}$. Then $\mu(Z) = 0$ and

$$|f_n(x) - f_{n'}(x)| \leq 3\|f_n - f_{n'}\|_{L^\infty_\mu} \quad \text{if} \quad x \in Z^c.$$

The sequence f_n converges uniformly on Z^c. Set $f_0(x) = \lim f_n(x)$ if $x \in Z^c$ and $f_0(x) = 0$ if $x \in Z$. Then $f_0 \in L^\infty_\mu$ and $\|f_n - f_0\|_{L^\infty} \to 0$. □

9.6 Proposition. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Suppose that $\mu(X) < +\infty$. Then $L^p_\mu(X, \mathcal{A}) \supset L^{p'}_\mu(X, \mathcal{A})$ if $1 \leq p < p' \leq +\infty$.*

PROOF. Use Hölder's inequality to write

$$\int_X |f|^p d\mu(x) = \int |f|^p \mathbf{1}_X d_\mu(x) \leq \| \ |f|^p \ \|_{L^r_\mu} \|\mathbf{1}_X\|_{L^s_\mu},$$

where r and s are conjugate exponents. If $p' < +\infty$, note that

$$\| \ |f|^p \ \|_{L^r} = \left(\int |f|^{rp}\right)^{1/r}$$

and take $r = \dfrac{p'}{p} > 1$. Then

$$(i) \qquad \|f\|_{L^p} \leq [\mu(X)]^\alpha \|f\|_{L^{p'}}, \quad \text{where} \quad \alpha = \frac{p' - p}{p'}.$$

This shows that every function in $L^{p'}$ is in L^p. If $p' = \infty$, note that

$$\int |f|^p \leq \|f\|^p_{L^\infty} \mu(X). \square$$

II

Borel Measures and Radon Measures

Introduction

The preceding chapter dealt with abstract measure theory; given an abstract set X, we rather arbitrarily prescribed the σ-algebra $\mathcal{B}$ of its measurable subsets. In this chapter, we work in a space X which is locally compact and can be written as a countable union of compact sets. A natural σ-algebra in this context is the Borel algebra $\mathcal{B}_X$. A *locally finite Borel measure* is a measure defined on $\mathcal{B}_X$ such that every compact set has finite measure. For X metrizable, we prove *Lusin's theorem*: If μ is a locally finite Borel measure and $A \in \mathcal{B}_X$, then for every $\epsilon > 0$ there exist an open set O and a closed set F such that $F \subset A \subset O$ and $\mu(O - F) < \epsilon$. Thus an arbitrary Borel set can be approximated to within ϵ by both an open and a closed set.

A natural vector space on X is the space $C_K(X)$ of continuous functions with compact support. A linear functional I on $C_K(X)$ is called *positive* if $I(f) \geq 0$ for every nonnegative function f. We prove the *Radon-Riesz theorem*, which constructs a bijection between the positive linear functionals on $C_K(X)$ and the locally finite Borel measures. In the Prologue, we showed that the Riemann integral on $\mathbf{R}$ defines a positive linear functional on $C_K(\mathbf{R})$. In this chapter, we apply the Radon-Riesz theorem to obtain a canonical translation-invariant Borel measure on $\mathbf{R}$, the *Lebesgue measure*. The theory of the Lebesgue integral appears as a special case of the theory of the abstract integral developed in Chapter I. We obtain the Lebesgue integral on $\mathbf{R}^n$ by constructing the product measure, and prove the change-of-variables formula for multiple integrals.

When Y is compact, the space of continuous functions on Y is a Banach space. We consider the dual vector space $(C(Y))^*$ of continuous linear functionals on Y, and show that every linear functional can be written as the *difference* of two positive linear functionals. This leads us to the concept of signed Radon measures.

Given a locally compact space X, we consider the Banach space $C_b(X)$ of bounded continuous functions on X and the closed subspace $C_0(X)$ of functions which vanish at infinity. $(C_0(X))^*$ is identified with the space $M^1(X)$ of finite signed Radon measures. Three topologies can be defined on this set by using the pairings with $C_K(X)$, $C_0(X)$, and $C_b(X)$. We compare the three corresponding notions of convergence.

The first section of this chapter is devoted to the construction of *partitions of unity*, which allow the passage from local to global considerations on X. It is purely topological, while the rest of the chapter describes measure theory on locally compact spaces.

1 Locally Compact Spaces and Partitions of Unity

1.0 Definition of locally compact spaces which are countable at infinity

Let X be a Hausdorff topological space which satisfies the following hypotheses:

1.0.1 X is *locally compact*, i.e. every point $x_0 \in X$ has a compact neighborhood.

1.0.2 X is *countable at infinity*, i.e. there exists a sequence $\{K_n\}$ of compact subsets of X such that

$$K_n \subset K_{n+1} \quad \text{and} \quad \bigcup_n K_n = X.$$

1.0.3 Proposition. *There exists a sequence H_m of compact sets such that*

$$H_m \subset \overset{\circ}{H}_{m+1} \quad (\textit{where } \overset{\circ}{A} \textit{ denotes the interior of } A)$$

and

$$\bigcup_{m=1}^{\infty} H_m = X.$$

PROOF. The proof is by induction. Set $H_1 = K_1$ and, assuming that H_q has been constructed, set $G_q = H_q \cup K_q$. Each $x \in G_q$ has a compact neighborhood $V(x)$; from the open cover of G_q formed by $\{\overset{\circ}{V}(x)\}$, extract a finite subcover.

This procedure gives points $x_{q,j} \in G_q$, $1 \leq j \leq m_q$, such that $H_q \subset \bigcup_{1\leq j\leq m_q} \overset{\circ}{V}(x_{q,j})$. Set $H_{q+1} = \bigcup_{1\leq j\leq m_q} V(x_{q,j})$. As the finite union of compact sets, H_{q+1} is compact. Furthermore,

$$\overset{\circ}{H}_{q+1} \supset \bigcup_{1\leq j\leq m_q} \overset{\circ}{V}(x_{q,j}) \supset H_q$$

and

$$\bigcup H_q \supset \bigcup K_q = X.\square$$

1.1 Urysohn's lemma

Lemma. *Let F_1 and F_2 be disjoint closed subsets of a locally compact space X. Then there exists a continuous function f on X such that*

$$\begin{array}{lll} f(x) = 1 & \textit{if and only if} & x \in F_1; \\ f(x) = 0 & \textit{if and only if} & x \in F_2; \\ 0 \leq f(x) \leq 1 & \textit{for all} & x \in X. \end{array}$$

PROOF. We restrict the proof to the relatively trivial special case where X is a *metric* space.

Let

$$f_i(x) = d(x, F_i) = \min(d(x, y_i)), \quad \text{where} \quad y_i \in F_i.$$

Then f_i $(i = 1, 2)$ is a positive continuous function and $f_i(x) = 0 \Leftrightarrow x \in F_i$.

Let a function Φ be defined on $Z = ([0, +\infty) \times [0, +\infty)) - (0, 0)$ by setting

$$\Phi(\xi, \eta) = \frac{\xi}{\xi + \eta}.$$

Then Φ is continuous since $(0, 0)$ is not in the domain of definition Z of Φ. Furthermore,

$$\begin{array}{l} 0 \leq \Phi \leq 1, \\ \Phi(\xi, 0) = 1 \quad \text{if} \quad \xi > 0, \text{ and} \\ \Phi(0, \eta) = 0 \quad \text{if} \quad \eta > 0. \end{array}$$

Let $f(x) = \Phi(f_1(x), f_2(x))$. Since $F_1 \cap F_2 = \emptyset$, the mapping into $(\mathbf{R}^+)^2$ defined by $x \mapsto (f_1(x), f_2(x))$ actually maps into Z. Thus f is the composition of continuous mappings and hence is continuous. $\square$

1.2 Support of a function

Definition. Let f be a continuous function on X. The support of f, denoted by supp(f), is the closed set

$$\text{supp}(f) = \text{closure}\,\{x : f(x) \neq 0\}.$$

1.2.1 Proposition. *The following statements are equivalent:*

(i) $z \notin supp\,(f)$.
(ii) *There exists a neighborhood* $V(z)$ *such that* $f(x) = 0\ \forall x \in V(z)$.

PROOF. Let $O = (\text{supp}\,(f))^c$; then O is an *open* set and

$$\{x : f(x) \neq 0\} \cap O \subset \text{supp}\,(f) \cap O = \emptyset,$$

whence, setting $O = V(z)$, we have shown [that] (ii) $\Rightarrow$ (i).
Conversely, if $V(z) \cap \{x : f(x) \neq 0\} = \emptyset$, then

$$\overset{\circ}{V}(z) \cap \overline{\{x : f(x) \neq 0\}} = \emptyset. \square$$

1.2.2 Proposition. *Suppose that* X *is a locally compact space,* F *is a closed subset of* X*, and* O *is an open subset of* X *such that* $F \subset O$*. Then there exists a continuous function* g *such that*

$$\begin{array}{lll} 0 \leq g(x) \leq 1 & \textit{for any} & x \in X; \\ g(x) = 1 & \textit{if and only if} & x \in F; \quad \textit{and} \\ supp\,(g) \subset O. & & \end{array}$$

PROOF. Set $F' = O^c$. Applying Urysohn's lemma (1.1), let f be the function associated with the pair of closed sets (F, F'). Set

$$F'' = f^{-1}([0, \frac{1}{2}]).$$

Then F'' is a *closed set* since f is a continuous function. Let g be the function associated by Urysohn's lemma with the pair (F, F''). Then $g(x) > 0$ implies $x \notin F''$, or $f(x) > \frac{1}{2}$, which may be written as

$$\{x : g(x) \neq 0\} \subset f^{-1}((\frac{1}{2}, 1]).$$

Hence supp $(g) \subset$ closure $(f^{-1})((\frac{1}{2}, 1])$.
Since $f^{-1}([\frac{1}{2}, 1])$ is closed, we have a fortiori

$$\text{supp}\,(g) \subset f^{-1}([\frac{1}{2}, 1]) \subset O. \square$$

1.3 Subordinate covers

1.3.0 **Definition.** Let $\{U_\alpha\}$ be an open cover of X. An open cover $\{V_n\}$ is said to be subordinate to $\{U_\alpha\}$ if, for any n, there exists $\alpha(n)$ such that

$$\overline{V}_n \subset U_{\alpha(n)}.$$

A cover $\{H_\gamma\}$ is said to be *locally finite* if, for every compact set K,

$$\text{card}\ \{\gamma : H_\gamma \cap K \neq \emptyset\} \quad \text{is finite.}$$

1.3.1 Theorem. *Let* X *be a locally compact space which is countable at infinity. Then every open cover has a locally finite subordinate open cover* $\{V_n\}$ *such that the* $\overline{V}_n$ *are compact.*

PROOF. Let $\{U_\alpha\}$ be an open cover of X and let $\{H_m\}$ be the sequence of compact sets defined in 1.0.3. Set

$$G_1 = \overline{H}_1 \quad \text{and} \quad G_m = \overline{(H_m - H_{m-1})}.$$

Then

$$G_m = \overline{H_m \cap (H_{m-1})^c} \subset H_m \cap \overline{H^c_{m-1}}.$$

But

$$\left(\overset{\circ}{H}_{m-1}\right)^c \supset H^c_{m-1}, \quad \text{whence} \quad \left(\overset{\circ}{H}_{m-1}\right)^c \supset \overline{H^c_{m-1}},$$

so that

$$G_m \subset H_m \cap \left(\overset{\circ}{H}_{m-1}\right)^c,$$

and thus $G_m \cap \overset{\circ}{H}_{m-1} = \emptyset$. Using 1.0.3,

$$(i) \qquad G_m \cap H_{m-2} = \emptyset.$$

Set

$$(ii) \qquad U_{\alpha,m} = U_\alpha \cap \overset{\circ}{H}_{m+1} \cap H^c_{m-2}.$$

Then $U_{\alpha,m}$ is an open cover of G_m.

For each $x \in G_m$, there is an open set $W_m(x)$ such that

$$(iii) \qquad \overline{W_m(x)} \subset U_{\alpha,m} \quad \text{where} \quad \alpha = \alpha(x).$$

The $W_m(x)$ form an open cover of the compact set G_m; from this cover we can extract a finite subcover, say $W_m(x_1), \ldots, W_m(x_j)$.

The family $\{W_m(x_k)\}$ is a countable family of open sets, which we denote by $\{V_n\}$. We have $\overline{V}_n \subset U_\alpha$, where $\alpha = \alpha(n)$. The $\{V_n\}$ cover G_m for every m, hence cover X. For fixed m, (i), (ii), and (iii) imply

$$(iv) \qquad \operatorname{card}\{n : V_n \cap G_m \neq \emptyset\} < +\infty.$$

We now prove a lemma.

1.3.2 Lemma. *Let K be a compact subset of X. Then there exists q such that $K \subset \overset{\circ}{H}_q$.*

PROOF. Set $F_r = \left(\overset{\circ}{H}_r\right)^c \cap K$; then $\cap_r F_r = \emptyset$.

The F_r form a decreasing sequence of closed subsets of the compact set K. Since their intersection is empty, there exists q such that

$$\emptyset = F_q = \left(\overset{\circ}{H}_q\right)^c \cap K. \square$$

1.3.3 CONCLUSION OF THE PROOF OF THEOREM 1.3.1. Given the compact set K, let q be determined by 1.3.2. Then (ii) and (iii) show that

$$W_m(x) \cap K = \emptyset \quad \text{if} \quad m \geq q - 2,$$

whence

$$\operatorname{card}\{n : V_n \cap K \neq \emptyset\} < +\infty. \square$$

1.4 Partitions of unity

1.4.0 **Definition.** A *partition of unity* on the space X is a sequence of continuous functions φ_n such that

(i) $0 \le \varphi_n \le 1$,
(ii) $\operatorname{supp}(\varphi_n)$is compact,
(iii) $\operatorname{card}\{n : K \cap \operatorname{supp}(\varphi_n) \ne \emptyset\} < +\infty$ for every compact set K, and
(iv) $\sum \varphi_n(x) = 1$.

REMARK. Condition (iii) is called a local finiteness condition. It implies that, for fixed x, the series (iv) contains only finitely many nonzero terms.

The partition of unity is said to be *subordinate* to the open cover U_α if

(v) for every n, there exists $\alpha(n)$ such that $\operatorname{supp}(\varphi_n) \subset U_{\alpha(n)}$.

1.4.1 Theorem. *Suppose that X is a locally compact space which is countable at infinity and $\{U_\alpha\}$ is an open cover of X. Then there exists a partition of unity subordinate to $\{U_\alpha\}$.*

PROOF. Let $\{V_n\}$ be the locally finite cover subordinate to $\{U_\alpha\}$ constructed in Theorem 1.3.1.

Since the V_n form a cover, by another application of 1.3.1 there is a locally finite cover $\{L_s\}$ subordinate to $\{V_n\}$ which satisfies

$$\overline{L}_s \subset V_n, \quad \text{where} \quad n = n(s).$$

Applying 1.2.2 to the pair $(\overline{L}_s, V_{n(s)})$, there is a function g_s such that $\operatorname{supp}(g_s) \subset V_n$ and $g_s(x) = 1$ if $x \in \overline{L}_s$. Since each $\overline{V}_n$ is compact and the cover $\{L_s\}$ is locally finite, only finitely many of the elements L_s are contained in any V_n. Since the cover $\{V_n\}$ is locally finite,

$$\operatorname{card}\{n : V_n \cap K \ne \emptyset\} < +\infty$$

for any compact set K. Hence, setting $I(K) = \{s : \operatorname{supp}(g_s) \cap K \ne \emptyset\}$, we obtain

$$\operatorname{card}(I(K)) < +\infty.$$

Thus the sequence $\{g_s\}$ satisfies condition (iii). Set

$$D(x) = \sum_s g_s(x).$$

To calculate $D(x)$ on a given compact set K, it suffices to let s range over $I(K)$. As this set is finite, $D(x)$ can be written on K as a sum of continuous functions; hence $D(x)$ is continuous on K. Together with the local compactness of the space X, this implies that D is continuous.

Furthermore, $\{L_s\}$ covers X. For every x, there exists s such that $x \in L_s$; that is,

$$D(x) \ge 1 \quad \text{for every} \quad x \in X.$$

Setting

$$\beta(x) = \frac{1}{D(x)}$$

gives a continuous function on X. Finally, set $\varphi_n = \beta g_n$. □

2 Positive Linear Functionals on $C_K(X)$ and Positive Radon Measures

2.0.1 Notation

Given a locally compact space X, $C_K(X)$ denotes the vector space of continuous functions with *compact support.* We write

$$f \geq 0 \quad \text{if} \quad f(x) \geq 0 \quad \text{for every } x.$$

2.0.2 **Definition.** A *positive linear functional* is a linear mapping $I : C_K(X) \to \mathbf{R}$ such that $I(f) \geq 0$ for every $f \geq 0$.

2.1 Borel measures

Let $\mathcal{B}_X$ denote the Borel algebra on X. A measure defined on $\mathcal{B}_X$ is called a *Borel measure*, and is said to be *locally finite* if

2.1.1 $$\mu(K) < +\infty \quad \text{for every compact set } K.$$

REMARK. Since K is closed, $K \in \mathcal{B}_X$.

2.1.2 Proposition. *Let μ be a locally finite Borel measure on X. Then every continuous function with compact support is integrable. Setting*

$$I_\mu(f) = \int f d\mu$$

defines a positive linear functional on $C_K(X)$.

PROOF. Since f is continuous, it is $\mathcal{B}_X$-measurable. Furthermore, $|f|$ is bounded by a constant M, and setting $K = \operatorname{supp}(f)$ yields

$$|f| \leq M\mathbf{1}_K.$$

By 2.1.1, $\mathbf{1}_K$ is integrable; by I-7.4.2, so is f. The positivity of I follows from I-6.8.3. □

2.2 Fundamental theorem of Radon-Riesz. *Let X be a metrizable locally compact space which is countable at infinity. Then the correspondence*

$$\mu \to I_\mu$$

of 2.1.2 defines a bijection which allows the locally finite Borel measures to be identified with the positive linear functionals on $C_K(X)$.

PROOF. This statement contains both an *existence* and a *uniqueness* theorem: Every positive linear functional is represented by an integral with respect to a locally compact Borel measure, and this representation is unique.

The proof of Theorem 2.2 occupies the rest of this section.

2.2.1 Approximation lemma. *Let X satisfy the hypotheses of 2.2. Then for every open set O in X there is an increasing sequence of compact sets K_n such that*

$$(i) \qquad O = \cup K_n \quad \text{and} \quad K_n \subset \overset{\circ}{K}_{n+1}.$$

For every compact set K in X, there is a decreasing sequence of open sets O_n such that $\overline{O}_n$ is compact,

$$(ii) \qquad K = \cap O_n, \quad \text{and} \quad \overline{O}_n \subset O_{n-1}.$$

PROOF. Set $G_n = \{x : d(x, O^c) \geq \frac{1}{n}\}$; then G_n is closed. Let $K_n = G_n \cap H_n$, where $\{H_n\}$ is the sequence of compact sets of 1.0.3.

Then $\overset{\circ}{K}_n \supset \overset{\circ}{G}_n \cap \overset{\circ}{H}_n \supset G_{n-1} \cap H_{n-1}$, and (i) is satisfied.

To prove (ii), let m be determined as in 1.3.2 so that $K \subset \overset{\circ}{H}_m$, and set

$$O_n = \overset{\circ}{H}_m \cap \left\{ x : d(x, K) < \frac{1}{n} \right\}. \square$$

2.3 Proof of uniqueness of the Riesz representation

Let μ and ν be locally finite Borel measures such that

$$2.3.0. \qquad \int f(x) d\mu(x) = \int f(x) d\nu(x), \quad \forall f \in C_K(X).$$

2.3.1 Proposition. *Suppose that 2.3.0 is satisfied. Then the measures μ and ν coincide on open sets and on sets which can be written as the intersection of an open set and a closed set.*

PROOF. Using the approximation lemma 2.2.1(i), we can write $O = \cup K_n$. For every pair $(K_n, \overset{\circ}{K}_{n+1})$, let g_n be determined as in 1.2.2:

$$\begin{array}{ll} g_n(x) = 1 & \text{if} \quad x \in K_n, \\ \operatorname{supp}(g_n) \subset K_{n+1}, & \text{and} \quad 0 \leq g_n \leq 1. \end{array}$$

Then $\mathbf{1}_{K_n} \leq g_n \leq \mathbf{1}_O$, whence

$$\mu(K_n) \leq \int g_n d\mu \leq \mu(O).$$

Since, by I-3.2.1,

$$\lim \mu(K_n) = \mu(O) \leq +\infty,$$

it follows that

$$(i) \qquad \lim \int g_n d\mu = \mu(O) \quad \text{and similarly} \quad \lim \int g_n d\nu = \nu(O).$$

Since $g_n \in C_K(X)$, 2.3.0 implies that the left-hand sides of the two equations are equal; thus $\nu(O) = \mu(O) \leq +\infty$.

Let $A = F \cap O$, where O is open and F is closed. Using the exhaustion principle (I-3.2.4) and setting

$$F_n = F \cap H_n \quad (H_n \text{ defined as in 1.0.3}),$$

we have

$$\mu(F \cap O) = \lim \mu(F_n \cap O),$$

whence it suffices to show that

$$\mu(K \cap O) = \nu(K \cap O)$$

for every compact set K.

By the approximation lemma, 2.2.1(ii), there exists a sequence $\{O_n\}$ of open sets with compact closures such that $K = \lim \downarrow O_n$. Since the $\overline{O}_n$ are compact, $\mu(O_1) < +\infty$; it follows from the principle of decreasing sequences (I-3.2.3) that

$$\mu(K \cap O) = \lim_n \mu(O \cap O_n),$$

and from the first half of the proof that

$$\mu(O \cap O_n) = \nu(O \cap O_n).\square$$

For convenient reference, we restate the first part of the proof of 2.3.1 in a more organized form.

2.3.2 Constructive definition of $\mu(O)$

Let O be an open set in X and let

$$T(O) = \{f \in C_K(X) : \operatorname{supp}(f) \subset O \text{ and } 0 \leq f \leq 1\}.$$

Then

$$\mu(O) = \sup \int f d\mu \quad \text{where} \quad f \in T(O).$$

PROOF. Set $L = \int f d\mu$, where $f \in T(O)$. Since $f \in T(O)$ implies $f \leq \mathbf{1}_O$, we have

$$\int f d\mu \leq \mu(O), \quad \text{whence} \quad L \leq \mu(O).$$

Furthermore, the g_n constructed in the proof of 2.3.1 satisfy $g_n \in T(O)$. Thus

$$\lim \int g_n d\mu \leq L, \quad \text{whence by 2.3.1(i)} \quad \mu(O) \leq L.\square$$

2.3.3 Terminology

Subsets of X which can be written as the intersection of an open set and a closed set are called *sets of type o.c.* Open sets and closed sets are special cases of o.c. sets. (Take their intersection with X.) A subset of X which can be written as a finite union of *disjoint* o.c. sets is called an *elementary set.* It follows from the additivity of μ and ν and from 2.3.1 that $\mu(\mathcal{E}) = \nu(\mathcal{E})$ for every elementary set $\mathcal{E}$.

Lemma. *The elementary sets form a Boolean algebra of subsets of X.*

REMARK. Compare I-1.5.4.

PROOF.

(i) Let R be an o.c. set. Then R^c is an elementary set, for if $R = O \cap F$, then $R^c = O^c \cup F^c$ and we can write

$$R^c = (O^c \cap F^c) \cup (O^c \cap F) \cup (O \cap F^c).$$

The three sets in parentheses are disjoint and each is of type o.c.

(ii) The intersection of two elementary sets is elementary. Indeed, let $\mathcal{E} = \cup_i R_i$ and $\mathcal{E}' = \cup_j R'_j$, where $R_i = O_i \cap F_i$ and $R'_j = O'_j \cap F'_j$. Then

$$\mathcal{E} \cap \mathcal{E}' = \cup_{i,j} R_i \cap R'_j.$$

Since the R_i, R'_j are disjoint, so are the $R_i \cap R'_j$. Moreover, $R_i \cap R'_j = (O_i \cap O'_j) \cap (F_i \cap F'_j)$ and hence is of type o.c.

(iii) The complement of an elementary set is an elementary set. If $\mathcal{E} = \cup R_i$ then $\mathcal{E}^c = \cap R_i^c$. By (i), each R_i^c is an elementary set. By (ii), $\mathcal{E}^c$, as the intersection of finitely many elementary sets, is elementary.

(iv) X is of type o.c. (hence elementary).

(v) A finite union of elementary sets is elementary. By (iii), it suffices to prove the statement for complements of elementary sets; but this follows from (ii). □

2.3.4 Proof of the Radon-Riesz theorem (uniqueness)

PROOF. Let $\mathcal{B} = \{A \in \mathcal{B}_X : \nu(A \cap H_n) = \mu(A \cap H_n)\ \forall n\}$ (where H_n was defined in 1.0.3).

We first show that $\mathcal{B}$ is a *monotone class.* This is immediate for increasing sequences, by I-3.2.1. Now let $\{A_s\}$ be a decreasing sequence, $A_s \in \mathcal{B}$. Then, by the compactness of H_n,

$$\mu(A_s \cap H_n) < +\infty \quad \text{and} \quad \nu(A_s \cap H_n) < +\infty.$$

Applying I-3.2.3,

$$\lim_s \mu(A_s \cap H_n) = \mu((\lim \downarrow A_s) \cap H_n),$$

whence

$$\mu((\lim \downarrow A_s) \cap H_n) = \nu(\lim(\downarrow A_s) \cap H_n).$$

As this is true for any n, we have $(\lim \downarrow A_s) \in \mathcal{B}$.

Furthermore, $\mathcal{B}$ contains the Boolean algebra of the elementary sets of X by 2.3.1. Therefore, by I-1.4, $\mathcal{B}$ coincides with the σ-algebra generated by the open sets and the closed sets; that is, $\mathcal{B} = \mathcal{B}_X$. □

2.4 Proof of existence of the Riesz representation

Given a positive linear functional I on $C_K(X)$, we would like to represent it in integral form. We begin by using a construction that appeared in the proof of uniqueness.

2.4.1 Measure of open sets

As in 2.3.2, we set

$$T(O) = \{f \in C_K(X) : \operatorname{supp}(f) \subset O \quad \text{and} \quad 0 \leq f \leq 1\}.$$

Given a positive linear functional I, we define

$$(i) \qquad I(O) = \sup I(f), \quad \text{where} \quad f \in T(O).$$

$I(O)$ is called the measure of the open set O relative to the linear form I. Note that

$$(ii) \qquad I(O_1) \leq I(O_2) \quad \text{if} \quad O_1 \subset O_2.$$

(iii) Proposition (Convexity inequality). *Let $\{O_n\}$ be a sequence of open subsets of X. Then*

$$I\left(\bigcup_n O_n\right) \leq \sum_n I(O_n).$$

PROOF. Set $W = \cup_n O_n$. Let $f \in T(W)$ and set

$$\Omega = (\operatorname{supp}(f))^c.$$

Then $\Omega, \{O_n\}$ form an open cover of X. Let φ_q be a partition of unity subordinate to this cover. Set

$$f_q = f\varphi_q.$$

Let $S = \{q : \varphi_q f \neq 0\}$. Since f has compact support, $\operatorname{card}(S) < +\infty$. If $q \in S$, let $q \mapsto \theta(q)$ be a mapping from S to $\mathbf{N}$ such that

$$\operatorname{supp}(\varphi_q) \subset O_{\theta(q)}.$$

Set

$$S(n) = \theta^{-1}(n).$$

The nonempty $S(n)$ form a partition of S. Set

$$\psi_n = \sum_{q \in S(n)} \varphi_q.$$

Set $J = \theta(S)$ and $f_n = f\psi_n$. Then $f \leq \sum_{n \in J} f_n$ and, since $f_n \in T(O_n)$, $I(f_n) \leq I(O_n)$ and

$$I(f) \leq \sum_{n \in J} I(O_n) \leq \sum_{n} I(O_n). \square$$

(iv) Proposition (Additivity of I). *Let O_i be a sequence of disjoint open sets, and set $O = \cup O_i$. Then*

$$I(O) = \sum I(O_i).$$

PROOF. Given n and ϵ, consider the nth partial sum of the series on the right-hand side and choose $f_i \in T(O_i)$ such that

$$I(f_i) \geq I(O_i) - \epsilon 2^{-i}.$$

Then $f = \sum_{i=1}^{n} f_i$ satisfies $f \in T(O)$, whence

$$I(O) \geq I(f) = \sum_{i=1}^{n} I(f_i) \geq \sum_{i=1}^{n} I(O_i) - \epsilon.$$

Since n and ϵ are arbitrary, we obtain

$$I(O) \geq \sum_{n=1}^{+\infty} I(O_i),$$

which together with the convexity inequality gives (iv). $\square$

2.4.2 Measure of compact sets

Let K be a compact subset of X and set

$$I(K) = \inf I(O), \quad O \text{ open}, \quad O \supset K.$$

Then

(i) $K_1 \subset K_2$implies$I(K_1) \leq I(K_2)$; and
(ii) if K is compact, O is open, and $K \subset O$, then $I(K) < I(O)$.

(iii) Proposition (Finite additivity). *Let K_1, K_2, ..., K_n be a finite collection of compact disjoint sets. Then*

$$I\left(\bigcup_{i=1}^{n} K_i\right) = \sum_{i=1}^{n} I(K_i).$$

PROOF. Let 2ϵ denote the infimum (minimum) of the distances from K_i to K_j and let

$$U_j = \{x : d(x, K_j) < \epsilon\}.$$

Then the U_j are disjoint open sets.

Choose O_i such that $O_i \supset K_i$ and $I(O_i) < I(K_i) - \epsilon 2^{-1}$, and set $O_i' = U_i \cap O_i$.

Let $K = \cup K_s$ and choose O such that $I(K) > I(O) - \epsilon$. Set $O_j'' = O \cap O_j'$; then $K \subset \cup O_j'' \subset O$, which implies that

$$I\left(\bigcup O_j''\right) - \epsilon \leq I(K) \leq I\left(\bigcup O_j''\right).$$

Since the O_j'' are disjoint, 2.4.1(iv) implies

$$I\left(\bigcup O_j''\right) = \sum I(O_j'').$$

Since

$$K_j \subset O_j'' \subset O_j,$$

we have

$$I(O_j'') - \epsilon 2^{-j} < I(K_j) < I(O_j''),$$

and thus

$$\sum_j I(O_j'') - \epsilon \leq \sum_j I(K_j) \leq \sum_j I(O_j''). \square$$

2.4.3 Inner measure and outer measure

We would like to define set functions for *arbitrary* subsets A of X. We set

$$\mu^*(A) = \inf I(O), \quad O \text{ open}, \quad O \supset A, \quad \text{and}$$
$$\mu_*(A) = \sup I(K), \quad K \text{ compact}, \quad K \subset A.$$

Then by 2.4.2(ii)

(i) $$\mu_*(A) \leq \mu^*(A).$$

$\mu_*(A)$ is called the *inner measure* of A and μ^* its *outer measure.*

(ii) Proposition (Convexity inequality for μ^*). *Let $\{A_n\}$ be a sequence of subsets of X. Then*

$$\mu^*\left(\bigcup_i A_i\right) \leq \sum_{i=1}^{n} \mu^*(A_i).$$

PROOF. Choose a sequence of open sets $\{O_i\}$ such that

$$A_i \subset O_i \quad \text{and} \quad I(O_i) < \mu^*(A_i) + \epsilon\, 2^{-i}.$$

Let $A = \cup_i A_i$ and let $O = \cup O_i$; then $A \subset O$, whence $\mu^*(A) \leq I(O)$.
By 2.4.1(iii),

$$I(O) \leq \sum I(O_i) \leq \sum \mu^*(A_i) + \epsilon. \square$$

(iii) Proposition (Concavity inequality for μ_*). *Let $\{A_i\}$ be a sequence of disjoint subsets of X. Then*

$$\mu_*\left(\bigcup_{i=1}^{+\infty} A_i\right) \geq \sum_{i=1}^{+\infty} \mu_*(A_i).$$

PROOF. Consider the nth partial sum of the series on the right-hand side. Fix compact sets K_i such that

$$K_i \subset A_i \quad \text{and} \quad I(K_i) \geq \mu_*(A_i) - \epsilon\, 2^{-i}.$$

Let $K = \cup_{i=1}^n K_i$; then K is compact. Since the A_i are disjoint, so are the K_i, and finite additivity (2.4.2(iii)) gives

$$I(K) = \sum_{i=1}^{n} I(K_i) \geq \sum_{i=1}^{n} \mu_*(A_i) - \epsilon.$$

Since K is compact and $K \subset A = \cup_i A_i$, we conclude that $\mu_*(A) \geq I(K)$.
$\square$

2.4.4 Construction of the measure (compact case)

Throughout this section, we assume that

(H) $\qquad X$ is compact.

Let

(i) $$\mathcal{B} = \{A \in \mathcal{P}(X) : \mu^*(A) = \mu_*(A)\}.$$

If $A \in \mathcal{B}$, we set

(ii) $$\mu(A) = \mu^*(A) = \mu_*(A).$$

(iii) Proposition. *$A \in \mathcal{B}$ if and only if for every $\epsilon > 0$ there exist a compact set K and an open set O such that*

$$K \subset A \subset O \quad \textit{with} \quad I(O) - \epsilon < I(K) < I(O).$$

PROOF. We prove sufficiency; the proof of necessity is similar. If $A \in \mathcal{B}$, there exists a compact set K such that $K \subset A$ and $I(K) + \frac{\epsilon}{2} > \mu_*(A)$.

There exists an open set O such that $O \supset K$ and $\mu^*(A) > I(O) - \frac{\epsilon}{2}$. Hence the fact that $\mu_*(A) = \mu^*(A)$ implies that

$$I(O) - \epsilon < I(K) < I(O).\square$$

(iv) Proposition. *Every closed set is in $\mathcal{B}$.*

PROOF. Let K be closed (hence compact). Then $\mu_*(K) = I(K)$ by definition, and

$$\mu^*(K) = \inf_{O \supset K} I(O) = I(K)$$

by definition of $I(K)$. $\square$

(v) Proposition. *Every open set is in $\mathcal{B}$.*

PROOF. Let O be an open set. Formally, $\mu^*(O) = I(O)$.

Furthermore, given $\epsilon > 0$, by the definition of $I(O)$ there exists $g \in T(O)$ such that $I(g) > I(O) - \epsilon$.

Let K be the support of g. Then $g \in T(\Omega)$ for every open set $\Omega \supset K$. Hence $I(g) < I(\Omega)\ \forall \Omega \supset K$; that is,

$$I(g) \leq \inf I(\Omega) = I(K).$$

Thus

$$I(K) \geq I(g) \geq I(O) - \epsilon$$

and therefore

$$\mu_*(O) \geq \mu^*(O) - \epsilon.\square$$

(vi) Proposition. *Let $\{A_n\}$ be a sequence of disjoint elements of $\mathcal{B}$. Then*

$$\cup_n A_n \in \mathcal{B} \quad \textit{and} \quad \mu\left(\cup_n A_n\right) = \sum_n \mu(A_n).$$

PROOF. $\mu^*\left(\cup A_n\right) \leq \sum_n \mu(A_n)$ by the convexity inequality, and $\mu_*\left(\cup A_n\right) \geq \sum_n \mu(A_n)$ by the concavity inequality.

Setting $Z = \cup A_n$, we thus have $\mu_*(Z) \geq \mu^*(Z)$, whence $Z \in \mathcal{B}$ and

$$\sum_n \mu(A_n) \leq \mu_*(Z) = \mu(Z) = \mu^*(Z) \leq \sum_n \mu(A_n).\square$$

We now refine criterion (iii).

(vii) Lusin's criterion. *Let $A \in \mathcal{P}(X)$. Then $A \in \mathcal{B}$ if and only if for every $\epsilon > 0$ there exist a compact set K and an open set O such that*

$$K \subset A \subset O \quad \textit{and} \quad \mu(O - K) < \epsilon.$$

PROOF. By (iii), we can find $K \subset A \subset O$ such that

$$\mu(O) < \mu(K) + \epsilon.$$

But $(O - K)$ and K are disjoint and belong to $\mathcal{B}$ (because $(O - K)$ is open and K is closed), whence by (vi)

$$\mu(O - K) + \mu(K) = \mu(O), \quad \text{or} \quad \mu(O - K) = \mu(O) - \mu(K) < \epsilon. \square$$

(viii) Proposition. *$\mathcal{B}$ is a Boolean algebra.*

PROOF. We will use Lusin's criterion (vii). We first show that $A^c \in \mathcal{B}$ if $A \in \mathcal{B}$. There exist a compact set K and an open set O such that

$$K \subset A \subset O \quad \text{with} \quad \mu(O - K) < \epsilon.$$

Then

$$O^c \subset A^c \subset K^c \quad \text{and} \quad K^c - O^c = O - K, \quad \text{whence}$$
$$\mu(K^c - O^c) = \mu(O - K) < \epsilon.$$

Similarly, let $A, A' \in \mathcal{B}$; then $K \cup K' \subset A \cup A' \subset O \cup O'$ and

$$\begin{aligned} (O \cup O') \cap (K \cup K')^c &= (O \cap (K \cup K')^c) \cup (O' \cap (K \cup K')^c) \\ &\subset (O \cap K^c) \cup (O' \cap K'^c). \end{aligned}$$

Hence, by the convexity inequality for the outer measure,

$$\mu^*((O \cup O') \cap (K \cup K')^c) \leq \mu^*(O - K) + \mu^*(O' - K').$$

Since all the sets in this expression are in $\mathcal{B}$, we can replace μ^* by μ to obtain that $A \cup A'$ satisfies (vii); hence $A \cup A' \in \mathcal{B}$. $\square$

(ix) Theorem. *Suppose that X is a compact space and $\mathcal{B}$ is the family of sets defined in (i). Then $\mathcal{B}$ is a σ-algebra containing the Borel algebra and μ defined in (ii) is a measure on $\mathcal{B}$. The σ-algebra $\mathcal{B}$ is μ-complete.*

PROOF. It must be shown that a countable union of sets $A_n \in \mathcal{B}$ is in $\mathcal{B}$. Set

$$B_1 = A_1, \quad B_n = A_n \cap \left(\cup_{j=1}^{n-1} A_j\right)^c.$$

Then $\cup B_n = \cup A_n$ and, since $\mathcal{B}$ is a Boolean algebra, $B_n \in \mathcal{B}$.

Since the B_n are disjoint, it follows from (vi) that their union is in $\mathcal{B}$. Thus $\mathcal{B}$ is a σ-algebra. By (vi), μ is a measure on $\mathcal{B}$. By (iv), $\mathcal{B}$ contains the closed sets; therefore $\mathcal{B}$ contains the Borel algebra $\mathcal{B}_X$. Next, let

$$Y \subset A, \quad \text{where} \quad A \in \mathcal{B} \quad \text{and} \quad \mu(A) = 0.$$

Then

$$\mu^*(Y) \leq \mu^*(A) = 0.$$

Furthermore, by 2.4.3(i),

$$\mu_*(Y) \leq \mu^*(Y),$$

whence

$$\mu_*(Y) = \mu^*(Y) = 0.$$

Thus $Y \in \mathcal{B}$, and hence $\mathcal{B}$ is μ-complete. □

(x) Definition. The measure μ constructed in Theorem 2.4.4(ix) is called the *Radon measure* associated with the positive linear functional I. The σ-algebra $\mathcal{B}$ on which the Radon measure μ is defined contains the Borel algebra $\mathcal{B}_X$. By *restricting* μ to $\mathcal{B}_X$, we can associate a Borel measure μ' with μ. *The σ-algebra $\mathcal{B}$ is the completion of $\mathcal{B}_X$ with respect to the measure* μ'; this will be proved in 3.4.2.

2.4.5 Proof of the representation theorem (compact case)

Theorem. *Let X be a compact space, let I be a positive linear functional on $C(X)$, and let μ be the Borel measure associated with I by 2.4.4(ix) and (x). Then*

$$\int f d\mu = I(f).$$

Proof. We will show that

$$(i) \qquad I(f) \leq \int f d\mu \quad \text{for every} \quad f \in C(X).$$

For a given $\epsilon > 0$, let

$$A_k = f^{-1}([k\epsilon, (k+1)\epsilon)), \quad \text{where} \quad |k| \leq N,$$

with N chosen so that $M = \max |f| < N\epsilon$. Set

$$O_k^n = f^{-1}\left(\left(\left(k - \frac{1}{n}\right)\epsilon, (k+1)\epsilon\right)\right).$$

Then O_k^n is open since f is continuous. $\cap O_k^n = A_k$, and hence the theorem on decreasing sequences gives

$$\lim_n \mu(O_k^n) = \mu(A_k).$$

Fix n so large that

$$(ii) \qquad \sum_{|k| \leq N} (k+1)[\mu(O_k^n) - \mu(A_k)] < 1.$$

Since the A_k form a partition of X, the O_k^n form an open cover of X. Let φ_k be a partition of unity subordinate to this cover. Set $f_k = \varphi_k f$; then $f = \sum f_k$ and moreover $f_k < (k+1)\epsilon\varphi_k$, whence $I(f_k) \leq (k+1)\epsilon I(\varphi_k)$.

Since $0 \leq \varphi_k \leq 1$ and supp $(\varphi_k) \subset O_k^n$, we have $I(\varphi_k) \leq \mu(O_k^n)$, whence

$$I(f) = \sum I(f_k) \leq \sum (k+1)\epsilon\mu(O_k^n).$$

Using (ii),

$$I(f) \leq \sum (k+1)\epsilon\mu(A_k) + \epsilon.$$

Furthermore,

$$\int f d\mu = \sum \int_{A_k} f.$$

But $f(x) \geq k\epsilon$ if $x \in A_k$, whence

$$\int f d\mu \geq \sum k\epsilon\mu(A_k)$$

and therefore

$$I(f) \leq \int f d\mu + \epsilon \left(1 + \sum_k \mu(A_k)\right).$$

Since A_k is a partition of X, $\sum \mu(A_k) = \mu(X)$. Thus

$$I(f) \leq \int f d\mu + \epsilon(1 + \mu(X)).$$

As ϵ is arbitrarily small, we have proved (i).

Now, applying (i) to $f' = -f$, we obtain the opposite inequality to (i); the two inequalities imply equality. □

2.4.6 Proof of the Radon-Riesz theorem (noncompact case)

Let X be a locally compact space which is countable at infinity. Let $\{H_m\}$ be the exhaustion sequence constructed in 1.0.3 and let u_m be the function associated by Urysohn's lemma with the pair $(H_{m-1}, \overline{(H_m^c)})$.

(i) Lemma. *Let $C(H_m)$ denote the functions defined and continuous on H_m. For $f \in C(H_m)$, define $u_m.f$ by*

$$\begin{array}{ll} (u_m.f)(x) = f(x)u_m(x) & \text{if} \quad x \in H_m; \\ (u_m.f)(x) = 0 & \text{if} \quad x \notin H_m. \end{array}$$

Then $u_m.f$ is a continuous function on X.

Proof. Only the behavior at the boundary of H_m must be checked. Let x_0 be a point in the boundary of H_m; then $u_m(x_0) = 0$ and there exists a neighborhood U of x_0 such that $|u_m(x)| < \epsilon$ if $x \in U$. Hence

$$|(u_m.f)(x)| < \epsilon \max |f|.\square$$

(ii) Corollary. *Let I be a positive linear functional on $C_K(X)$. Set*

$$I_m(f) = I(u_m.f), \quad f \in C(H_m).$$

Then I_m is a positive linear functional on H_m.

(iii) By the compact case of Riesz's theorem, proved in 2.4.5, there exists a measure μ_m defined on the Borel algebra $\mathcal{B}_{H_m}$ of H_m such that

$$I_m(f) = \int f d\mu_m, \quad \forall f \in C(H_m).$$

(iv) Let $f \in C_K(X)$; then there exists p such that

$$\operatorname{supp}(f) \subset \overset{\circ}{H}_p.$$

Hence $u_m.f = f$ if $m > p$ and $I(u_m.f) = I(f)$, and thus

$$\int f d\mu_m = I(f) \quad \text{if} \quad m > p.$$

(v) Let O be an open subset of X such that $\overline{O}$ is compact; then there exists p such that $O \subset \overset{\circ}{H}_p$.

Hence, letting

$$T(O) = \{f \in C_K(X) : \operatorname{supp}(f) \subset O\},$$

we have

$$\mu_m(O) = \sup I_m(f), \quad \text{where} \quad f \in T(O).$$

By (iv),

$$I_m(f) = I(f) \quad \text{if} \quad m > p,$$

whence

$$\mu_m(O) = \sup I(f) = \mu_{m'}(O) \quad \text{if} \quad m \text{ and } m' > p.$$

(vi) *The measures μ_m and $\mu_{m'}$ coincide on the Borel algebra $\mathcal{B}_{H_p}$ if m, $m' > p$.*

PROOF. Let

$$\mathcal{Z} = \{A \in \mathcal{B}_{H_p} : \mu_m(A) = \mu_{m'}(A)\}.$$

Let O_q be a decreasing sequence of open sets such that

$$\cap O_q = H_p \quad \text{and} \quad O_q \subset \overset{\circ}{H}_{p+1}.$$

Then $\mu_m(O_q) = \mu_{m'}(O_q)$ by (v). Hence

$$\mu_m(H_p) = \mu_{m'}(H_p).$$

Let $B \in \mathcal{Z}$; then, since $B^c = H_p - B$,

$$\mu_m(B^c) = \mu_m(H_p) - \mu_m(B) = \mu_{m'}(H_p) - \mu_{m'}(B) = \mu_{m'}(B^c).$$

Hence $B \in \mathcal{Z}$ implies $B^c \in \mathcal{Z}$.

Let G be an open subset of H_p. Then there exists $G' \subset \overset{\circ}{H}_{p+1}$ such that G' is open in X and

$$G' \cap H_p = G,$$

whence

$$G = \lim G' \cap O_q.$$

By (v),

$$\mu_m(G' \cap O_q) = \mu_{m'}(G' \cap O_q).$$

Hence $\mathcal{Z}$ contains the open subsets of H_p. Taking complements shows that $\mathcal{Z}$ contains the closed subsets.

We now use 2.4.4(vii) (Lusin's criterion) and 2.4.4(ix). Given a Borel set A and an $\epsilon > 0$, there exist a closed set K and an open set O such that $K \subset A \subset O$ and $\mu_m(O) < \mu_m(K) + \epsilon$.

Since $\mu_m(O) = \mu_{m'}(O)$ and $\mu_m(K) = \mu_{m'}(K)$, it follows that

$$\begin{aligned} &\mu_{m'}(K) \le \mu_{m'}(A) < \mu_{m'}(O) = \mu_m(O) < \mu_m(K) + \epsilon, \\ &\mu_m(K) \le \mu_m(A) < \mu_m(O) < \mu_m(K) + \epsilon. \end{aligned}$$

Hence

$$|\mu_m(A) - \mu_{m'}(A)| < \epsilon.$$

Since ϵ is arbitrarily small, $\mu_m(A) = \mu_{m'}(A)$. □

(vii) Definition of Borel measure.

Let $\{H_m\}$ be the exhaustion sequence defined in 1.0.3. For $A \in \mathcal{B}_X$, set

$$\mu(A) = \lim \mu_{m+2}(A \cap H_m).$$

By (vi),

$$\mu_{m+2}(A \cap H_{m-1}) = \mu_{m+1}(A \cap H_{m-1}),$$

whence the inclusion $A \cap H_{m-1} \subset A \cap H_m$ implies that the sequence $\{\mu_{m+2}(A \cap H_m)\}$ is *increasing*. Hence its limit exists and is finite or equal to $+\infty$.

We first prove *finite additivity*. Let $A_1,\ A_2 \in \mathcal{B}_X$, $A_1 \cap A_2 = \emptyset$. Then, setting $A = A_1 \cup A_2$,

$$\mu_{m+2}(A \cap H_m) = \mu_{m+2}(A_1 \cap H_m) + \mu_{m+2}(A_2 \cap H_m).$$

Hence, passing to the limit,

$$\mu(A) = \mu(A_1) + \mu(A_2).$$

To prove *σ-additivity*, it suffices to show that μ is continuous on increasing sequences.

Let $B_1 \subset B_2 \subset \ldots \subset B_q, \ldots$, where $B_q \in \mathcal{B}_X$, and set $B_\infty = \cup B_q$.

Suppose first that $\mu(B_\infty) = +\infty$. Let M be a positive real number; then there exists m such that

$$\mu(B_\infty \cap H_m) > M.$$

By (vi),

$$\mu(B_\infty \cap H_m) = \mu_{m+2}(B_\infty \cap H_m).$$

Since μ_{m+2} is continuous on increasing sequences, there exists q such that

$$\mu_{m+2}(B_q \cap H_m) > M,$$

whence

$$\mu(B_q) > M.$$

As this is true for all M, $\lim \mu(B_q) = +\infty$.

We now consider the case $\mu(B_\infty) = a < +\infty$. Let $\epsilon > 0$ be given. There exists m such that

$$a - \epsilon < \mu(B_\infty \cap H_m) < a.$$

By (vi), $\mu(B_\infty \cap H_m) = \mu_{m+2}(B_\infty \cap H_m)$. Since μ_{m+2} is continuous on increasing sequences, we have

$$\lim_q \mu_{m+2}(B_q \cap H_m) = \mu_{m+2}(B_\infty \cap H_m).$$

Hence there exists r such that

$$\mu_{m+2}(B_r \cap H_m) > \mu_{m+2}(B_\infty \cap H_m) - \epsilon,$$

and thus $\mu(B_r) > \mu(B_r \cap H_m)$ implies that $\mu(B_r) > \mu(B_\infty) - 2\epsilon$. □

(viii) Representation formula.

Let $f \in C_K(X)$; then there exists m such that supp $(f) \subset \overset{\circ}{H}_m$. By (iv),

$$I(f) = \int_{H_m} f d\mu_{m+2}.$$

But $d\mu_{m+2}$ is equal to $d\mu$ on H_m, whence

$$I(f) = \int f d\mu.$$

(ix) Definition of the associated Radon measure.

Completing the measure space $(X, \mathcal{B}_X, \mu)$ yields a measure $\overline{\mu}$, called the Radon measure associated with the linear functional I.

3 Regularity of Borel Measures and Lusin's Theorem

3.0.1 **Hypothesis.** We assume that the space X is locally compact, metrizable, and countable at infinity.

3.0.2 **Definition.** A measure μ defined on a σ-algebra $\mathcal{B}$ containing the Borel algebra $\mathcal{B}_X$ of X is called *regular* if *for every $A \in \mathcal{B}$ and for every $\epsilon > 0$ there exist an open set O and a closed set F such that $F \subset A \subset O$ and $\mu(O - F) < \epsilon$.*

3.1 Proposition. *Let X satisfy 3.0.1 and let ρ be a locally finite Borel measure on $\mathcal{B}_X$. Then there exists a Radon measure ν such that $\rho(A) = \nu(A)$ for every $A \in \mathcal{B}_X$.*

PROOF. Let $f \in C_K(X)$. Since the indicator function of any compact set is integrable, the inequality $|f| \leq M\mathbf{1}_K$, where $K = \operatorname{supp}(f)$ and $M = \max|f|$, implies that f is integrable (see I-7.4.3).

Hence a positive linear functional can be defined on $C_K(X)$ by setting

$$I(f) = \int f d\rho.$$

By the uniqueness theorem (2.3.4), the linear functional I determines the measure; that is, if ν denotes the Radon measure associated with the form I by Riesz's theorem, then

$$\rho(A) = \nu(A) \quad \text{for any} \quad A \in \mathcal{B}_X. \square$$

3.2 Theorem. *Let X be a locally compact space satisfying the hypothesis of 3.0.1. Then every Radon measure μ on X is regular.*

PROOF. If X is compact, regularity follows from Lusin's criterion, 2.4.4(vii).

If X is only locally compact, let A be a measurable subset of X and let H_n be the exhaustion of the space constructed in 1.0.3. Set $A_n = G_n \cap A$, where $G_n = \overline{(H_n - H_{n-1})}$. Using Lusin's criterion on the compact set H_{n+1}, fix a closed set F_n and an open set O_n of X such that

$$F_n \subset H_{n+1}, \quad F_n \subset A_n \subset O_n, \quad \text{and} \quad \mu(O_n - F_n) < \epsilon 2^{-n}.$$

Note that F_n is compact. Set $O = \cup O_n$; then O is open and $O \supset A$. Similarly, set $F = \cup F_n$. By 1.3.2, this union is locally finite (that is, any compact set meets only a finite number of F_n); hence F is closed. Clearly $F \subset A$ and $\mu(O - F) < \epsilon$. $\square$

3.3 Theorem. *Let X satisfy the hypothesis of 3.0.1. Then any locally finite Borel measure ρ on X is regular.*

PROOF. By 3.1, ρ is the restriction to the Borel algebra of a Radon measure ν. Since ν is regular by 3.2, a fortiori so is ρ. $\square$

3.4 The classes $\mathcal{G}_\delta(X)$ and $\mathcal{F}_\sigma(X)$

3.4.0 **Definition.** The class of subsets of X which can be written as a *countable intersection of open sets* is called $\mathcal{G}_\delta(X)$.

A countable intersection of elements of $\mathcal{G}_\delta(X)$ is in $\mathcal{G}_\delta(X)$.

The class of subsets of X which can be written as a *countable union of closed sets* is called $\mathcal{F}_\sigma(X)$.

A countable union of elements of $\mathcal{F}_\sigma(X)$ is in $\mathcal{F}_\sigma(X)$.

Clearly $\mathcal{G}_\delta(X)$ and $\mathcal{F}_\sigma(X)$ are subclasses of the Borel algebra.

3.4.1 Proposition. *Let μ be a regular measure defined on the σ-algebra $\mathcal{B}$ of the locally compact space X. Then for every $A \in \mathcal{B}$ there exist*

$$\Gamma \in \mathcal{G}_\delta(X) \quad \textit{and} \quad \Phi \in \mathcal{F}_\sigma(X)$$

such that

$$\Phi \subset A \subset \Gamma \quad \textit{and} \quad \mu(\Gamma - \Phi) = 0.$$

PROOF. By 3.0.2, we can find a sequence $\{F'_n\}$ of closed sets and a sequence $\{O'_n\}$ of open sets such that

$$F'_n \subset A \subset O'_n \quad \text{and} \quad \mu(O'_n - F'_n) < n^{-1}.$$

Set

$$O_n = \cap_{q \le n} O'_q \quad \text{and} \quad F_n = \cup_{q \le n} F'_q.$$

Then $F_n \subset A \subset O_n$ and $\{O_n - F_n\}$ is a decreasing sequence. Furthermore, $O_n - F_n \subset O'_n - F'_n$, whence $0 = \lim \mu(O_n - F_n) = \lim \mu(\downarrow (O_n - F_n))$. Set $\Gamma = \lim \downarrow O_n$ and $\Phi = \lim \uparrow F_n$.

Then $\Gamma - \Phi = \lim \downarrow (O_n - F_n)$, whence $\mu(\Gamma - \Phi) = 0$. Finally,

$$\lim \downarrow O_n \in \mathcal{G}_\delta(X) \quad \text{and} \quad \lim \uparrow F_n \in \mathcal{F}_\sigma(X).\square$$

3.4.2 Corollary. *Let μ be a regular measure defined on a σ-algebra $\mathcal{B}$ on X, let μ' be the restriction of μ to the Borel algebra $\mathcal{B}_X$, and let $\overline{\mu}'$ denote the measure defined by extending μ' to the completion $\overline{\mathcal{B}}_X$. Then $\overline{\mathcal{B}}_X \supset \mathcal{B}$ and μ equals the restriction of $\overline{\mu}'$ to $\mathcal{B}$.*

REMARK. Cf. I-4.2.2.

3.4.3 Lusin's theorem. *Suppose that X is a locally compact space, ν is a regular measure defined on the σ-algebra $\mathcal{B} \supset \mathcal{B}_X$, and f is a $\mathcal{B}$-measurable function. Then for every compact set H and every $\epsilon > 0$ there exists a compact set K such that $K \subset H$, $\nu(H - K) < \epsilon$, and the restriction of f to K is continuous.*

PROOF. Set $G_n = \{x : |f(x)| \ge n\} \cap H$. Then $\{G_n\}$ is a decreasing sequence and $\nu(G_n) < +\infty$, whence

$$\lim \ \nu(G_n) = \nu(\cap G_m) = 0.$$

Hence we can find m such that $\nu(G_m) < 2^{-1}\epsilon$.

Considering $f' = f\mathbf{1}_{G_m}$ reduces the proof to the case of *bounded* f. In this case, f is the uniform limit of a sequence $\{g_n\}$ of simple functions (cf. I-6.4.1, of which we follow the notation). Let m_0 be such that $|f(x)| < m_0$ for all x. Setting

$$J_{k,n} = \{x \in H : f(x) \in [kn^{-1}, (k+1)n^{-1}]\},$$

we may take

$$g_n = \sum_k kn^{-1}\mathbf{1}_{J_{k,n}}, \quad \text{where} \quad -nm_0 < k < nm_0.$$

Using the regularity of ν, we can find a compact set $K_{k,n}$ such that

$$K_{k,n} \subset J_{k,n} \quad \text{and} \quad \sum_k \nu(J_{k,n} - K_{k,n}) < 2^{-n-1}\epsilon.$$

Let $V_n = \cup_k K_{k,n}$, where $|k| < nm_0$.

Then V_n is a finite union of compact sets and hence compact. Furthermore, $\nu(H \cap V_n^c) < 2^{-n-1}\epsilon$. Let $W = \cup_n V_n^c \cap H$.

The convexity inequality (I-3.3) gives

$$(i) \qquad \nu(W) < \epsilon \quad \text{and} \quad W^c \cap H = \cap V_n.$$

Set $V_\infty = \cap V_n$. Then V_∞ is compact, whence

$$(ii) \qquad K'_{k,n} = V_\infty \cap K_{k,n} \quad \text{is compact.}$$

Moreover,

$$(iii) \qquad K_{k,n} \quad \text{is } \textit{open} \text{ in} \quad V_n$$

since $K_{k,n}^c \cap V_n = \cup_{j \neq k} K_{j,n}$, $|j| < nm_0$.

By (iii), there exists an open subset Ω of X such that $\Omega \cap V_n = K_{k,n}$ and hence $K'_{k,n} = \Omega \cap V_\infty$; it follows that

$$(iv) \qquad K'_{k,n} \quad \text{is } \textit{open} \text{ in} \quad V_\infty.$$

It follows from (ii) and (iv) that the indicator function of $K'_{k,n}$ is continuous on V_∞. This, with the fact that $J_{k,n} \cap V_\infty = K_{k,n} \cap V_\infty$, gives

(v) The restriction of g_n to V_∞ is continuous. Since g_n converges uniformly on V_∞ to f, the restriction of f to V_∞ is continuous.

Furthermore, (i) shows that $\nu(H - V_\infty) < \epsilon$. □

3.5 Density theorem. *Let X be a locally compact space satisfying the hypothesis 3.0.1 and let ν be a Radon measure on X. Then for every p, $1 \leq p < +\infty$, $C_K(X)$ is dense in $L^p(X, \nu)$.*

PROOF. Let $\{H_n\}$ be the exhaustion of X defined in 1.0.3 and let T_n be the truncation operator, defined in I-6.7, associated with this exhaustion. Let $f \in L^p$ be given. Then

$$(T_n)(f)(x) \to f(x) \quad \text{for every} \quad x \in X$$

and

$$|T_n f - f|^p \leq |f|^p.$$

By the dominated convergence theorem (I-7.6),

$$\|T_n f - f\|_{L^p} \to 0.$$

Let m be such that

$$\|T_m f - f\|_{L^p} < \epsilon.$$

Set $f_m = T_m f$; then f_m is bounded by m and its support is contained in the compact set H_m. Set $\eta = (m^{-1}\epsilon)^p$.

Let K be a compact set, depending on m, such that the restriction φ_m of f_m to K is continuous and such that $\nu(H_m - K) < \frac{\eta}{2}$. Let O be an open set such that $O \supset H_m$ and $\nu(O - H_m) < \frac{\eta}{2}$.

By a theorem of Urysohn,[1] we can find $u \in C_K(X)$ such that supp $(u) \subset O$,

$$u(x) = \varphi_m(x) \quad \text{if} \quad x \in K, \quad \text{and} \quad u(x) \leq m \quad \text{for all } x.$$

On K, $f_m = \varphi_m = u$, whence $f_m - u = (f_m - u)\mathbf{1}_{K^c}\mathbf{1}_O$. Since $|f_m - u| \leq 2m$,

$$\|f_m - u\|_{L^p}^p \leq (2m)^p \nu(O \cap K^c) \leq (2m)^p (m^{-1}\epsilon)^p.$$

Hence $\|f_m - u\|_{L^p} \leq 2\epsilon$, and finally $\|f - u\|_{L^p} \leq 3\epsilon$. □

3.6 REMARK. The regularity of Radon measures allows us to approximate L^p functions by continuous functions, and measurable sets by open or closed sets.

4 The Lebesgue Integral on $\mathbf{R}$ and on $\mathbf{R}^n$

4.1 Definition of the Lebesgue integral on $\mathbf{R}$

We first consider $C_K(\mathbf{R})$, the continuous functions on $\mathbf{R}$ with compact support. The Riemann integral (see the Prologue) defines a positive linear functional on $C_K(\mathbf{R})$ by

$$I(f) = \int f(t)dt.$$

Hence there exists by II-2 a Radon measure μ such that

$$I(f) = \int f(t)d\mu(t).$$

This μ is called the *Lebesgue measure* on $\mathbf{R}$, and functions measurable in this sense are called *Lebesgue measurable.*

---

[1]See, for example, N. Bourbaki, *General Topology* (New York: Springer-Verlag, 1989), IX.4.2.

4.2 Properties of the Lebesgue integral

We include here only properties specific to the Lebesgue integral. Its most important properties are common to all Radon measures, and were established in Sections 2 and 3 of this chapter.

4.2.1 Proposition. *Let $a, b \in \mathbf{R}$, $a < b$. Then*

$$\mu([a,b]) = \mu((a,b)) = b - a.$$

PROOF. $\mu((a,b)) = \sup I(f)$ where $0 \leq f \leq 1$ and $\text{supp}\,(f) \subset (a,b)$. Setting

$$\begin{array}{lll} f = 1 & \text{on} & [a + 2\epsilon, b - 2\epsilon], \\ f = 0 & \text{if} & t < a + \epsilon \quad \text{or} \quad t > b - \epsilon, \end{array}$$

and f linear on $[a + \epsilon, a + 2\epsilon]$ and $[b - 2\epsilon, b - \epsilon]$, we obtain

$$\mu((a,b)) \geq b - a - 3\epsilon.$$

Hence, since ϵ is arbitrary,

$$\mu((a,b)) \geq b - a.$$

The opposite inequality follows from the mean value theorem for the Riemann integral. □

4.2.2 Theorem. *Let O be an open subset of $\mathbf{R}$. Then O is a countable union of disjoint intervals:*

$$(i) \qquad O = \cup_k (a_k, b_k); \quad \textit{and}$$

$$(ii) \qquad \mu(O) = \sum (b_k - a_k).$$

PROOF. Let $x \in O$ and set

$$\alpha(x) = \sup\{y : y < x,\ y \notin O\},$$
$$\beta(x) = \inf\{y : y > x,\ y \notin O\}.$$

Since O^c is closed, $\alpha(x) \in O^c$ if $\alpha(x)$ is finite and $\beta(x) \in O^c$ if $\beta(x)$ is finite. It follows that $(\alpha(x), \beta(x)) \subset O$ and that there exists no open interval which strictly contains $(\alpha(x), \beta(x))$ and is itself contained in O. Moreover, $x \in (\alpha(x), \beta(x))$, whence

$$O = \bigcup_{x \in O} (\alpha(x), \beta(x)).$$

Define an equivalence relation on O by

$$x \sim x' \quad \text{if} \quad (\alpha(x), \beta(x)) = (\alpha(x'), \beta(x')).$$

Since every open interval in $\mathbf{R}$ contains at least one rational number, the set of equivalence classes is countable and (i) follows. We obtain (ii) by using the σ-additivity of μ and 4.2.1. □

4.2.3 Corollary. *Every open set has strictly positive Lebesgue measure.*

4.2.4 Theorem (Characterization of negligible sets). *A subset E of $\mathbf{R}$ is negligible with respect to Lebesgue measure if and only if, for every $\epsilon > 0$, there exists a sequence of intervals (c_k, d_k) such that*

$$\bigcup_k (c_k, d_k) \supset E \quad \text{and} \quad \sum (d_k - c_k) < \epsilon.$$

PROOF. The sufficiency of the condition follows from 4.2.1 and the convexity inequality (1.3.3). Its necessity follows from the regularity of Radon measures (3.2) and from 4.2.2. □

4.2.5 Corollary. *Let $x \in \mathbf{R}$ and let $A = \{x\}$. Then $\mu(A) = 0$.*

PROOF. $O_n = \left(x - \frac{1}{n}, x + \frac{1}{n}\right)$ satisfies $\mu(O_n) < 2n^{-1}$. □

REMARK. We can summarize 4.2.5 by saying that a "point" of $\mathbf{R}$ has Lebesgue measure zero.

4.2.6 Translation invariance

For fixed $a \in \mathbf{R}$, translation by the vector a is the mapping τ_a of $\mathbf{R}$ into $\mathbf{R}$ defined by

$$\tau_a : x \mapsto x + a.$$

Proposition. *Let B be a Lebesgue-measurable subset of $\mathbf{R}$. Then $\tau_a(B)$ is Lebesgue measurable and $\mu(\tau_a(B)) = \mu(B)$.*

PROOF. It follows from the definition of the integral I in 4.1 that $I(\tau_a(f)) = I(f)$, where $(\tau_a f)(x) = f(x - a)$. The uniqueness of the Radon measure associated with a positive linear functional implies the result. □

4.2.7 Notation

By abuse of language, we write

$$\int_{\mathbf{R}} f(t)dt \quad \text{for} \quad \int_{\mathbf{R}} f(t)d\mu(t).$$

We thus use the same notation for the Riemann integral and the Lebesgue integral that extends it. Translation invariance is written

$$(i) \qquad \int_{\mathbf{R}} f(t-a)dt = \int_{\mathbf{R}} f(t)dt.$$

The vector space of Lebesgue-integrable functions defined on $\mathbf{R}$ will be denoted by $L^1(\mathbf{R})$. The next statement follows from the translation invariance of Lebesgue measure.

(ii) If $f \in L^p(\mathbf{R})$, then $\tau_a f \in L^p(\mathbf{R})$ and

$$\|\tau_a f\|_{L^p} = \|f\|_{L^p}, \quad 1 \leq p \leq +\infty.$$

4.3 Lebesgue measure on $\mathbf{R}^n$

4.3.1 Definitions and notation

To simplify notation, we begin by constructing Lebesgue measure on $\mathbf{R}^2$.

We denote by $(\mathbf{R}, \mathcal{A}, \mu)$ the real numbers equipped with Lebesgue measure μ and the σ-algebra of Lebesgue-measurable subsets. Let $(\mathbf{R}, \mathcal{A}, \mu_1)$ and $(\mathbf{R}, \mathcal{A}, \mu_2)$ be two copies of the measure space $(\mathbf{R}, \mathcal{A}, \mu)$.

Let $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$ and let $\mathcal{B}$ denote the tensor product σ-algebra :

$$\mathcal{B} = \mathcal{A}_1 \otimes \mathcal{A}_2.$$

Then $\mathcal{B}$ contains the Borel algebra of $\mathbf{R}^2$ (I-2.4.2). Let $\mu_1 \otimes \mu_2$ be the product measure defined on $\mathcal{B}$ by I-8.4.1.

Lebesgue measure on $\mathbf{R}^2$ is the measure ν obtained by completing $\mu_1 \otimes \mu_2$ (cf. I-4.2.3). The completion of $\mathcal{B}$ is the σ-algebra of *Lebesgue-measurable* subsets of $\mathbf{R}^2$. We denote by $L^1(\mathbf{R}^2)$ the space of Lebesgue-integrable functions on $\mathbf{R}^2$.

If $f \in L^1(\mathbf{R}^2)$, we write

$$\int_{\mathbf{R}^2} f d\nu = \int\int_{\mathbf{R}^2} f(t_1, t_2) dt_1 dt_2.$$

Then, by Fubini's theorem (I-8.5),

$$\int_{\mathbf{R}^2} f(t_1, t_2) dt_1 dt_2 = \int_{\mathbf{R}} dt_2 \left[\int_{\mathbf{R}} f(t_1, t_2) dt_1 \right].$$

Lebesgue measure on $\mathbf{R}^n$ is constructed recursively, by writing $\mathbf{R}^n = \mathbf{R} \times \mathbf{R}^{n-1}$. For $f \in L^1(\mathbf{R}^n)$, the integral thus obtained is written as

$$\int_{\mathbf{R}^n} f(t_1, t_2, \ldots, t_n) dt_1 dt_2 \ldots dt_n,$$

and Fubini's theorem reduces the calculation of this integral to the calculation of n successive integrals on $\mathbf{R}$.

4.3.2 Lebesgue measure on $\mathbf{R}^n$ and the Radon-Riesz theorem

To simplify notation, we restrict to the case where $n = 2$. Let a positive linear functional be defined on $C_K(\mathbf{R}^2)$ by

$$(i) \qquad I(\varphi) = \int dt_2 \left[\int \varphi(t_1, t_2) dt_1 \right], \quad \varphi \in C_K(\mathbf{R}^2).$$

By the Radon-Riesz theorem, there exists a Radon measure ρ such that

$$I(\varphi) = \int \varphi d\rho.$$

By the uniqueness part of the Riesz representation theorem,

$$\rho(A) = \nu(A) \quad \text{for every Borel set } A.$$

Furthermore, since

$$\nu([-R, +R] \times [-R, +R]) = 4R^2,$$

Lebesgue measure is locally finite and hence *regular* by 3.2.

The measures ρ and ν are complete regular measures defined on the Borel algebra.

Lebesgue measure on R^2 may be regarded as the Radon measure associated with (i).

4.3.3 Translation invariance

This is proved as in 4.2.6, by using 4.3.2.

4.3.4 Proposition. *Every open set in R^n has strictly positive Lebesgue measure.*

PROOF. We restrict to the case where $n = 2$. Let O be a nonempty open set and let $(t_1, t_2) \in O$. Then there exists $\epsilon > 0$ such that

$$Q = (t_1 - \epsilon, t_1 + \epsilon) \times (t_2 - \epsilon, t_2 + \epsilon) \subset O.$$

The product measure of the square Q is the product of the measures of its components (I-8.1(i)), whence

$$\nu(Q) = 4\epsilon^2 > 0 \quad \text{and} \quad \nu(O) \geq \nu(Q).\square$$

4.4 Change of variables in the Lebesgue integral on $\mathbf{R}^n$

4.4.0 Some facts from differential calculus

Let O be an open set in $\mathbf{R}^n$. A mapping $f = (f^1, \ldots, f^n)$ is said to be a diffeomorphism if

(i) $f(O)$ is an open subset O' of $\mathbf{R}^n$ and f is a homeomorphism of O onto O' (i.e. a bicontinuous bijection); and

(ii) f and g have continuous first partial derivatives, where g denotes the inverse homeomorphism. The *Jacobian matrix* of f is the matrix

$$J_f = \frac{\partial f^i}{\partial x_k}, \quad 1 \leq i \leq n,\ 1 \leq k \leq n.$$

We then have the following *composition law*:

If f and h are diffeomorphisms for which the composition $h \circ f = q$ is defined, then q is a diffeomorphism and the Jacobian matrix of q is the product of the Jacobian matrices,

$$J_q = J_h J_f. \tag{iii}$$

In particular, $J_g = J_f^{-1}$.

Thus the Jacobian matrix of a diffeomorphism is invertible: $\det(J_f(x))$ is a continuous function that is nowhere zero, and hence has constant sign on a connected component of O.

4.4.1 Change-of-variables theorem

Theorem. *Let O and O' be open subsets of R^n and let f be a diffeomorphism from O onto O'.*

Let $C_K(O')$ denote the continuous functions which have compact support contained in O'. Then

$$\int_O \varphi(f(x))|det\ J_f(x)|dx = \int_{O'} \varphi(x')dx' \quad if \quad \varphi \in C_K(O'). \tag{i}$$

REMARKS.

(ii) Since f is a homeomorphism, $\varphi \in C_K(O')$ implies $(\varphi \circ f) \in C_K(O)$. Since $\det(J_f(x))$ is a continuous function, the integrands on both sides of (i) are continuous functions with compact support and therefore integrable.

(iii) Using a partition of unity on O', we can write $\varphi = \sum \varphi_s$, where the φ_s are supported in arbitrarily small open sets. It thus suffices to prove the theorem for each φ_s. This means that we may assume throughout that φ *has sufficiently small support.*

(iv) *Functoriality.* Suppose that $f = g \circ h$, where g and h are diffeomorphisms. If the change-of-variables formula is proved for the diffeomorphisms g and h, then the result will hold for f in view of the identity

$$|\det\ J_f| = |\det\ J_g|\ |\det\ J_h|.$$

4.4.2 Lemma. *The change-of-variables formula holds when $n = 1$.*

PROOF. In this case, the formula becomes

$$\int \varphi(f(x))|f'(x)|dx = \int \varphi(x')dx'.$$

Using (iii), we can reduce the proof to the case where the support of φ is small enough that $f'(x)$ has constant sign. By the mapping $x \to -x$, this can be further reduced to the case $f'(x) > 0$. Then the formula is

$$\int \varphi(f(x))f'(x)dx = \int \varphi(x')dx'.$$

Set

$$\begin{aligned} F(t) &= \int_0^t \varphi(f(x))f'(x)dx, \\ G(t) &= \int_{f(0)}^{f(t)} \varphi(x')dx'. \end{aligned}$$

Then, differentiating the integrals, we obtain

$$\begin{aligned} G'(t) &= \varphi(f(t))f'(t), \\ F'(t) &= \varphi(f(t))f'(t). \end{aligned}$$

Hence $F(t) - G(t)$ is a constant.

Setting $t = 0$ shows that this constant is zero. □

4.4.3 Proof of the change-of-variables theorem

We proceed by induction on n. Assume that the result holds for $m < n$. Writing $x \in \mathbf{R}^n$ in the form $x = (\xi, y)$, where $\xi \in \mathbf{R}$, $y \in \mathbf{R}^{n-1}$, set

$$\begin{aligned} h(x) &= (\xi', y'), &\quad \text{where} \quad & \xi' = f^1(\xi, y),\ y' = y, \quad \text{and} \\ g(x') &= (\xi', \theta(x')), &\quad \text{where} \quad & \theta = (\widehat{f^2}(\xi, y'),\ \ldots,\ \widehat{f^n}(\xi, y')). \end{aligned}$$

The notation $\widehat{f^2}(\xi, y)$ means that ξ has been replaced in this expression by ξ', by inverting the relation $\xi' = f^1(\xi, y)$.

By the implicit function theorem, this inversion is possible in a neighborhood of x_0 if

$$(ii) \qquad \frac{\partial f^1}{\partial x^1}(x_0) \neq 0.$$

But the fact that det $J_f \neq 0$ implies that the column vector $\left(\frac{\partial f^1}{\partial x^k}\right)_{1 \leq k \leq n}$ is nonzero, and we can renumber the coordinates so that (ii) holds. Thus $g(x')$ can be defined, and it follows from (i) that

$$f = g \circ h.$$

Using 4.4.1(iv), it suffices to prove the theorem for g and for h. Next, we calculate

$$\int_{\mathbf{R} \times \mathbf{R}^{n-1}} \varphi(\xi', \theta(x'))(\det\ J_g)\ d\xi' dy'.$$

By Fubini's theorem, this equals

$$(iii) \qquad \int_{\mathbf{R}} d\xi' \int_{\mathbf{R}^{n-1}} \varphi(\xi', y')(\det\ J_g)dy'.$$

Note that the Jacobian matrix J_g has some row for which all entries are zero except the diagonal entry, which equals 1.

Thus $\det J_g = \det J_{\theta_{\xi'}}$, where $\theta_{\xi'} : y' \to \theta(\xi', y')$.

By the induction hypothesis,

$$\int_{\mathbf{R}^{n-1}} \varphi(\xi', y')\det(J_{\theta_{\xi'}})dy' = \int_{\mathbf{R}^{n-1}} \varphi(\xi', y'')dy'',$$

and substituting this into (iii) proves the theorem for the change of variables defined by g.

It remains to prove the theorem for h. Note that, by Fubini,

$$(iv) \quad \int_{\mathbf{R}^n} \varphi(f^1(\xi, y), y)\det(J_h)d\xi dy = \int_{\mathbf{R}^{n-1}} dy \int_{\mathbf{R}} \varphi(f^1(\xi, y), y)\det J_h d\xi.$$

But $\det J_h = \frac{\partial f^1}{\partial \xi}$ and, by 4.4.2,

$$\int_{\mathbf{R}} \varphi(f^1(\xi, y), y)\frac{\partial f^1}{\partial \xi}d\xi = \int_{\mathbf{R}} \varphi(\eta, y)d\eta.$$

The result follows by substitution into (iv). □

REMARK. This proof can be given the following geometric interpretation. Let $p : \mathbf{R}^n \to \mathbf{R}$, where p is differentiable and $\nabla p \neq 0$ everywhere. Then the volume element $dv_{\mathbf{R}^n}$ can be written locally as the volume element on the hypersurface $p =$ constant, "multiplied" by $\frac{dp}{\|\nabla p\|}$. The induction hypothesis allows us to treat the change of variable on the hypersurface; the other change of variable occurs in one dimension.

5 Linear Functionals on $C_K(X)$ and Signed Radon Measures

In Section 2 we studied *positive* linear functionals on $C_K(X)$. We now drop the hypothesis of positivity and substitute the more general hypothesis of *continuity.*

5.1 Continuous linear functionals on $C(X)$ (X compact)

Throughout this section, X is a compact space. Then $C_K(X)$ is the space $C(X)$ of all continuous functions. A norm is defined on $C(X)$ by setting

$$\|f\|_C = \max_{x\in X} |f(x)|.$$

Convergence in this norm corresponds to *uniform convergence*. $C^*(X)$ denotes the Banach space of continuous linear functionals l on $C(X)$; that is, those for which there exists a constant a such that

$$|l(f)| \leq a\|f\|_C.$$

Setting $\|l\|_{C^*} = \sup_{\|f\|_C \leq 1} |l(f)|$ yields $|l(f)| \leq \|l\|_{C^*}\|f\|_C$.

5.1.1 Proposition. *If l is positive, then l is continuous.*

PROOF. Indeed,

$$-\|f\|_C \leq f(x) \leq \|f\|_C$$

implies

$$-\|f\|_C l(\mathbf{1}_X) \leq l(f) \leq \|f\|_C l(\mathbf{1}_X),$$

whence

(*i*) $$\|l\|_{C^*} = l(\mathbf{1}_X).\square$$

5.2 Decomposition theorem

Theorem. *Let X be a compact space and let $l \in C^*(X)$. Then there exist positive linear functionals l^+ and l^- such that*

5.2.1 $$l = l^+ - l^- \quad \textit{and}$$

5.2.2 $$\|l\|_{C^*} = \|l^+\|_{C^*} + \|l^-\|_{C^*}$$

and such a decomposition is unique.

PROOF. For each nonnegative f in $C(X)$, let

$$H(f) = \{u \in C(X) : 0 \leq u \leq f\}$$

and let

(*i*) $$l^+(f) = \sup l(u), \quad \text{where} \quad u \in H(f).$$

Let f_1, $f_2 \geq 0$. Since $u_1 \in H(f_1)$ and $u_2 \in H(f_2)$, $u_1 + u_2 \in H(f_1 + f_2)$; hence $H(f_1) + H(f_2) \subset H(f_1 + f_2)$.

We now prove the opposite inclusion. Let $u \in H(f_1 + f_2)$ be given. Set $v = \min\{u, f_1\} = \frac{1}{2}(u + f_1 - |u - f_1|)$; then $v \in C(X)$, $v \in H(f_1)$, and $w = u - v \in H(f_2)$.

Thus $u = v + w$ with $v \in H(f_1)$, $w \in H(f_2)$, and we have shown that

(*ii*) $$H(f_1 + f_2) = H(f_1) + H(f_2).$$

This implies

(*iii*) $$l^+(f_1 + f_2) = l^+(f_1) + l^+(f_2), \quad f_1, f_2 \geq 0.$$

Any $g \in C(X)$ can be written as

(*iv*) $$g = g_1 - g_2 \quad \text{with} \quad g_1, g_2 \geq 0.$$

(For example, we can take $g_1 = \max(g, 0)$.)

Define

$$(v) \qquad l^+(g) = l^+(g_1) - l^+(g_2).$$

We will justify this definition by showing that the right-hand side is independent of the choice of the decomposition (iv). Let

$$g = g_3 - g_4, \quad g_3, g_4 \geq 0.$$

Then $g_1 - g_2 = g_3 - g_4$, or $g_1 + g_4 = g_3 + g_2$. Using (iii),

$$l^+(g_1) + l^+(g_4) = l^+(g_3) + l^+(g_2),$$

or

$$l^+(g_1) - l^+(g_2) = l^+(g_3) - l^+(g_4),$$

which justifies definition (v).

It therefore follows from (iii) and (v) that

$$l^+(g + g') = l^+(g) + l^+(g').$$

Similarly, it follows from (i) that

$$l^+(\alpha f) = \alpha l^+(f) \quad \text{if} \quad \alpha \geq 0, \ f \geq 0.$$

Since $0 \in H(f)$, we have $l^+(f) \geq 0$ if $f \geq 0$, whence

(vi) l^+ is a positive linear functional on $C(X)$.

Setting $l^- = l^+ - l$, we have $l^- \in C^*(X)$. Furthermore, let $f \in C(X)$, $f \geq 0$. Then

$$l^-(f) = (\sup l(u)) - l(f) = \sup(l(u - f)), \quad \text{where} \quad u \in H(f).$$

For $f \geq 0$, set $G(f) = \{v \in C(X) : -f \leq v \leq 0\}$. Then the mapping $u \mapsto u - f$ defines a bijection of $H(f)$ onto $G(f)$; hence $l^-(f) = \sup l(v)$, where $v \in G(f)$.

Since $0 \in G(f)$, $l^-(f) \geq 0$ and we have thus obtained the decomposition 5.2.1.

Let 1 denote the indicator function of the full set X; then, by 5.1(i),

$$\|l^+\|_{C^*(X)} = l^+(\mathbf{1}) \quad \text{and} \quad \|l^-\|_{C^*(X)} = l^-(\mathbf{1}).$$

There exist $u_n \in H(\mathbf{1})$ and $v_n \in G(\mathbf{1})$ such that

$$l^+(\mathbf{1}) = \lim l(u_n) \quad \text{and} \quad l^-(\mathbf{1}) = \lim l(v_n).$$

It is straightforward to show that

$$\|l^+\|_{C^*} + \|l^-\|_{C^*} = \lim(l(u_n) + l(v_n)).$$

We have $0 \le u_n \le 1$, $-1 \le v_n \le 0$, and $-1 \le u_n(x) + v_n(x) \le 1$, or

$$\|u_n + v_n\|_C \le 1.$$

Hence

$$|l(u_n + v_n)| \le \|l\|_{C^*},$$

and we have shown that

$$\|l^+\|_{C^*} + \|l^-\|_{C^*} \le \|l\|_{C^*}.$$

Since the opposite inequality follows from the triangle inequality, this proves 5.2.2. □

5.2.3 Uniqueness of the decomposition

Let

(i) $l = \varphi - \psi$ where φ, ψ are positive linear functionals.

Then

$$l^+(f) = \sup\{\varphi(u) - \psi(u)\} \quad \text{with} \quad u \in H(f).$$

But

$$\varphi(u) - \psi(u) \le \varphi(u)$$

since $u \ge 0$, and thus

$$\sup\{\varphi(u) - \psi(u)\} \le \sup \varphi(u) = \varphi(f).$$

That is,

$$l^+(f) \le \varphi(f) \quad \text{for every } f \ge 0.$$

Set $\varphi - l^+ = \theta$; then θ is a *positive linear functional*, and it follows from (i) that

(ii)
$$\varphi = l^+ + \theta \quad \text{and} \quad \psi = l^- + \theta.$$

Then $\|\varphi\|_{C^*} = \varphi(\mathbf{1}) = l^+(\mathbf{1}) + \theta(\mathbf{1}) = \|l^+\|_{C^*} + \|\theta\|_{C^*}$; similarly $\|\psi\|_{C^*} = \|l^-\|_{C^*} + \|\theta\|_{C^*}$.

Suppose that the decomposition $\varphi - \psi$ satisfies 5.2.2; then

$$\|l\|_{C^*} = \|\varphi\|_{C^*} + \|\psi\|_{C^*} = \|l^+\|_{C^*} + \|l^-\|_{C^*} + 2\|\theta\|_{C^*}.$$

Furthermore, by 5.2.2, $\|l\|_{C^*} = \|l^+\|_{C^*} + \|l^-\|_{C^*}$. Subtracting these two equations shows that $2\|\theta\|_{C^*} = 0$; thus $\theta = 0$. □

5.2.4 Corollary. *Given $l \in C^*(X)$, there are two Borel measures μ_1 and μ_2 uniquely determined by*

$$l(f) = \int f d\mu_1 - \int d\mu_2 \qquad \textit{and}$$

$$\|l\|_{C^*} = \mu_1(X) + \mu_2(X).$$

PROOF. By the decomposition theorem (5.2) and the Radon-Riesz theorem. □

5.3 Signed Borel measures

In this section, we establish the equivalent of Theorem 5.2 for Borel measures.

5.3.1 **Definition.** A signed Borel measure on the compact space X is a mapping

$$\nu : \mathcal{B}_X \to \mathbf{R}$$

that can be written in the form

$$(i) \qquad \nu(A) = \rho_1(A) - \rho_2(A),$$

where ρ_1, ρ_2 are finite Borel measures. The decomposition (i) is clearly not unique; adding the same Borel measure θ to ρ_1 and ρ_2 will not change the mapping ν.

5.3.2 Mutually singular measures

Two Borel measures ν_1 and ν_2 are said to be *mutually singular* if there exists a Borel set $A \in \mathcal{B}_X$ such that

$$(i) \qquad \nu_1(A) = \nu_1(X) \quad \text{and} \quad \nu_2(A) = 0.$$

The relation is symmetric, for A^c carries all the mass of ν_2 and has ν_1-measure zero.

5.3.3 Theorem. *If X is a compact space, there exists a bijection between the continuous linear functionals on $C(X)$ and the signed Borel measures. The decomposition of a linear functional given in Theorem 5.2 corresponds to the decomposition of the signed Borel measure as a difference of two mutually singular Borel measures.*

PROOF. We use 5.2.4. The only statement still needing proof is the equivalence of the following two properties:

$$(i) \qquad \|l\|_{C^*} = \rho_1(X) + \rho_2(X).$$

$$(ii) \qquad \rho_1 \text{ and } \rho_2 \text{ are mutually singular.}$$

We first show that (ii) $\Rightarrow$ (i). If ρ_1 and ρ_2 are mutually singular, let A be an element of $\mathcal{B}_X$ such that

$$\rho_1(A) = \rho_1(X) \quad \text{and} \quad \rho_2(A) = 0.$$

Set

$$\varphi = \mathbf{1}_A - \mathbf{1}_{A^c}.$$

Then

$$\int \varphi d\rho_1 = \rho_1(X) \quad \text{and} \quad \int \varphi d\rho_2 = -\rho_2(X),$$

whence

$$\int \varphi(d\rho_1 - d\rho_2) = \rho_1(X) + \rho_2(X).$$

By Theorem 3.5, we can find $f \in C(X)$ such that

$$\|f - \varphi\|_{L^1} < \epsilon, \quad \text{where} \quad L^1 = L^1(X, \mathcal{B}_X, \rho_1 + \rho_2).$$

Set

$$\begin{array}{lll} \tilde{f}(x) = f(x) & \text{if} & |f(x)| \leq 1; \\ \tilde{f}(x) = \text{signum}\, f(x) & \text{if} & |f(x)| > 1. \end{array}$$

Then $\|\tilde{f} - \varphi\|_{L^1} < 2\epsilon$ and

$$\int \tilde{f}(d\rho_1 - d\rho_2) > \rho_1(X) + \rho_2(X) - 2\epsilon, \quad \text{where} \quad \tilde{f} \in C(X), \ \|\tilde{f}\|_C \leq 1.$$

Conversely, we show that (i) $\Rightarrow$ (ii). There exists a sequence $\{\varphi_n\}$ in the closed unit ball of $C(X)$ such that $l(\varphi_n) \to \|l\|_{C^*}$. Set $\varphi_n = \varphi_n^+ - \varphi^-$; then

$$l(\varphi_n) = \left[\int \varphi_n^+ d\rho_1 + \int \varphi_n^- d\rho_2\right] - \left[\int \varphi_n^- d\rho_1 + \int \varphi_n^+ d\rho_2\right].$$

Since $\varphi_n^+ \leq 1$ and $\varphi_n^- \leq 1$, the first term in brackets is at most equal to $\rho_1(X) + \rho_2(X) = \|l\|_{C^*}$ by (i). Hence the convergence of $l(\varphi_n)$ to $\|l\|_{C^*}$ implies that

$$\int \varphi_n^+ d\rho_1 \to \rho_1(X) \quad \text{and} \quad \int \varphi_n^- d\rho_1 \to 0.$$

Since

$$\|1 - \varphi_n^+\|_{L^1(\rho_1)} = \int (1 - \varphi_n^+) d\rho_1,$$

we conclude that

$$\|1 - \varphi_n^+\|_{L^1(\rho_1)} \to 0 \quad \text{and} \quad \|\varphi_n^-\|_{L^1(\rho_1)} \to 0.$$

Passing to a subsequence $\{\varphi_{n_1}\}$, we may replace the convergence in $L^1(\rho_1)$ of $\{1 - \varphi_n^+\}$ by convergence ρ_1-a.e. Passing to a new subsequence $\{\varphi_{n_1}\}$ reduces the proof to the case where $\psi_s = \varphi_{n_1}^+$ satisfies

$$\begin{array}{lll} \psi_s & \text{converges to } 1 & \rho_1\text{-a.e.}; \\ \psi_s & \text{converges to } 0 & \rho_2\text{-a.e.} \end{array}$$

Let

$$A = \{x : \lim \psi_s^+(x) = 1\}.$$

Then

$$\|1 - \mathbf{1}_A\|_{L^1(\rho_1)} = 0 \quad \text{and} \quad \|\mathbf{1}_A\|_{L^1(\rho_2)} = 0,$$

or

$$\rho_1(A) = \rho_1(X) \quad \text{and} \quad \rho_2(A) = 0. \square$$

5.3.4 Proposition. *Let ν be a signed Borel measure. Then there exists a decomposition*

$$\nu = \rho_1^0 - \rho_2^0$$

such that ρ_1^0 and ρ_2^0 are mutually singular. Such a decomposition is unique. We set

$$|\nu| = \rho_1^0 + \rho_2^0$$

and call $|\nu|$ the absolute value of ν.

PROOF. Let a continuous linear functional on $C(X)$ be defined by setting

$$l(\varphi) = \int \varphi d\nu = \int \varphi d\rho_1 - \int \varphi d\rho_2.$$

Then, by 5.3.3, the decomposition of ν as a difference of mutually singular Borel measures corresponds to the decomposition of l given by 5.2.4. This decomposition exists and is unique by 5.2.4. □

5.3.5 Signed Radon measures

Given a signed Borel measure ν on the compact space X, let $\rho_1^0 - \rho_2^0$ be its canonical decomposition. Let $\mathcal{B}$ be the completion of the Borel algebra $\mathcal{B}_X$ with respect to $|\nu|$. We define a signed measure on $\mathcal{B}$ by setting

$$\mu(B) = \rho_1^0(B) - \rho_2^0(B), \quad \forall B \in \mathcal{B}.$$

μ is called the signed Radon measure associated with the signed Borel measure ν.

If X is a locally compact space, a signed Radon measure ν on X is given by two *mutually singular* Borel measures ν_1 and ν_2. We set $|\nu| = \nu_1 + \nu_2$ and define the σ-algebra $\mathcal{B}_\nu$ by completing the Borel algebra $\mathcal{B}_X$ with respect to $|\nu|$. Then, if $A \in \mathcal{B}_\nu$ and $|\nu|(A) < +\infty$, we define $\nu(A) = \nu_1(A) - \nu_2(A)$.

5.3.6 Important remark on terminology

Let X be a locally compact space. We denote by $M(X)$ the vector space of signed Radon measures and by $M^+(X)$ the *Radon measures* on X; that is, the measures associated with positive linear forms. In the usual terminology, $M(X)$ is called the space of Radon measures and $M^+(X)$ the space of *positive* Radon measures. From the point of view of grammatical accuracy, this terminology is better than ours; a noun modified by an adjective should describe a *narrower* class of objects than the noun alone. Our use throughout Chapter I of the word "measure" to mean a positive measure may justify our ignoring this rule now.

5.3.7 Complex measures

We denote by $C(X; \mathbf{C})$ the space of continuous *complex-valued* functions on the compact space X. Separating real and imaginary parts, we can write

$$C(X; \mathbf{C}) = C(X) \oplus C(X).$$

A $\mathbf{C}$-linear functional l on $C(X;\mathbf{C})$ is determined by restricting $\mathrm{Re}(l)$ to each summand of the direct sum. Since X is assumed compact, specifying l is equivalent to specifying two signed Radon measures μ_1 and μ_2. Setting $\|\mu\| = \|\mu_1\| + \|\mu_2\|$ and $l = \mu_1 + i\mu_2$, we have

$$l(f + ih) = \int f d\mu_1 - \int h d\mu_2 + i \int f d\mu_2 + h d\mu_1.$$

$\mu_1 + i\mu_2$ is called the *complex* measure associated with this form.

5.4 Dirac measures and discrete measures

5.4.1 Dirac measures

Let X be a locally compact space X and let $x_0 \in X$. The Dirac measure at x_0 is the linear functional

$$l_{x_0}(f) = f(x_0), \quad \forall f \in C_K(X_0).$$

This positive linear functional is represented by a Borel measure δ_{x_0} whose completion is defined on the σ-algebra $\mathcal{P}(X)$ consisting of all the subsets of X. We have

$$\begin{aligned} \delta_{x_0}(A) &= 1 \quad \text{if} \quad x_0 \in A \\ \delta_{x_0}(A) &= 0 \quad \text{if} \quad x_0 \notin A. \end{aligned}$$

5.4.2 Discrete measures

Now let $x_1, \dots, x_j, \dots \in X$ and $\alpha_j \in \mathbf{R}$. Suppose that, for every compact set K,

(*i*) $$\sum_{j \in S_K} |\alpha_j| < +\infty, \quad \text{where} \quad S_K = \{j : x_j \in K\}.$$

A locally finite signed Borel measure ν is defined by setting, for $B \in \mathcal{B}_X$,

$$\nu(B) = \sum \alpha_j, \quad \text{where} \quad j \in S_B = \{j : x_j \in B\}.$$

This series is absolutely convergent by (i). Let $\nu^+ = \sum_{\alpha_j > 0} \alpha_j \delta_{x_j}$ and let $\nu^- = \sum_{\alpha_j < 0} -\alpha_j \delta_{x_j}$. Then ν^+ and ν^- are *locally finite* Borel measures. Completing the Borel algebra with respect to $|\nu| = \nu^+ + \nu^-$, we recover the σ-algebra of subsets $\mathcal{P}(X)$; hence

$$|\nu|(C) \leq +\infty \quad \text{is defined} \quad \forall C \in \mathcal{P}(X).$$

In contrast, $\nu(C)$ is defined only for those $C \in \mathcal{P}(X)$ which also satisfy $|\nu|(C) < +\infty$.

We denote by $M_d(X)$ the discrete measures on A and by $M_d^1(X)$ the finite discrete measures: $M_d^1(X) = M^1(X) \cap M_d(X)$.

5.5 Support of a signed Radon measure

5.5.1 **Definition.** Let $\mu \in M(X)$. The *support* of μ, written supp (μ), is the smallest closed set F such that $|\mu|(F^c) = 0$. Let us show that this set exists. Taking complements, finding F is equivalent to finding the largest open set H such that $|\mu|(H) = 0$.

The hypothesis 3.0.1 implies that X satisfies the second separability axiom of I-2.4.1. Therefore we can find a countable family of open sets O_n which forms a basis for the open sets. Set

$$S = \{n : |\mu|(O_n) = 0\} \quad \text{and} \quad H = \cup_{n \in S} O_n.$$

Then H, as a countable union of sets of measure zero, has measure zero: $|\mu|(H) = 0$.

Let O' be an open set such that $|\mu|(O') = 0$; then $O' = \cup_{n \in T} O_n$ (since $\{O_n\}$ is a basis for the open sets). The hypothesis $|\mu|(O') = 0$ implies that $|\mu|(O_n) = 0$, whence $T \subset S$ and $O' \subset H$.

5.5.2 Proposition. *Suppose that X is a locally compact space, $f \in C_K(X)$, and $\mu \in M(X)$. Then*

$$\int f d\mu = 0 \quad \textit{if} \quad \operatorname{supp}(f) \cap \operatorname{supp}(\mu) = \emptyset.$$

PROOF. Let $\mu = \mu_1 - \mu_2$ with $|\mu| = \mu_1 + \mu_2$, and let $H = (\operatorname{supp}(\mu))^c$. Then $f = 0$ $|\mu|$-a.e., whence $f = 0$ a.e. μ_i, $i = 1, 2$, which implies that $\int f d\mu_i = 0$, $i = 1, 2$.

6 Measures and Duality with Respect to Spaces of Continuous Functions on a Locally Compact Space

6.1 Definitions

We consider the following three vector spaces of continuous functions on X:

$C_K(X)$,	the continuous functions with compact support;
$C_0(X)$,	the continuous functions which vanish at infinity; and
$C_b(X)$,	the bounded continuous functions.

(i) Recall that a function f is said to vanish at infinity if, for every $\epsilon > 0$, there exists a compact set K such that $|f(x)| < \epsilon$ for $x \notin K$. We have the following inclusions:

(ii) $$C_K(X) \subset C_0(X) \subset C_b(X).$$

If X is compact, these three spaces coincide; if X is not compact, each of the inclusions is strict. A norm is defined on $C_b(X)$ by setting, for $f \in C_b(X)$,

$$(iii) \qquad \|f\|_{C_b} = \sup |f(x)|, \quad x \in X.$$

This norm defines, by *restriction*, norms on $C_0(X)$ and $C_K(X)$. The restriction of $\| \ \|_{C_b}$ to $C_0(X)$ will sometimes be denoted by $\| \ \|_{C_0}$. We then have

$$(iv) \qquad \|h\|_{C_0} = \max |h(x)|, \quad x \in X.$$

The difference between (iii) and (iv) is that, although the supremum may not be attained in (iii), it is attained in (iv) and gives a maximum.

6.2 Proposition. *The space $C_b(X)$ equipped with the norm (iii) is complete. The space $C_0(X)$ is a closed subspace of $C_b(X)$ and is therefore complete. The space $C_K(X)$ is a dense subspace of $C_0(X)$.*

PROOF. Only the third (and hardest) assertion will be proved here.[2]

Let $\{H_n\}$ be the exhaustion sequence of compact sets constructed in 1.0.3. Recall that $H_n \subset \overset{\circ}{H}_{n+1}$. For each n, let φ_n, ψ_n be a partition of unity subordinate to the open cover consisting of the two sets $\overset{\circ}{H}_{n+1}$ and H_n^c. Then, since $\varphi_n + \psi_n = 1$ on X,

$$\varphi_n = 1 \quad \text{on } H_n.$$

Given $h \in C_0(X)$, set $h_n = h\varphi_n$. Then $h_n \in C_K(X)$ and $\|h - h_n\|_{C_0} = \|h\psi_n\|_{C_0} \to 0$ as $n \to \infty$, since supp $(\psi_n) \subset H_n^c$ and $h(x) \to 0$ as x tends to infinity. □

6.3 The Alexandroff compactification

Given a locally compact space X, we can associate with it a *compact* space Y and a homeomorphism ψ of X onto Y with one point removed. Y is called the *Alexandroff compactification* of X. The construction consists of adjoining a point at infinity to X by setting $Y = X \cup \{\infty\}$, where ∞ is a new element. The complements of compact subsets of X are taken as a system of open neighborhoods of ∞.

Having thus defined Y from the set-theoretic point of view, we now construct a topology on Y in a more precise way by specifying its closed subsets.

---

[2]For the first two, see for example E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, 3rd ed. (New York: Springer-Verlag, 1975).

A subset F of Y is closed if and only if it satisfies the following conditions:

(i) $F \cap X$ is closed; and
(ii) if $F \cap X$ is not compact, then $\infty \in F$.

Let p be the injection of X into Y; then, by (i), $p^{-1}(F)$ is closed in X if F is closed in Y. Let H be closed in X; then, by (ii), H or $H \cup \{\infty\}$ is closed in Y. Intersecting with $\{\infty\}^c$ shows that H is a relatively closed set in $\{\infty\}^c$. Thus p is a homeomorphism of X onto $\{\infty\}^c$.

The open neighborhoods of ∞ are the complements of closed sets that do not contain ∞; that is, the complements of compact subsets of X. It follows easily that the topology of Y is Hausdorff.

We now show that Y is compact. Let O_γ be an open cover of Y. There exists γ_0 such that $\infty \in O_{\gamma_0}$; hence there exists a compact set K such that $O_{\gamma_0} = K^c$. The sets $O_\gamma \cap K$ form an open cover of K. Let $O_{\gamma_1} \cap K, \ldots, O_{\gamma_n} \cap K$ be a finite subcover. Then $O_{\gamma_0}, \ldots, O_{\gamma_n}$ form a finite subcover of Y.

6.4 Proposition. *Let X be a locally compact space and let Y be its Alexandroff compactification. Set*

$$V = \{f \in C(Y) : f(\infty) = 0\}.$$

For every function $f \in V$, let $\tilde{f}$ denote its restriction to X. Then

$$f \to \tilde{f}$$

is a linear mapping which is an isometry of V onto $C_0(X)$.

PROOF. Let $f \in V$; then the restriction $\tilde{f}$ of f to X defines an element $\tilde{f} \in C_b(X)$. Furthermore, since f is continuous at ∞, for every $\epsilon > 0$ there exists a compact set K such that $|f(x) - f(\infty)| < \epsilon$ if $x \notin K$. Hence $\tilde{f} \in C_0(X)$.

Conversely, let $h \in C_0(X)$. Then h can be extended to Y by setting $h_1(\infty) = 0$ and setting $h_1(x) = h(x)$ if $x \in X$. Since $h \in C_0(X)$, h_1 is continuous at the point ∞ and hence continuous everywhere. □

6.5 *The space $M^1(X)$*

(i) We denote by $M^1(X)$ the set of signed Radon measures ν on X such that $|\nu|$ is finite, and define a norm on $M^1(X)$ by setting

$$\|\nu\|_{M^1} = \int d|\nu| = |\nu|(X).$$

Moreover, for every Borel set A of X, $\nu(A) = \nu_1(A) - \nu_2(A)$ is well defined. (See 5.3.5.)

(ii) Proposition. *Let Y be the compactification of X and let*

$$W = \{\nu \in M(Y) : \nu(\{\infty\}) = 0\}.$$

Let a mapping $\nu \to \tilde{\nu}$

$$W \mapsto M^1(X)$$

be defined by setting

$$\tilde{\nu}(A) = \nu(A) \quad \forall A \in \mathcal{B}_X \subset \mathcal{B}_Y.$$

This mapping is an isometric bijection of W onto $M^1(X)$.

PROOF. It suffices to note that every $B \in \mathcal{B}_Y$ can be written either as $B = A \cup \{\infty\}$ or as $B = A$, for some $A \in \mathcal{B}_X$. In the first case, the additivity of ν gives $\nu(B) = \nu(A) + \nu(\{\infty\}) = \nu(A)$ since $\nu(\{\infty\}) = 0$. □

6.6 Theorem. *$M^1(X)$ is the Banach space dual of $C_0(X)$.*

PROOF. With the notation of 6.4, $C_0(X) \simeq V \subset C(Y)$. Let $l \in C^*(Y)$; then its restriction to V defines a continuous linear form on V. By the Hahn-Banach theorem, every linear functional on V can be written in this way. Thus

$$(V)^* \simeq C^*(Y)/H,$$

where H is the space of linear functionals which vanish identically on V. Since V has codimension 1, H has dimension 1 and is therefore the vector subspace generated by δ_∞, the Dirac measure at infinity. But, in the notation of 6.5(ii), $W \simeq M(Y)/H$, whence $(C_0(X))^* \simeq M(Y)/H \simeq W \simeq M^1(X)$. All these identifications are isometric. In particular, for every $\mu \in M^1(X)$,

$$\sup_{\|f\|_{C_0} \leq 1} \int f d\mu = \|\mu\|_{M^1}. \square$$

6.7 Defining convergence by duality

The following three spaces of continuous functions are associated with a locally compact space X:

$$C_K(X) \subset C_0(X) \subset C_b(X).$$

Convergence in $M(X)$. A sequence $\{\mu_n\}$, $\mu_n \in M(X)$, is said to *converge vaguely* to $\mu_0 \in M(X)$ if

$$(i) \qquad \int f d\mu_n \to \int f d\mu_0, \quad \forall f \in C_K(X).$$

Convergence in $M^1(X)$. Given a sequence $\{\nu_n\}$, $\nu_n \in M^1(X)$, we have two new concepts of convergence.

ν_n is said to *converge weakly* to ν_0 if

$$(ii) \qquad \int h d\nu_n \to \int h d\nu_0, \quad \forall h \in C_0(X).$$

ν_n is said to *converge narrowly* to ν_0 if

$$(iii) \qquad \int k d\nu_n \to \int k d\nu_0, \quad \forall k \in C_b(X).$$

Since $M^1(X) \subset M(X)$, vague convergence can be defined on $M^1(X)$ as well. Thus $M^1(X)$ is provided with *four notions of convergence*, which imply each other according to the following diagram:

(convergence in norm) $\Rightarrow$ (narrow convergence) $\Rightarrow$ (weak convergence) $\Rightarrow$ (vague convergence).

6.8 Theorem. *Let $\mu_n \in M^1(X)$. Consider the following statements:*

(i) $\{\mu_n\}$ converges weakly.
(ii) $\{\mu_n\}$ converges narrowly.
(iii) There exist a constant c and a dense set $D \subset C_0(X)$ such that

$$\|\mu_n\|_{M^1} \leq c \quad \textit{and} \quad \int g d\mu_n \textit{ converges for every } g \in D.$$

(iv) For every $\epsilon > 0$, there exists a compact set K such that

$$|\mu_n|(K^c) < \epsilon \quad \textit{for all sufficiently large } n.$$

(v) Each μ_n is positive, $\{\mu_n\}$ converges weakly to μ, and $\int d\mu_n \to \int d\mu < +\infty$.

Then

$$\begin{aligned} (iii) &\Leftrightarrow (i), \\ (i) \text{ and } (iv) &\Leftrightarrow (ii), \\ (v) &\Rightarrow (ii). \end{aligned}$$

REMARK. To simplify the exposition, we prove only the direct implications, which are the easiest; these are practically the only ones used in what follows.

PROOF THAT (iii) $\Rightarrow$ (i). The family of linear functionals on $C_0(X)$

$$l_n(f) = \int f d\mu_n$$

satisfies

$$|l_n(f - f')| \leq c\|f - f'\|_{C_0}.$$

It is thus an equicontinuous family. Since it converges on a dense subset D, by Ascoli's theorem[3] it converges on all of C_0. Let $l_\infty(f) = \lim l_n(f)$. Using 6.6, we find that l_∞ is defined by a Radon measure $\mu_\infty \in M^1(X)$ and that $\{\mu_n\}$ converges weakly to μ_∞. □

[3] See Bourbaki, *General Topology*, X.2.5.

PROOF THAT (i) AND (iv) $\Rightarrow$ (ii). Let $f \in C_b(X)$. Let K be the compact set determined by (iv), and let φ denote a function with compact support such that $\varphi(x) = 1$ if $x \in K$. Then $f = f\varphi + u$, where supp $(u) \subset K^c$; hence

$$\int f d\mu_n = \int f\varphi d\mu_n + \int u d\mu_n.$$

By (i), the first integral converges to $\int \varphi f d\mu_\infty$, where $d\mu_\infty$ is the weak limit of $\{d\mu_n\}$. Moreover,

$$\left|\int u d\mu_n\right| \leq \|u\|_{C_b}|\mu_n|(K^c) \leq \epsilon\|f\|.\square$$

PROOF THAT (v) $\Rightarrow$ (ii). It will suffice to prove that (v) implies (iv). Given $\epsilon > 0$, let K be a compact subset of X such that $\mu(\overline{(K^c)}) < \epsilon$. Let f be a function with support contained in K such that $0 \leq f \leq 1$ and

$$\int f d\mu > \|\mu\| - \epsilon.$$

Let n_0 be such that, if $n > n_0$,

$$\left|\int d\mu_n - \int d\mu\right| < \epsilon \quad \text{and} \quad \left|\int f\ d\mu_n - \int f\ d\mu\right| < \epsilon.$$

Then

$$\mu_n(\overline{(K^c)}) \leq \|\mu_n\| - \int f\ d\mu_n, \quad \text{whence} \quad \mu_n(\overline{(K^c)}) < 3\epsilon \text{ if } n \geq n_0.$$

$\square$

6.9 Theorem. *Let X be a locally compact space and let $M^1_{d,f}(X)$ denote the finite linear combinations of the Dirac measure on X. Then, for any $\mu \in M^1(X)$, there exists a sequence $\{\mu_n\}$, $\mu_n \in M^1_{d,f}(X)$, such that $\{\mu_n\}$ converges narrowly to μ.*

PROOF. Let $\{\varphi_n\}$ be an increasing sequence of functions with compact support such that $0 \leq \varphi_n \leq 1$ and $\lim \varphi_n = 1$. Then $\|\varphi_n\mu - \mu\|_{M^1} \to 0$ by Lebesgue's dominated convergence theorem. Hence it suffices to prove the theorem when μ has compact support K. Let $\{O_{n,j} : j \in [1, s_n]\}$ be a finite cover of K by balls of radius $\frac{1}{n}$. Let $\tilde{A}_{n,1} = O_{n,1}$, $\tilde{A}_{n,2} = O_{n,2} \cap O^c_{n,1}$, and set $A_{n,q} = \tilde{A}_{n,q} \cap K$. Then each $A_{n,q}$ has diameter $< \frac{1}{n}$ and the $A_{n,q}$ form a partition of K. Restricting to $A_{n,q} \neq \emptyset$, choose $x_{n,q} \in A_{n,q}$.

μ_n is constructed by setting

$$\mu_n = \sum_q \mu(A_{n,q})\delta_{x_{n,q}}.$$

Let $f \in C_b(X)$; then

$$\int f d\mu = \sum \int_{A_{n,q}} f\ d\mu.$$

Since f is uniformly continuous on the compact set K, there exists a sequence $\{\eta_n\}$ which tends to zero as $n \to \infty$ and satisfies

$$|f(x) - f(x')| < \eta_n \quad \text{if} \quad d(x, x') < \frac{1}{n}.$$

Hence

$$\int \mathbf{1}_{A_{n,q}} f \, d\mu = f(x_{n,q}) \int \mathbf{1}_{A_{n,q}} d\mu \ + \ \theta_q \eta_n |\mu|(A_{n,q}),$$

where $|\theta_q| < 1$. Summing over q gives

$$\left| \int f \, d\mu - \int f \, d\mu_n \right| \le \eta_n |\mu|(K). \square$$

III

Fourier Analysis

Introduction

Fourier analysis can be illustrated by analogies from optics. Given a light beam, the goal of *spectral analysis* is to determine the *monochromatic* beams it contains; that is, the beams of the form $\exp(\frac{2i\pi}{\lambda}t)$. Once a spectral analysis has been carried out, one can ask whether the analysis is *exhaustive*: is all the energy of the beam really concentrated in the band of frequencies where the spectral analysis was done? One can also ask whether the beam can be reconstructed from its monochromatic components: can *spectral synthesis* be performed?

It is well known that quantum mechanics determines the possible energy levels of a system as the eigenvalues of a hermitian operator defined on a Hilbert space $\mathcal{H}$. More generally, given a system of pairwise-commuting hermitian operators, the eigenvalues of the system are the possible values of the associated "observables".

In the general setting of spectral theory, the problems of spectral analysis, conservation of energy, and spectral synthesis remain completely meaningful. Taking the space $L^2(\mathbf{R}^n)$ as a Hilbert space $\mathcal{H}$ and the hermitian operators generated by the translations as a family of operators, one naturally recovers Fourier analysis as a special case; what is more surprising is that general spectral theory can be obtained as a classical theorem of Fourier analysis, Bochner's theorem. This will be done in Appendix I.

Since differentiation operators on $L^2(\mathbf{R}^n)$ appear as limits of translation operators, Fourier analysis realizes their spectral decomposition as well.

Thus partial differential equations with constant coefficients are subject to the methods of real Fourier analysis (or complex Fourier analysis, but we will not pursue this point).

Studying the domains of definition of the Laplace operator and its iterates in $L^2(\mathbf{R}^n)$ leads to the construction of Sobolev scales, a theory that is stable under local diffeomorphisms and thus well suited to the local theory of partial differential equations with *variable* coefficients. In dealing with the theory of distributions, we use the approaches of Sobolev and Schwartz simultaneously. The chapter ends with the local inversion of elliptic operators with variable coefficients, by means of Calderon's theory of pseudo-differential operators.

1 Convolutions and Spectral Analysis on Locally Compact Abelian Groups

1.1 NOTATION. Let G be an abelian (commutative) group . The group operation will usually be written additively:

$$(g_1, g_2) \mapsto g_1 + g_2.$$

With this notation, the identity element will be denoted by 1 and the inverse of g by $-g$.

A locally compact abelian group is an abelian group which is given the structure of a locally compact topological space compatible with the group operation. That is, the mapping from $G \times G$ to G defined by

$$(i) \qquad (g_1, g_2) \mapsto g_1 - g_2$$

is continuous. It can be shown that a metrizable group G has a translation-invariant metric d; that is, d satisfies

$$(ii) \qquad d(g_0 + g, g_0 + g') = d(g, g').$$

1.2 Examples

1.2.0 The integers $\mathbf{Z}$ form a group under addition. Given the distance defined by $d(n, m) = |n - m|$, they form a locally compact group.

1.2.1 $\mathbf{R}^n$, with vector addition, is a locally compact group.

1.2.2 The one-dimensional torus

Let

$$\mathbf{T} = \{z \in \mathbf{C} : |z| = 1\}.$$

$\mathbf{T}$ is the set of complex numbers of modulus 1. (From the set-theoretic point of view, $\mathbf{T}$ is a circle.) $\mathbf{T}$ is given the group operation defined by the multiplication of complex numbers. If $z_1, z_2 \in \mathbf{T}$, then $z_1 z_2 \in \mathbf{T}$ and $z_1^{-1} = \bar{z}_1 \in \mathbf{T}$. Thus $\mathbf{T}$ is an abelian group; when endowed with the topology induced by $\mathbf{C}$, it is compact.

1.2.3 The n-dimensional torus

We denote by $\mathbf{T}^n$ the product of n copies of $\mathbf{T}$, endowed with the product topology and the product group operation.

1.2.4 A homomorphism from $\mathbf{R}$ onto $\mathbf{T}$

With $\theta \in \mathbf{R}$, we associate the element

$$u(\theta) = \mathrm{e}^{i\theta} \in \mathbf{T}.$$

Then $u(\theta + \theta') = u(\theta)u(\theta')$, i.e. u is a *homomorphism* of $\mathbf{R}$ onto $\mathbf{T}$. The kernel of u is

$$u^{-1}(1) = \{\theta : \mathrm{e}^{i\theta} = 1\} = 2\pi\mathbf{Z},$$

where $\mathbf{Z}$ is the subgroup of $\mathbf{R}$ consisting of the integers. Let $C(\mathbf{T})$ denote the functions defined and continuous on $\mathbf{T}$ and let $C_b(\mathbf{R})$ denote the bounded continuous functions on $\mathbf{R}$. Let u^* be the map from $C(\mathbf{T})$ into $C_b(\mathbf{R})$ defined by

$$(u^* f)(\theta) = f(u(\theta)), \quad \forall \theta \in \mathbf{R}.$$

Then the image of u^* consists of those functions $h \in C_b(\mathbf{R})$ that are *periodic* with period 2π; that is, functions satisfying

$$h(\theta + 2\pi) = h(\theta).$$

1.2.5 A homomorphism from $\mathbf{R}^n$ onto $\mathbf{T}^n$

With $x = (x_1, \ldots, x_n)$ we associate

$$v(x) = (\mathrm{e}^{ix_1}, \ldots, \mathrm{e}^{ix_n}).$$

The kernel of v is $2\pi\mathbf{Z}^n$. The operation

$$f \mapsto f \circ v = v^* f$$

maps $C(\mathbf{T}^n)$ onto the n-fold periodic functions on $\mathbf{R}^n$; that is, functions h satisfying

$$h(x + y) = h(x), \quad \forall y \in (2\pi\mathbf{Z})^n.$$

1.3 The group algebra

$M^1(G)$ denotes the Banach space of signed Radon measures on G which have finite total mass.

1.3.1 Discrete measures

Let δ_g denote the Dirac measure at the point g and let

$$M_d^1(G) = \left\{ \mu \in M^1(G) : \mu = \sum_{k=1}^{+\infty} \beta_k \delta_{g_k}, \quad \text{where} \quad \sum |\beta_k| < +\infty \right\}.$$

1.3.2 Convolution in $M_d^1(G)$

The convolution of two Dirac measures δ_{g_1} and δ_{g_2} is defined by

$$\delta_{g_1} * \delta_{g_2} = \delta_{g_1+g_2}.$$

That is, the convolution product is the Dirac measure at the point $g_1 + g_2$. This definition is extended to $M_d^1(G)$ by *bilinearity*. Given $\mu = \sum \beta_k \delta_{g_k}$ and $\mu' = \sum \beta'_k \delta_{g'_k}$ in M_d^1, we set

$$\mu * \mu' = \sum_{k,s} \beta_k \beta'_s \delta_{g_k + g'_k}.$$

Note that the convolution product is commutative, associative, and bilinear:

$$\mu * \mu' = \mu' * \mu,$$
$$(\mu * \mu') * \mu'' = \mu * (\mu' * \mu''),$$
$$(\mu + \nu) * \mu' = \mu * \mu' + \nu * \mu'.$$

Moreover,

$$\|\mu * \mu'\|_{M^1(G)} \le \sum_{k,s} |\beta_k \beta'_s| = \left(\sum_k |\beta_k| \right) \left(\sum_s |\beta'_s| \right) = \|\mu\| \ \|\mu'\|.$$

(Strict inequality can occur only if $g_k + g'_s = g_{k'} + g'_{s'}$, with $(k,s) \neq (k',s')$.) We would like to extend the convolution operator from $M_d^1(G)$ to all of M^1 by an explicit formula realizing this extension. Let $C_0(G)$ denote the continuous functions on G which vanish at infinity.

1.3.3 Fundamental lemma. *Let μ, $\mu' \in M_d^1(G)$ and let $\rho = \mu * \mu'$. Then*

$$\int_G f(z) d\rho(z) = \int_G \int_G f(x+y) d\mu(x) d\mu'(y), \quad \forall f \in C_0(G).$$

PROOF. The right-hand side, which we denote by II, can be written as

$$II = \sum_{k,k'} f(g_k + g'_k) \beta_k \beta'_k.$$

Grouping together all terms such that $g_k + g'_{k'} = g''_{k''}$, we obtain

$$II = \sum f(g''_{k''}) \sum_{g_k + g'_{k'} = g''_{k''}} \beta_k \beta'_{k'} = \int f(z) d\rho(z).$$

1.3.4 **Definition of the convolution product on $M^1(G)$.** Given ν, $\nu' \in M^1(G)$, we define a linear functional on $C_0(G)$ by setting

$$l(f) = \int_G \int_G f(x+y) d\nu(x) d\nu'(y).$$

This integral converges, since

$$|l(f)| \le \int_G \int_G |f(x+y)|\ |d\nu(x)|\ |d\nu'(y)| \le \|f\|_{C_0} \int_G \int_G d\lambda \otimes d\lambda',$$

where $\lambda = |\nu|$, $\lambda' = |\nu'|$. By Fubini,

$$\int_G \int_G d\lambda \otimes d\lambda' = \lambda(G)\lambda'(G) = \|\nu\|_{M^1} \|\nu'\|_{M^1},$$

and hence

(*i*) $$|l(f)| \le \|f\|_{C_0(G)} \|\nu\|_{M^1} \|\nu'\|_{M^1}.$$

Thus l is a linear functional on $C_0(G)$ which is continuous in the norm topology. By II-6.6, there exists a measure $\sigma \in M^1(G)$ such that $\int f d\sigma = l(f)$. We set $\sigma = \nu * \nu'$, and call σ the convolution product of ν and ν'.

1.3.5 Theorem (Properties of the convolution product). *Let G be a locally compact group and let $M^1(G)$ be the Banach space of finite Radon measures on G. The convolution product is defined on $M^1(G)$ by the formula*

(*i*) $$\int_G f(z) d\lambda(z) = \int_G \int_G f(x+y) d\nu(x) d\nu'(y), \quad \forall f \in C_0(G),$$

where ν, $\nu' \in M^1(G)$ and $\lambda = \nu * \nu'$.

It has the following properties.

(*ii*) $$\|\nu * \nu'\| \le \|\nu\|\ \|\nu'\|$$

(*iii*) $$\nu * \nu' = \nu' * \nu \quad \textit{(commutativity)}$$

(*iv*) $$(\nu * \nu') * \nu'' = \nu * (\nu' * \nu'') \quad \textit{(associativity)}$$

(*v*) $$(\nu + \nu') * \nu'' = \nu * \nu' + \nu * \nu'' \quad \textit{(linearity)}$$

Furthermore,

(vi) if $\{\nu_n\}$ and $\{\nu'_n\}$ converge narrowly to ν_0 and ν'_0, then $\nu_n * \nu'_n$ converges narrowly to $\nu_0 * \nu'_0$.

PROOF. Formula (ii) follows from 1.3.4(i). In order to prove (vi), note that the narrow convergence of ν_n and ν'_n and Fubini's theorem imply that $\nu_n \otimes \nu'_n$ converges narrowly to $\nu_0 \otimes \nu'_0$. Let $f \in C_b(G)$ and set $u(x, y) = f(x + y)$. Then $u \in C_b(G \times G)$, and

$$\lim \int_{G\times G} u \, d\nu_n \otimes d\nu'_n = \int_{G\times G} u \, d\nu_0 \otimes d\nu'_0$$

can be written as

$$\lim \int_G \int_G f(x+y) d\nu_n(x) \otimes d\nu'_n(y) = \int_G \int_G f(x+y) d\nu_0(x) \otimes d\nu'_0(y),$$
$$\forall f \in C_b(G).$$

Thus (vi) is proved.

The algebraic properties (iii), (iv), and (v) can be proved by passing to the limit and using (vi), since these properties hold on $M^1_d(G)$ by 1.3.2. By II-6.9, $M^1_d(G)$ is dense in the topology of narrow convergence on $M^1(G)$. (Or this could easily be proved directly.)

1.3.6 Support of the convolution product

If F_1 and F_2 are subsets of G, we set

$$F_1 + F_2 = \{g : g = g_1 + g_2 \ \text{ with } \ g_i \in F_i\}.$$

Proposition. *Let* $\nu_1, \nu_2 \in M^1(G)$. *Then* $supp\,(\nu_1 * \nu_2) \subset \overline{supp\,(\nu_1) + supp\,(\nu_2)}$.

PROOF. $\int \varphi(x+y) d\nu_1(x) d\nu_2(y) = 0$ if φ is zero on supp $(\nu_1) +$ supp (ν_2). □

Equality holds if both measures are positive.

1.4 The dual group. The Fourier transform on M^1

1.4.1 Characters

Let G be a locally compact abelian group and let $\mathbf{T}$ be the multiplicative group of complex numbers of modulus 1 considered in 1.2.2. A *character* on G is a mapping

$$\chi : G \to \mathbf{T}$$

such that

(i) χ is continuous, and
(ii) χ is a homomorphism: $\chi(g + g') = \chi(g)\chi(g')$.

1.4.2 The dual group

The set of characters of G is denoted by $\widehat{G}$, and is given a group structure by defining the product χ_3 of two characters χ_1 and χ_2 as follows:

$$\chi_3(g) = \chi_1(g)\chi_2(g), \quad \forall g \in G.$$

The inverse χ_4 of χ_1 is defined by the formula

$$\chi_4(g) = \overline{\chi_1(g)} = \frac{1}{\chi_1(g)}.$$

Thus $\widehat{G}$ is an abelian group. The identity element is the trivial character χ_0 defined by

$$\chi_0(g) = 1, \quad \forall g \in G.$$

1.4.3 The Fourier transform on $M^1(G)$

Given $\mu \in M^1(G)$, we assign to it a function defined on $\widehat{G}$ by

$$\widehat{\mu}(\chi) = \int_G \chi(g) d\mu(g).$$

$\widehat{\mu}$ is called the Fourier transform of μ.

1.4.4 Fundamental theorem (Trivialization of the convolution product). *Let μ, $\nu \in M^1(G)$. Then*

$$\widehat{\mu * \nu} = \widehat{\mu}\widehat{\nu};$$

that is, the Fourier transform maps the convolution product of measures to the usual product of functions.

PROOF. Let $\rho = \mu * \nu$. Then

$$\begin{aligned}
\int_G \chi(z) d\rho(z) &= \int_G \int_G \chi(x+y) d\mu(x) d\nu(y) \\
&= \int_G \int_G \chi(x)\chi(y) d\mu(x) d\nu(y) \\
&= \left(\int_G \chi(x) d\mu(x)\right)\left(\int_G \chi(y) d\nu(y)\right) \\
&= \widehat{\mu}(\chi)\widehat{\nu}(\chi). \square
\end{aligned}$$

The first equality follows from 1.3.5, the second from the identity $\chi(x+y) = \chi(x)\chi(y)$, and the third from Fubini's theorem.

REMARK. Let δ_0 denote the Dirac measure concentrated at 0. Then

$$\widehat{\delta_0}(\chi) = \chi(0) = 1, \quad \forall \chi \in \widehat{G}.$$

Moreover,

$$\delta_0 * \mu = \mu, \quad \forall \mu \in M^1(G);$$

that is, δ_0 is the identity element of the algebra $M^1(G)$.

1.5 Invariant measures. The space L^1

1.5.1 Translation-invariant measures

A measure $\mu \in M^1(G)$ is said to be *translation invariant* if

$$(i) \qquad \int f(g+g_0)d\mu(g) = \int f(g)d\mu(g), \quad \forall g_0 \in G.$$

1.5.2 Proposition. *Suppose that μ satisfies (i) and that G is compact. Then*

$$(ii) \qquad \widehat{\mu}(\chi) = 0 \quad \textit{for every nontrivial character.}$$

PROOF. Let χ be a nontrivial character. Then there exists $g_0 \in G$ such that $\chi(g_0) \neq 1$. Condition (i) can be written in the form

$$(iii) \qquad \delta_{g_0} * \mu = \mu.$$

(iv) Since G is compact, $\mu(G) < \infty$ and thus $\mu \in M^1(G)$. Under these conditions, 1.4.4 can be applied:

$$(\delta_{g_0} * \mu)^\wedge(\chi) = \widehat{\delta}_{g_0}(\chi)\widehat{\mu}(\chi) = \chi(g_0)\widehat{\mu}(\chi),$$

whence

$$\chi(g_0)\widehat{\mu}(\chi) - \widehat{\mu}(\chi) = 0 \Rightarrow \widehat{\mu}(\chi) = 0. \square$$

1.5.3 Corollary. *Suppose that G is a compact group, μ is a translation-invariant Radon measure on G, and $L^2(G;\mu)$ is the associated Hilbert space. Then any two distinct characters of G are orthogonal. If the measure μ is also normalized by the condition*

$$\int d\mu = 1,$$

then the characters of G form an orthonormal system.

PROOF. Given $\chi_1, \chi_2 \in \widehat{G}$, we evaluate

$$(\chi_1|\chi_2)_{L^2} = \int_G \chi_1(g)\overline{\chi_2(g)}d\mu(g).$$

$$\chi_1(g)\overline{\chi_2(g)} = \chi_1(g)(\chi_2(g))^{-1} = \chi_3(g),$$

where $\chi_3(g) \in \widehat{G}$. By 1.5.2, the integral $\int \chi_3(g)d\mu(g)$ is zero if χ_3 is not identically equal to 1, that is if $\chi_1 \neq \chi_2$. Finally, if μ is normalized,

$$\|\chi_1\|_{L^2}^2 = \int_G \chi_1(g)\overline{\chi_1(g)}d\mu(g) = \int_G d\mu(g) = 1. \square$$

1.5.4 Haar's theorem. *Let G be a locally compact abelian group. Then there exists a translation-invariant positive Radon measure μ_0 on G, and this measure is unique up to a multiplicative constant.*

REMARK. If μ_0 is an invariant measure and c is a positive constant, it is clear that $c\mu_0$ is an invariant measure.

We assume without proof this general theorem of Haar, and restrict ourselves to constructing invariant measures in the special cases of the groups $\mathbf{R}$, $\mathbf{T}$, and $\mathbf{Z}$.

1.5.5 Examples of Haar measure

(i) *Counting measure on* $\mathbf{Z}$

Let $\mathbf{Z}$ be the set of integers. Consider the measure μ_0 such that

$$\mu_0(\{n\}) = 1 \quad \text{for every} \quad n \in \mathbf{Z}.$$

Then μ_0 is translation invariant.

(ii) *Lebesgue measure on* $\mathbf{R}$

Let $\mathbf{R}$ be the additive group of real numbers. The Lebesgue measure μ_0 is translation invariant (II-4.2.6) and hence a Haar measure.

(iii) *Haar measure on* $\mathbf{T}$

Let $\varphi : \mathbf{R} \to \mathbf{T}$ be defined by setting

$$\varphi(\theta) = \mathrm{e}^{i\theta}.$$

Let a mapping $\sigma : \mathbf{T} \to \mathbf{R}$ be defined by

$$\sigma(\zeta) = \arg \zeta, \quad \text{where} \quad \arg \zeta \in [0, 2\pi).$$

Then $\sigma(\zeta)$ is a Borel mapping from $\mathbf{T}$ into $\mathbf{R}$. Set

$$\nu_0(A) = \frac{1}{2\pi}\mu_0(\sigma(A));$$

then ν_0 is a Borel measure on $\mathbf{T}$. Moreover,

$$\int_{\mathbf{T}} f d\nu_0 = \int_0^{2\pi} f(\mathrm{e}^{i\theta})\frac{d\theta}{2\pi} \quad \text{and} \quad \int d\nu_0 = 1.$$

Lemma. *The measure ν_0 is translation invariant.*

PROOF. Let $\theta_0 \in [0, 2\pi)$ and set

$$I_{\theta_0} = \int_0^{2\pi} f\left(\mathrm{e}^{i(\theta+\theta_0)}\right)\frac{d\theta}{2\pi} = \int_{\theta_0}^{\theta_0+2\pi} f\left(\mathrm{e}^{i\lambda}\right)\frac{d\lambda}{2\pi} = \int_{\theta_0}^{2\pi} + \int_{2\pi}^{2\pi+\theta_0}.$$

Setting $\lambda - 2\pi = u$ in the last integral yields

$$\int_{2\pi}^{\theta_0+2\pi} f(e^{i\lambda})\frac{d\lambda}{2\pi} = \int_0^{\theta_0} f(e^{iu})\frac{du}{2\pi},$$

whence

$$\int_{\mathbf{T}} f\ d\nu_0 = I_{\theta_0}.\square$$

Uniqueness of the Haar measure in (i) is clear. For case (iii), it will be proved in 2.2.8.

(iv) *The product structure*

The measures on $\mathbf{Z}^n$, $\mathbf{R}^n$, and $\mathbf{T}^n$ are the products of the Haar measures on each factor.

1.5.6 Notation

The Haar measure of the group G will be denoted by dg. If G is locally compact, this measure is defined up to a normalizing factor. If G is compact, the factor is chosen so that G has measure 1.

1.6 The space $L^1(G)$

1.6.1 Identification of $L^1(G)$ with a vector subspace of $M^1(G)$

We denote by $L^1(G)$ the space of functions integrable with respect to Haar measure on G, and define an injection

$$j : L^1(G) \to M^1(G)$$

by associating with the function $f \in L^1(G)$ the Radon measure

(i) $$\mu_f = f(g)dg.$$

1.6.2 The convolution product on $L^1(G)$

Proposition. *Let f, $h \in L^1(G)$ and let μ_f and μ_h be the Radon measures associated with them by 1.6.1(i). Then there exists $k \in L^1(G)$ such that*

(i) $$\mu_f * \mu_h = \mu_k \quad (L^1(G) \textit{ is a subalgebra of } M^1(G)).$$

k is defined by

(ii) $$k(g_0) = \int f(g_0 - g)h(g)dg = \int h(g_0 - g)f(g)dg,$$

where the two integrals converge almost everywhere in g_0 with respect to Haar measure. We write

$$k = f * h.$$

(iii) REMARK. Since the convolution product on $L^1(G)$ is the restriction of the product on $M^1(G)$, it satisfies the identities 1.3.5(ii) to (v).

PROOF. Let $\varphi \in C_0(G)$; then

$$\langle \varphi, \mu_f * \mu_h \rangle = \int_G \int_G \varphi(g_1 + g_2) f(g_1) h(g_2) dg_1 dg_2.$$

Using Fubini's theorem yields

$$\langle \varphi, \mu_f * \mu_h \rangle = \int_G h(g_2) dg_2 \left[\int_G \varphi(g_1 + g_2) f(g_1) dg_1 \right].$$

Set $g_1 = g_3 - g_2$ inside the brackets. Since dg_1 is invariant under translation, $dg_1 = dg_3$ *for fixed* g_2, whence

$$\langle \varphi, \mu_f * \mu_h \rangle = \int_G h(g_2) dg_2 \left[\int_G \varphi(g_3) f(g_3 - g_2) dg_3 \right].$$

Using Fubini again, we obtain

$$\langle \varphi, \mu_f * \mu_h \rangle = \int_G \varphi(g_3) dg_3 \left[\int_G h(g_2) f(g_3 - g_2) dg_2 \right].$$

Fubini's theorem implies that the integral in brackets converges for almost every g_3 and is an integrable function $k \in L^1(G)$. We have thus shown that

$$\langle \varphi, \mu_1 * \mu_2 \rangle = \int \varphi(g_3) k(g_3) dg_3. \square$$

1.6.3 The Fourier transform on L^1

The Fourier transform on L^1 is obtained by *restriction* from the Fourier transform on M^1 and thus is written

$$(i) \qquad \widehat{f}(\chi) = \int_G f(g)\chi(g) dg, \quad \forall \chi \in \widehat{G}.$$

Theorem 1.4.4, on the trivialization of the convolution product, gives by restriction

$$(ii) \qquad \widehat{f * h}(\chi) = \widehat{f}(\chi)\widehat{h}(\chi).$$

1.6.4 Bessel's inequality. *Let G be a compact abelian group and let $f \in L^2(G)$. Then $f \in L^1(G)$ and*

$$\|f\|^2_{L^2(G)} \geq \sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^2.$$

PROOF. Since $\mu(G) < +\infty$, I-9.6 implies that $L^1(G) \supset L^2(G)$. Moreover,

$$\widehat{f}(\chi) = (f|\overline{\chi})_{L^2}.$$

Let S be a finite subset of $\widehat{G}$, let V_S denote the vector subspace generated in $L^2(G)$ by $\{\chi : \chi \in S\}$, and let f_S denote the orthogonal projection of f onto V_S. Then $f = f_S + f - f_S$, where $f - f_S$ is orthogonal to f_S. Hence

$$\|f\|_{L^2}^2 = \|f_S\|_{L^2}^2 + \|f - f_S\|_{L^2}^2,$$

and therefore

$$\|f\|_{L^2}^2 \geq \|f_S\|_{L^2}^2.$$

But it follows from 1.5.3 that

$$f_S = \sum_{\chi \in S} \chi(f|\chi) = \sum_{\chi \in S} \widehat{f}(\chi^{-1})\chi \quad \text{and}$$

$$\|f_S\|_{L^2}^2 = \sum_{\chi \in S} |\widehat{f}(\chi^{-1})|^2. \square$$

1.7 The translation operator

1.7.1 The translation operator on $L^p(G)$

Given a function f defined on G and a fixed $g_0 \in G$, we denote by $\tau_{g_0} f$ the function defined by

$$(i) \qquad (\tau_{g_0} f)(g) = f(g - g_0).$$

By the translation invariance of dg, $f \in L^p(G)$ implies $(\tau_g f) \in L^p(G)$, and moreover

$$(ii) \qquad \|\tau_g f\|_{L^p} = \|f\|_{L^p}.$$

Furthermore,

$$(iii) \qquad \tau_{g_1} \circ \tau_{g_2} = \tau_{g_1 + g_2}.$$

We summarize the last identity by saying that $g \mapsto \tau_g$ is a *representation of G in $L^p(G)$*; that is, the mapping is a homomorphism of G into the group of linear automorphisms of $L^p(G)$. We define the translate of a set A by an element g_0 of G to be $\tau_{g_0}(A) = A + g_0$.

If u_A is the indicator function of the set A ($u_A(x) = 1$ if $x \in A$ and $u_A(x) = 0$ if $x \notin A$), then $\tau_{g_0}(u_A) = u_{\tau_{g_0}(A)}$.

1.7.2 Fundamental theorem (Trivialization of the translation operator on $L^1(G)$ under the Fourier transform). *Let $f \in L^1(G)$. Then*

$$\widehat{\tau_{g_0} f}(\chi) = \chi(g_0)\widehat{f}(\chi) \quad \forall \chi \in \widehat{G}.$$

PROOF.
$$\widehat{\tau_{g_0} f}(\chi) = \int_G f(g - g_0)\chi(g)dg.$$
The change of variables $g \mapsto g - g_0 = g'$ leaves the Haar measure invariant: $dg = dg'$. Making this change of variables gives
$$\int f(g - g_0)\chi(g)dg = \int f(g')\chi(g' + g_0)dg' = \chi(g_0)\int f(g')\chi(g')dg'. \square$$

1.7.3 Continuity of the translation operator

Let $C_K(G)$ denote the compactly supported continuous functions on G, with the norm
$$\|f\|_{C_K} = \max|f(g)|, \ g \in G.$$

Continuity theorem. *(i) Let $f \in C_K(G)$. Then the mapping from G to $C_K(G)$ defined by $g \mapsto \tau_g f$ is uniformly continuous.*
(ii) Similarly, let $u \in L^p(G)$, where $1 \le p < +\infty$. Then the mapping from G to $L^p(G)$ defined by $g \mapsto \tau_g u$ is uniformly continuous.

PROOF.(i) Since f is continuous and compactly supported, f is *uniformly* continuous. Given $\epsilon > 0$, there exists η such that
$$|f(g_1) - f(g_2)| < \epsilon \quad \text{if} \quad d(g_1, g_2) < \eta.$$
Hence
$$|\tau_{g_0}(f)(g) - \tau_{g_0'}(f)(g)| = |f(g-g_0) - f(g-g_0')| < \epsilon \quad \text{if} \quad d(g-g_0, g-g_0') < \eta.$$
But it follows from the invariance of the distance under translation (cf. 1.1(ii)) that $d(g - g_0, g - g_0') = d(g_0, g_0')$, whence
$$\|\tau_{g_0}(f) - \tau_{g_0'}(f)\|_{C_b} < \epsilon \quad \text{if} \quad d(g_0, g_0') < \eta. \square$$
(ii) We now consider the case where $u \in L^p$. Since $p < +\infty$, by II-3.5 there exists $f \in C_K(G)$ such that $\|f - u\|_{L^p} < \frac{\epsilon}{3}$. Let us write
$$\tau_g u - \tau_{g'} u = \tau_g f - \tau_{g'} f + \tau_{g'}(f - u) - \tau_g(f - u).$$
Using 1.7.1(ii),
$$\|\tau_g(f - u)\|_{L^p} = \|f - u\|_{L^p} < \frac{\epsilon}{3},$$
whence
$$\|\tau_g u - \tau_{g'} u\|_{L^p} < \frac{2}{3}\epsilon + \|\tau_g f - \tau_{g'} f\|_{L^p}.$$
Let $A = \text{supp}\,(f)$. Then
$$\text{supp}\,(\tau_g f - \tau_{g'} f) \subset \tau_g(A) \cup \tau_{g'}(A),$$
$$\text{meas}\,(\text{supp}\,(\tau_g f - \tau_{g'} f)) < 2 \text{ meas}\,(A),$$
$$\|\tau_g f - \tau_{g'} f\|_{L^p} \le \|\tau_g f - \tau_{g'} f\|_{C_b}(2 \text{ meas}\,(A))^{1/p}.$$
The right-hand side of the last inequality tends to zero as $d(g, g') \to 0$ by the first part of the theorem. $\square$

1.8 Extensions of the convolution product

In this section, we give other cases where formula 1.6.2 converges.

1.8.1 The convolution product and the dual pairing

Let $\widetilde{f}$ denote the function defined by

$$\widetilde{f}(g) = f(-g).$$

Formula 1.6.2(ii) can be written formally as

$$(i) \qquad k(g_0) = \langle \tau_{g_0}\widetilde{f}, h\rangle = \langle f, \tau_{g_0}\widetilde{h}\rangle.$$

Lemma. *Let $f \in L^p(G)$ and $h \in L^q(G)$, where $1 \leq p \leq +\infty$ and p and q are conjugate exponents. Then, for every $g_0 \in G$, the integral*

$$(ii) \qquad \int f(g_0 - g)h(g)dg$$

converges and defines a function $k(g_0)$ which is uniformly continuous and bounded and which satisfies

$$(iii) \qquad \|k\|_{C_b} \leq \|f\|_{L^p}\|h\|_{L^q}.$$

PROOF. By symmetry, we may assume that $p \leq q$; then, since p and q are conjugate, $1 \leq p \leq 2$.

Using (i), we have

$$|k(g)| = |\langle \tau_g\widetilde{f}, h\rangle| \leq \|\tau_g f\|_{L^p}\|h\|_{L^q} = \|f\|_{L^p}\|h\|_{L^q},$$

and moreover

$$|k(g_0) - k(g_1)| = |\langle \tau_{g_0}\widetilde{f} - \tau_{g_1}\widetilde{f}, h\rangle| \leq \|\tau_{g_0}\widetilde{f} - \tau_{g_1}\widetilde{f}\|_{L^p}\|h\|_{L^q}.$$

Since $p < +\infty$, it follows from 1.7.3(ii) that the first term tends to zero when $d(g_0, g_1) \to 0$. □

1.8.2 Theorem (Action of $M^1(G)$ on $L^p(G)$ $(1 \leq p \leq +\infty)$). *Let $\mu \in M^1(G)$ and let $f \in L^p(G)$. Then the integral*

$$(i) \qquad h(g_0) = \int f(g_0 - g)d\mu(g)$$

converges almost everywhere in g_0 with respect to Haar measure and defines a function in L^p. Furthermore,

$$(ii) \qquad \|h\|_{L^p} \leq \|f\|_{L^p}\|\mu\|_{M^1}.$$

PROOF. Let $f' = |f|$ and let $\mu' = |\mu|$. Let $u \in L^q$, $u \geq 0$, and consider the double integral

$$I = \int \int f'(g_0 - g) u(g_0) dg_0 \, d\mu'(g).$$

Choose a Borel representative of the equivalence class of f in $\mathcal{L}^p(G)$. For this fixed choice, $f'(g_0 - g)$ is a Borel function and hence measurable with respect to the product measure $dg_0 \otimes d\mu(g)$. Thus Fubini's theorem can be applied once we have shown the convergence of

$$\int d\mu'(g) \left[\int f'(g_0 - g) u(g_0) dg_0 \right].$$

By 1.8.1, the integral in brackets is convergent and bounded above by $\|f\|_{L^p} \|u\|_{L^q}$, whence

$$(iii) \qquad |I| \leq \|\mu\|_{M^1} \|f\|_{L^p} \|u\|_{L^q}.$$

Letting u equal the indicator function of a compact set K, it follows from Fubini's theorem that the integral (i) converges dg-almost everywhere on K. Since K is arbitrary, (i) converges dg-a.e. on G. Let $h(g)$ be the function thus obtained. By (iii),

$$\left| \int h(g) u(g) dg \right| \leq \|\mu\|_{M^1} \|f\|_{L^p} \|u\|_{L^q}, \quad \forall u \in L^q.$$

If $p > 1$, then $q < +\infty$ and we define a linear functional on $L^q(G)$ by

$$l(u) = \int h(g) u(g) dg.$$

This form is bounded, since $|l(u)| \leq C \|u\|_{L^q}$. By the duality theorem (IV-6.3), it follows that $h \in L^p$. If $p = 1$, take $u(g) = \operatorname{sign}(h(g))$ if $h(g) \neq 0$ and $u(g) = 0$ otherwise. Then (iii) implies that

$$\int |h(g)| dg \leq \|\mu\|_{M^1} \|f\|_{L^1} < +\infty.$$

Thus $h \in L^1$.

1.8.3 The translation operator as a convolution operator

Note that if $\mu = \delta_{g_1}$, then

$$\int f(g_0 - g) d\mu(g) = f(g_0 - g_1) = (\tau_{g_1} f)(g_0).$$

In particular, $\tau_0 f = f$. Thus the action of $M^1(G)$ on $L^1(G)$ is a generalization of the translation operator. More generally, if $\mu \in M_d^1$ (cf. 1.3.1), then $\mu = \sum \beta_k \delta_{g_k}$ and

$$\int f(g_0 - g) d\mu(g) = \left(\sum \beta_k \tau_{g_k} f\right)(g_0).$$

NOTATION. Let $\mu \in M^1(G)$ and $f \in L^p(G)$ $(1 \leq p < +\infty)$. We make the notational convention that

$$(\tau_\mu f)(g_0) = \int f(g_0 - g) d\mu(g).$$

Then

$$\|\tau_\mu(f)\|_{L^p} \leq \|\mu\|_{M^1} \|f\|_{L^p}.$$

1.9 Convergence theorem. *Let $\{\mu_n\}$ be a sequence of measures in $M^1(G)$ satisfying hypotheses (iii) and (iv) of Theorem II-6.8 and converging narrowly to ν. Then*

$$\|\tau_{\mu_n} f - \tau_\nu f\|_{L^p} \to 0, \quad \forall f \in L^p, \quad 1 \leq p < +\infty.$$

If in addition $f \in C_0(G)$, then $\|\tau_{\mu_n} f - \tau_\nu f\|_{C_0} \to 0$.

REMARK. Using the converse of Theorem II-6.8, it would suffice to assume that $\{\mu_n\}$ converges narrowly to ν. Because this converse was not proved, we prefer to give the rather awkward statement above.

PROOF. Since $p < +\infty$, we can find $h \in C_K(G)$ such that

$$\|f - h\|_{L^p} < \epsilon.$$

By hypothesis II-6.8(iii), $\sup \|\mu_n\|_{M^1} = c < +\infty$, whence

$$\|\tau_{\mu_n}(f - h)\|_{L^p} \leq \|\mu_n\|_{M^1} \|f - h\|_{L^p} \leq c\|f - h\|_{L^p}.$$

It thus suffices to show that

$$(i) \qquad \|\tau_{\mu_n} h - \tau_\nu h\|_{L^p} \to 0.$$

Hypothesis II-6.8(iv) implies that for every $\epsilon > 0$ there exists a compact set H such that, for sufficiently large n,

$$\int_{H^c} d|\mu_n| < \epsilon \quad \text{and} \quad \int_{H^c} d|\nu| < \epsilon.$$

Let φ be an element of $C_K(G)$ such that supp $(\varphi) \subset K_1$ and $\varphi = 1$ on H. Set

$$\mu_n' = \varphi\mu_n, \quad \nu' = \varphi\nu, \quad \mu_n'' = (1 - \varphi)\mu_n, \quad \nu'' = (1 - \varphi)\nu.$$

Then $\|\nu''\| < \epsilon$, $\|\mu_n''\| < \epsilon$, and the proof is reduced to proving (i) for μ_n' and ν'. Furthermore, since μ_n' converges narrowly to ν', it suffices to show that (i) holds when the μ_n are supported in a fixed compact set K_1. Let K_2 be the support of h; then the support of $\tau_{\mu_n} h$ lies in $K_3 = K_1 + K_2$.

But K_3 is a compact subset of G. Moreover, by the definition of narrow convergence, for every fixed g

$$\int h(g - g')d\mu_n(g') \to \int h(g - g')d\nu(g').$$

That is, $u_n(g) = (\tau_{\mu_n} h - \tau_\nu h)(g)$ satisfies $u_n(g) \to 0$ everywhere. It follows from the bound $|\tau_{\mu_n} h| \le c\|h\|_{C_b} \mathbf{1}_{K_3}$ that

$$|u_n(g)| \le 2c\|h\|_{C_b} \mathbf{1}_{K_3} = c_1 \mathbf{1}_{K_3}.$$

Hence, by Lebesgue's dominated convergence theorem,

$$\int |u_n(g)|^p dg \to 0.\square$$

If $f \in C_0(G)$, we now determine $h \in C_K(G)$ by the condition $\|h - f\|_{C_0} < \epsilon$. As above, we reduce the proof to showing the result when the μ_n are supported in a fixed compact set K_3. Setting $\widetilde{h}(\xi) = h(-\xi)$, we write

$$\int h(g - \xi)d\mu_n(\xi) = \int (\tau_g \widetilde{h})(\xi)d\mu_n(\xi).$$

The mapping $\Phi : g \to \tau_g \widetilde{h}$ from G to $C_0(G)$ is continuous. Hence the image under Φ of the compact set K_3 is a compact set $\widetilde{H} \subset C_0(G)$. By hypothesis II-6.8(iii), there exists a constant c such that $\|\mu_n\|_{M^1} < c$. Consider the functions u_n defined on $\widetilde{H}$ by

$$u_n(y) = \int y(\xi)d\mu_n(\xi), \quad y \in \widetilde{H}.$$

Since $\|\mu_n\|_{M^1} < c$, these functions are equicontinuous. By the definition of narrow convergence,

$$u_n(y) \to \int y(\xi)d\nu(\xi), \quad \forall y \in C_0(G).$$

Since the functions u_n are equicontinuous and converge for every $y \in \widetilde{H}$, the compactness of $\widetilde{H}$ implies that they converge uniformly. $\square$

1.9.1 Corollary. *Let $\{\mu_n\}$ be a sequence of measures which converge narrowly to δ_0 and satisfy hypotheses (iii) and (iv) of II-6.8. Then $\|\tau_{\mu_n} f - f\|_{L^p} \to 0$.*

1.9.2 Corollary. *$L^p(G)$ is an $M^1(G)$-module; that is,*

(i) $$(\tau_\mu \circ \tau_\nu)(f) = \tau_{\mu * \nu}(f) = (\tau_\nu \circ \tau_\mu)(f).$$

In particular, if $g_0 \in G$,

(ii) $$\tau_{g_0}(\tau_\mu f) = \tau_\mu(\tau_{g_0} f) = \tau_{\mu * \delta_{g_0}}(f).$$

PROOF. It suffices to verify (i) in the case of discrete measures, where everything is obvious; the general case follows from the narrow density of $M^1_d(G)$ in $M^1(G)$ combined with Theorem 1.9.

For (ii), we note that

$$(iii) \qquad \tau_{g_0} u = \delta_{g_0} * u,$$

where δ_{g_0} denotes the Dirac measure at g_0, and use (i). □

2 Spectral Synthesis on $\mathbf{T}^n$ and $\mathbf{R}^n$

In Section 1 we introduced the Fourier transform, defined on the dual group $\widehat{G}$. We were not concerned with whether the dual group of G contained other elements than the trivial character, everywhere equal to 1. If $\widehat{G}$ were trivial, Fourier transform theory would have a very limited scope. We now exhibit the characters on $\mathbf{T}^n$ and $\mathbf{R}^n$ and use them to prove the injectivity of the Fourier transform. In certain cases, we will be able to characterize its image and give an explicit inversion formula.

2.1 The character groups of $\mathbf{R}^n$ *and* $\mathbf{T}^n$

(i) *The characters on* $\mathbf{R}$ *are of the form*

$$\chi_t(x) = e^{itx}, \quad \textit{where} \quad t \in \mathbf{R}, \quad t \textit{ fixed.}$$

Hence $\widehat{\mathbf{R}} = \mathbf{R}$.

PROOF. It is clear that an imaginary exponential satisfies the equation $\mathrm{e}^{it(x+y)} = \mathrm{e}^{itx}\mathrm{e}^{ity}$ and is a complex number of modulus 1. What must be proved is the converse. Let $x \mapsto \chi(x)$ be a character of $\mathbf{R}$; then, since $\chi(0) = 1$ and a is continuous, there exists an interval $[-a, a]$ such that

$$\mathrm{Re}(\chi(x)) > 0 \quad \text{if} \quad x \in [-a, a].$$

Hence we can define a function $l(x)$ without ambiguity by

$$\log \chi(x) = il(x), \quad \frac{\pi}{2} < l(x) < \frac{\pi}{2}, \quad x \in [-a, a].$$

Then $l(x)$ is continuous and

$$l(x+y) = l(x) + l(y) \quad \text{if} \quad x,\ y, \text{ and } x + y \in [-a, a].$$

It follows from this equation that $l(mx) = ml(x)$ if m is an integer such that $|mx| \le a$, and similarly that $l\left(\frac{y}{m}\right) = \frac{1}{m} l(y)$ if $|y| \le a$. Hence $l(ra) = rl(a)$ for every rational number r such that $|r| \le 1$.

By continuity, $l(xa) = xl(a)$ if $x \in \mathbf{R}$, $|x| \le 1$. Hence

$$\chi(y) = \mathrm{e}^{i\alpha y}, \quad \text{where} \quad \alpha = \frac{1}{a} l(a) \text{ and } |y| \le a.$$

For any $y_1 \in \mathbf{R}$ there exists an integer m such that $y_1 = my$ with $|y| \leq a$; thus

$$\chi(y_1) = (\chi(y))^m = \mathrm{e}^{i\alpha ym} = \mathrm{e}^{i\alpha y_1}.\square$$

(ii) *The characters on $\mathbf{R}^n$ are of the form*

$$\chi_t(x) = \exp\left(i\sum_{k=1}^{n} t_k x^k\right),$$

where $x = (x^1, \ldots, x^n) \in \mathbf{R}^n$ and $t = (t_1, \ldots, t_n) \in \mathbf{R}^n$. Hence $\widehat{\mathbf{R}}^n = \mathbf{R}^n$.

PROOF. The imaginary exponentials are obviously characters. It must be shown that every character is of this form. Let $e_k = (0, 0, 1, 0, \ldots)$ be the kth element of the canonical basis of $\mathbf{R}^n$. Then $\lambda \mapsto \lambda e_k$ is a homomorphism from $\mathbf{R}$ to $\mathbf{R}^n$ and hence $\lambda \mapsto \chi(\lambda e_k)$ is a character on $\mathbf{R}$. By (i), we can write

$$\chi(\lambda e_k) = \mathrm{e}^{it_k\lambda}.$$

Writing $x = \sum_k x^k e_k$, it follows that

$$\chi(x) = \prod_k \chi(x^k e_k) = \prod \mathrm{e}^{it_k x^k}.\square$$

(iii) *The characters on $\mathbf{T}^n$ are of the form*

$$\chi_m(\theta) = \exp\left(i\sum_{k=1}^{n} m_k \theta^k\right),$$

where $m = (m_1, \ldots, m_n) \in \mathbf{Z}^n$ and $(e^{i\theta^1}, \ldots, e^{i\theta^n}) \in \mathbf{T}^n$. Hence $\widehat{\mathbf{T}^n} = \mathbf{Z}^n$.

PROOF. The numbers $(\theta^1, \ldots, \theta^n)$ are each defined only up to a multiple of 2π; this indeterminacy has no effect on the value of $\chi_m(\theta)$ since $m \in \mathbf{Z}^n$, and thus $\chi_m(\theta)$ is indeed a character on $\mathbf{T}^n$.

Conversely, let χ be a character on $\mathbf{T}^n$. We define (cf. 1.2.5) a homomorphism $v : \mathbf{R}^n \to \mathbf{T}^n$ by setting $v(x) = \left(\mathrm{e}^{ix^1}, \ldots, \mathrm{e}^{ix^n}\right)$. Then $\chi \circ v$ is a character on $\mathbf{R}^n$ and hence, by (ii), is of the form

$$\chi(v(x)) = \exp\left(i\sum t_k x^k\right).$$

Suppose that $v(x) = 1$. Then $\chi(v(x)) = 1$; hence

$$\sum_k t_k x^k \equiv 0 \text{ modulo } 2\pi.$$

Setting x equal successively to $2\pi e_1, 2\pi e_2, \ldots, 2\pi e_n$ shows that $t_1, t_2, \ldots, t_n \in \mathbf{Z}$.
$\square$

2.2 Spectral synthesis on $\mathbf{T}$

2.2.1 The Poisson kernel

Given a number $r \in [0,1)$, the *Poisson kernel* on $\mathbf{T}$ is the function defined by the series

$$(i) \qquad P_r(\theta) = \sum_{n\in\mathbf{Z}} r^{|n|} e^{in\theta}.$$

Not only is this series uniformly convergent, but its sum can be calculated:

$$P_r(\theta) = \sum_{n=0}^{+\infty} (re^{i\theta})^n + \sum_{p=1}^{+\infty} (re^{-i\theta})^p.$$

Using the formula for the sum of a geometric series, we obtain

$$P_r(\theta) = \frac{1}{1-re^{i\theta}} + \frac{re^{-i\theta}}{1-re^{-i\theta}}.$$

Thus

$$(ii) \qquad P_r(\theta) = \frac{1-r^2}{1-2r\cos\theta + r^2} = \frac{1-r^2}{(1-r\zeta)(1-r\overline{\zeta})}, \quad \text{where} \quad \zeta = e^{i\theta}.$$

2.2.2 Proposition. *Let $e = 1$ denote the identity element of $\mathbf{T}$ and let $d\nu(\zeta)$ denote the Haar measure on $\mathbf{T}$ defined in 1.5.4. Then*

$$P_r(\zeta)d\nu(\zeta) \to \delta_e \quad \textit{narrowly as} \quad r \to 1$$

and, moreover, satisfies hypotheses (iii) and (iv) of II-6.8.

PROOF. Let f be a continuous periodic function, with period 2π. We must show that

$$\text{(i)} \qquad \int_{-\pi}^{+\pi} P_r(\theta) f(\theta) \frac{d\theta}{2\pi} \to f(0) \quad \text{as} \quad r \to 1.$$

(ii) Note that, by 2.2.1(ii), $P_r(\theta) > 0$.

Integrating the uniformly convergent series 2.2.1(i) term by term shows that $\displaystyle\int_{-\pi}^{+\pi} P_r(\theta)\frac{d\theta}{2\pi} = 1$. Hence, since $P_r(\theta) > 0$,

$$\text{(iii)} \qquad \|P_r\|_{L^1} = 1.$$

(iv) For fixed $\eta > 0$, $\max_{\eta \le |\theta| \le \pi} P_r(\theta) = P_r(\eta)$, which approaches zero as $r \to 1$.

Set $f_1(\theta) = f(\theta) - f(0)$. Since

$$\int_{-\pi}^{+\pi} f(0) P_r(\theta) \frac{d\theta}{2\pi} = f(0),$$

it suffices to show that

$$\int_{-\pi}^{+\pi} f_1(\theta)P_r(\theta)\frac{d\theta}{2\pi} \to 0.$$

Let $\epsilon > 0$ be given. Then there exists η such that $|f_1(\theta)| < \frac{\epsilon}{2}$ if $|\theta| < \eta$. Fixing η, we split the integral in two:

$$\int_{-\pi}^{+\pi} f_1(\theta)P_r(\theta)\frac{d\theta}{2\pi} = \int_{-\eta}^{+\eta} f_1(\theta)P_r(\theta)\frac{d\theta}{2\pi} + \int_{[-\eta,+\eta]^c\cap[-\pi,+\pi]} f_1(\theta)P_r(\theta)\frac{d\theta}{2\pi}.$$

The first integral is bounded above by $\frac{\epsilon}{2}\|P_r\|_{L^1}$, which equals $\frac{\epsilon}{2}$ by (iii), and the second by $\|f\|_{C(\mathbf{T})}P_\epsilon(\eta)$, which approaches zero by (iv).

Hypothesis (iv) of II-6.8 clearly holds since $\mathbf{T}$ is compact, and hypothesis (iii) since $P_r d\nu$ has total mass 1. □

(v) Corollary. *Let $d\mu_r$ denote the measure on $M^1(\mathbf{T})$ defined by $P_r(\theta)d\theta/2\pi$.*
If $f \in L^p(\mathbf{T})$ $(1 \le p < +\infty)$, then $\|\tau_{\mu_r}f - f\|_{L^p} \to 0$ as $r \to 1$.
If $f \in C(\mathbf{T})$, then $\|\tau_{\mu_r}f - f\|_{C(\mathbf{T})} \to 0$.

PROOF. By 2.2.2 and 1.9.1.

2.2.3 Proposition. *Let $f \in L^1(\mathbf{T})$ and let $\widehat{f}(n)$, $n \in \mathbf{Z}$, be its Fourier transform. Then*

$$(f * P_r)(\theta) = \sum_{m\in\mathbf{Z}} \widehat{f}(m)r^{|m|}e^{-im\theta}.$$

PROOF.

$$\begin{aligned}(f * P_r)(\theta) &= \int_{-\pi}^{+\pi} f(\varphi)P_r(\theta-\varphi)\frac{d\varphi}{2\pi}\\ &= \int_{-\pi}^{+\pi} f(\varphi)\sum_{n\in\mathbf{Z}} r^{|n|}e^{in(\theta-\varphi)}\frac{d\varphi}{2\pi}.\end{aligned}$$

The uniformly convergent series $\sum_{n\in\mathbf{Z}} r^{|n|}e^{in(\theta-\varphi)}$ can be integrated term by term, giving

$$\begin{aligned}(f * P_r)(\theta) &= \sum_{n\in\mathbf{Z}} r^{|n|}e^{in(\theta)}\int_{-\pi}^{\pi} e^{-in(\varphi)}f(\varphi)\frac{d\varphi}{2\pi}\\ &= \sum_{n\in\mathbf{Z}} r^{|n|}e^{in(\theta)}\widehat{f}(-n).\end{aligned}$$

The result follows by setting $-n = m$. □

2.2.4 Spectral synthesis theorem.

(i) Let $f \in L^p(\mathbf{T})$, $1 \le p < +\infty$. Set

$$g_r(\theta) = \sum_{n\in\mathbf{Z}} r^{|n|}\widehat{f}(n)e^{-in\theta}.$$

Then $\|f - g_r\|_{L^p} \to 0$ *as* $r \to 1$.

(ii) Let $f \in C(\mathbf{T})$. *Then* $\|f - g_r\|_{C(\mathbf{T})} \to 0$ *as* $r \to 1$.

REMARK. Since g_r is defined in terms of the Fourier transform of f, the theorem shows that f can be reconstructed from its Fourier transform.

PROOF. By 2.2.2(v) and 2.2.3.

2.2.5 Theorem on conservation of energy. *Let* $f \in L^2(\mathbf{T})$. *Then*

(i)
$$\|f\|^2_{L^2(\mathbf{T})} = \sum_{n\in\mathbf{Z}} |\widehat{f}(n)|^2 \quad \textit{and}$$

(ii)
$$\left\| f(\theta) - \sum_{n=-p}^{p} \widehat{f}(n)e^{-in\theta} \right\|_{L^2} \to 0 \quad \textit{as} \quad p \to +\infty.$$

Plancherel's theorem. *Let* $\ell^2(\mathbf{Z})$ *denote the set of sequences such that* $\sum |a_n|^2 < +\infty$.

(iii) The mapping $f \to \widehat{f}$ *defines an isometric isomorphism from* $L^2(\mathbf{T})$ *onto* $\ell^2(\mathbf{Z})$.

PROOF. Since the characters on $\mathbf{T}$ are mutually orthogonal,

$$\|g_r\|^2_{L^2} = \sum r^{2|n|} |\widehat{f}(n)|^2.$$

By Bessel's inequality,

$$\|f\|^2_{L^2} \geq \sum_{n\in\mathbf{Z}} |\widehat{f}(n)|^2.$$

For a proof by contradiction, assume that the inequality is *strict.* Since $\|f - g_r\|_{L^2} \to 0$ by 2.2.4, $\|g_r\|_{L^2} \to \|f\|_{L^2}$. Hence

$$\lim_{r\uparrow 1} \sum_{n\in\mathbf{Z}} r^{2|n|} |\widehat{f}(n)|^2 > \sum_{n\in\mathbf{Z}} |\widehat{f}(n)|^2,$$

a contradiction; Bessel's inequality is in fact an equality and (i) is proved.

Let V_p denote the vector subspace of L^2 generated by those $\mathrm{e}^{in\theta}$ for which $-p \leq n \leq p$. Then (cf. 1.6.4 and 1.5.3) the orthogonal projection of f onto V_p can be written as

$$s_p(\theta) = \sum_{|n|\leq p} \widehat{f}(n)\mathrm{e}^{-in\theta}.$$

By the Pythagorean theorem,

$$\|f - s_p\|^2_{L^2} + \|s_p\|^2_{L^2} = \|f\|^2_{L^2},$$

whence

$$\|f - s_p\|^2_{L^2} = \|f\|^2_{L^2} - \|s_p\|^2_{L^2} = \sum_{|n|>p} |\widehat{f}(n)|^2,$$

where the second equality follows from (i). Since the last expression tends to zero, (ii) is proved.

To prove (iii), let $U : L^2(\mathbf{T}) \to \ell^2(\mathbf{Z})$ be defined by $U(f) = \{\widehat{f}(n)\}$. Then U is an isometry by (i). It follows that the image of U is a complete subspace of $l^2(\mathbf{Z})$ and hence is closed.

Let $W = \{\{a_n\} \in \ell^2 : a_n = 0$ except for finitely many $n\}$. The function that maps $\{a_n\} \in W$ to the trigonometric polynomial $\sum a_n \mathrm{e}^{-in\theta}$ is continuous, since the sum is finite. Because the function is continuous, it lies in L^2; thus $U(L^2) \supset W$. Since W is dense in ℓ^2 and $U(L^2)$ is closed, we conclude that $U(L^2) = \ell^2$. □

2.2.6 The Fourier inversion formula

If we are given $\widehat{f}$ and want to evaluate the function f at a point, the only result at our disposal so far is 2.2.4(ii). The drawback of this formula is that it involves a double limit: we must first sum a series, then let r tend to 1.

We would like to obtain results on the convergence of the partial sums of the Fourier series of f, that is the sums

$$s_p(\theta) = \sum_{n=-p}^{n=+p} \widehat{f}(n)\mathrm{e}^{-in\theta}.$$

Theorem 2.2.5(ii) is a convergence theorem for the L^2 norm.

Lennart Carleson showed in 1965 that the partial sums of the Fourier series of a function f in $L^2(\mathbf{T})$ converge almost everywhere to f. He thus resolved a problem that had remained open for fifty years. The following is an elementary result.

2.2.7 Fourier inversion theorem. *Let $f \in L^1(\mathbf{T})$. Assume*

(i) $$\sum |\widehat{f}(n)| < +\infty.$$

Then

(ii) $$f(\theta) = \sum_{n\in\mathbf{Z}} \widehat{f}(n)e^{-in\theta} \quad \text{for almost every } \theta.$$

If f is also continuous, equality holds everywhere.

PROOF. Set

$$g_r = f * P_r = \sum \widehat{f}(n)r^{|n|}\mathrm{e}^{-in\theta} \quad \text{and} \quad \varphi(\theta) = \sum_{n\in\mathbf{Z}} \widehat{f}(n)\mathrm{e}^{-in\theta}.$$

Then $\varphi \in C(\mathbf{T})$ since the series converges uniformly. We now show that

(iii) $$\|\varphi - g_r\|_{C(\mathbf{T})} \to 0.$$

Clearly

$$\|\varphi - g_r\|_{C(\mathbf{T})} \le \sum_n |\widehat{f}(n)|(1 - r^{|n|}).$$

Given ϵ, fix p so that $\sum_{|n|>p} |\widehat{f}(n)| < \epsilon/2$. Then $\sum_{|n|\le p} |\widehat{f}(n)|(1 - r^{|n|})$ is the sum of $2p+1$ terms, each of which tends to zero. This proves (iii).

It follows from the inequality $\|g_r - \varphi\|_{L^1} \le \|g_r - \varphi\|_{C(\mathbf{T})}$ that $\lim_{r\to 0} \|g_r - \varphi\|_{L^1} = 0$. By 2.2.4,

$$\|f - \varphi\|_{L^1} = 0.$$

Thus f and φ are equal a.e., and (ii) is proved.

Suppose that f is continuous; then, since φ is continuous, so is $f - \varphi = u$. If u were not identically 0, $\{u \ne 0\}$ would contain an interval, contradicting (ii); hence $u = 0$ everywhere. $\square$

(iv) REMARK. As an element of L^1, f is defined only up to a set of measure zero. (ii) means that the equivalence class of f under the relation of equality almost everywhere contains a continuous function, namely φ. It is reasonable to take this continuous function as a representative of the equivalence class of f.

2.2.8 Density of the trigonometric polynomials

A finite linear combination of exponentials is called a trigonometric polynomial.

Proposition. *The trigonometric polynomials are dense in the normed spaces $L^p(\mathbf{T})$ $(1 \le p < +\infty)$ and $C(\mathbf{T})$.*

PROOF. Since $C(\mathbf{T})$ is dense in L^p by II-3.5, it suffices to prove density in $C(\mathbf{T})$, recalling that $\|\ \|_{L^p} \le \|\ \|_{C(\mathbf{T})}$.

Let $h \in C(\mathbf{T})$ and let $\epsilon > 0$ be given. Using 2.2.4(ii), fix r such that $\|h - h_r\|_{C(\mathbf{T})} < \frac{\epsilon}{2}$. Decompose h_r as

$$h_r(\theta) = \sum_{|n|\le p} \widehat{h}(n) r^{|n|} \mathrm{e}^{-in\theta} + \sum_{|n|>p} \widehat{h}(n) r^{|n|} \mathrm{e}^{-in\theta}.$$

Note that $|\widehat{h}(n)| \le \|h\|_{L^1(\mathbf{T})} \le \|h\|_{C(\mathbf{T})}$; this implies

$$\left| \sum_{|n|>p} \widehat{h}(n) r^{|n|} \mathrm{e}^{-in\theta} \right| \le \|h\|_{C(\mathbf{T})} \frac{2r^{p+1}}{1-r}.$$

Since r is fixed, this expression is less than $\frac{\epsilon}{2}$ for sufficiently large p. Thus

$$\left\| h - \sum_{|n|\le p} \widehat{h}(n) r^{|n|} \chi_{-n} \right\| < \epsilon. \square$$

Corollary (Injectivity of the Fourier transform on measures). *Let μ, $\nu \in M^1(\mathbf{T})$ satisfy*

$$\widehat{\mu}(n) = \widehat{\nu}(n) \quad \text{if} \quad n \in \mathbf{Z}.$$

Then $\mu = \nu$.

PROOF. Let Q be a trigonometric polynomial. By linearity, $\int Q d\nu = \int Q d\mu$. Since the trigonometric polynomials are dense in $C(\mathbf{T})$, it follows that

$$\int f d\nu = \int f d\mu \quad \forall f \in C(\mathbf{T}).\square$$

Corollary (Uniqueness of Haar measure on T). *Let ρ be a Haar measure on $\mathbf{T}$. Then there exists a constant c such that $\rho = c\frac{d\theta}{2\pi}$.*

PROOF. By 1.5.2, $\widehat{\rho}(n) = 0$ if $n \neq 0$. It thus suffices to use the preceding corollary. $\square$

2.3 Extension of the results to $\mathbf{T}^n$

The Poisson kernel is defined on $\mathbf{T}^n$ by

$$P_r(\zeta) = \prod_{k=1}^{n} P_r(\zeta^k), \quad \zeta = (\zeta^1, \ldots, \zeta^n) \in \mathbf{T}^n.$$

Since the Haar measure $d\nu(\zeta) = d\nu(\zeta^1) \otimes \cdots \otimes d\nu(\zeta^n)$ is a product measure,

$$P_r(\zeta) d\nu(\zeta) = P_r(\zeta^1) d\nu(\zeta^1) \otimes \cdots \otimes P_r(\zeta^n) d\nu(\zeta^n).$$

By 2.2.2 each term converges narrowly to δ_{e_k}; hence $P_r(\zeta)d\nu(\zeta)$ converges narrowly to δ_e.

It can be shown as in 2.2.3 that, for all $f \in L^1(\mathbf{T}^n)$,

2.3.1
$$(f * P_r)(\theta) = \sum_{m \in \mathbf{Z}^n} \widehat{f}(m) r^{\|m\|} e^{-im.\theta},$$

where $\|m\| = |m_1| + |m_2| + \cdots + |m_n|$ and $m.\theta = \sum_{k=1}^{n} m_k \theta^k$. The following theorems are proved as in 2.2.

2.3.2 Spectral synthesis theorem. *Let $f \in L^p(\mathbf{T})$ $(1 \leq p < +\infty)$. Set $g_r(\theta) = \sum \widehat{f}(m) r^{\|m\|} e^{-im.\theta}$. Then*

$$\|f - g_r\|_{L^p(\mathbf{T}^n)} \to 0.$$

If $f \in C(\mathbf{T})$, then

$$\|f - g_r\|_{C(\mathbf{T})} \to 0.$$

2.3.3 Theorem on conservation of energy. *Let $f \in L^2(\mathbf{T}^n)$. Then*

(*i*)
$$\|f\|_{L^2}^2 = \sum_{m \in \mathbf{Z}^n} |\widehat{f}(m)|^2.$$

Set $s_p(\theta) = \sum_{m \in S_p} \widehat{f}(m) e^{-im.\theta}$, where $S_p = \{m \in \mathbf{Z}^n : |m_k| \leq p \ \forall k\}$. Then

(*ii*)
$$\|f - s_p\|_{L^2(\mathbf{T}^n)} \to 0 \quad as \quad p \to +\infty.$$

(iii) (Plancherel) The mapping $f \to \widehat{f}$ is a bijection of $L^2(\mathbf{T}^n)$ onto $\ell^2(\mathbf{Z}^n)$.

2.3.4 Fourier inversion theorem. *Let $f \in L^1(\mathbf{T}^n)$. Suppose that*

(*i*)
$$\sum_{m \in \mathbf{Z}^n} |\widehat{f}(m)| < +\infty.$$

Then

(*ii*)
$$f(\theta) = \sum_{m \in \mathbf{Z}^n} \widehat{f}(m) e^{-im.\theta} \quad \text{for almost every } \theta.$$

(iii) If f is continuous, equality holds everywhere.

2.4 Spectral synthesis on $\mathbf{R}$

2.4.0 Regularity of the Fourier transform on $\mathbf{R}^n$

Let $\mu \in M^1(\mathbf{R}^n)$. Its Fourier transform is defined by

$$\widehat{\mu}(t) = \int_{\mathbf{R}^n} e^{it.x} d\mu(x).$$

2.4.0.1 Proposition. *The Fourier transform $\widehat{\mu}(t)$ is a bounded continuous function and*

(*i*)
$$\|\widehat{\mu}\|_{C_b(\mathbf{R}^n)} \leq \|\mu\|_{M^1}.$$

PROOF. Set $\mu = k|\mu|$ with $k \in L^1_\mu$. Then

$$\widehat{\mu}(t_n) = \int e^{it_n.x} k(x) d|\mu|(x).$$

If the sequence $\{t_n\}$ converges to t_0, the sequence of functions $\{e^{it_n.x}k(x)\}$ converges everywhere to $e^{it_0.x}k(x)$. Since it is bounded in modulus by $1 \in L^1_{|\mu|}$, Lebesgue's dominated convergence theorem implies that $\widehat{\mu}(t_n) \to \widehat{\mu}(t_0)$. Finally,

$$|\widehat{\mu}(t)| \leq \int d|\mu| = |\mu|(\mathbf{R}^n) = \|\mu\|_{M^1}.\square$$

2.4.0.2 Theorem (Lebesgue). *If $f \in L^1(\mathbf{R}^n)$, then its Fourier transform $\widehat{f}(t) = \int_{\mathbf{R}^n} e^{it.x} f(x)dx$ is a continuous function that vanishes at infinity, and*

$$(i) \qquad \|\widehat{f}\|_{C_0(\mathbf{R}^n)} \leq \|f\|_{L^1(\mathbf{R}^n)}.$$

PROOF. Since $f(x)dx \in M^1(\mathbf{R}^n)$, the only new property to be proved is that

$$\widehat{f}(t) \to 0 \quad \text{as} \quad \|t\| \to +\infty.$$

Let $\epsilon > 0$ be given. Since the translation operator is continuous on $L^1(\mathbf{R}^n)$, there exists η such that

$$(ii) \qquad \|\tau_y f - f\|_{L^1} < \epsilon \quad \text{if} \quad \|y\| < \eta.$$

It follows from the property $\widehat{\tau_y f}(t) = e^{iy.t}\widehat{f}(t)$ that $(\tau_y f - f)^\wedge(t) = (e^{iy.t} - 1)\widehat{f}(t)$.

Using (i) and (ii),

$$(iii) \qquad |(e^{iy.t} - 1)\widehat{f}(t)| < \epsilon \quad \text{if} \quad \|y\| < \eta.$$

If t satisfies $\|t\| > \pi\eta^{-1}$, we can find y such that $y.t = \pi$ and $\|y\| < \eta$. Hence, by (iii),

$$2|\widehat{f}(t)| < \epsilon \quad \text{if} \quad \|t\| > \pi\eta^{-1}.\square$$

2.4.1 Dilations and the Fourier transform

A dilation on $\mathbf{R}$ is multiplication by a positive number λ:

$$x \mapsto \lambda x \quad \forall x \in \mathbf{R}, \quad \lambda \text{ fixed, } \lambda > 0.$$

Given a function u defined on $\mathbf{R}$, let

$$(i) \qquad u_\lambda(x) = \lambda^{-1}u(\lambda^{-1}x).$$

Take $u \in L^1(\mathbf{R})$ and set $\lambda^{-1}x = y$. Then $\int u_\lambda(x)dx = \int u(y)dy$. In particular,

$$(ii) \qquad \|u_\lambda\|_{L^1} = \|u\|_{L^1}.$$

Similarly, again setting $\lambda^{-1}x = y$,

$$(iii) \qquad \widehat{u}_\lambda(t) = \int u_\lambda(x)e^{itx}dx = \int u(y)e^{it\lambda y}dy = \widehat{u}(\lambda t).$$

2.4.2 Lemma. *Let $u \in L^1(\mathbf{R})$ and assume that $\int u(x)dx = 1$. Then, as $\lambda \to 0$, $u_\lambda(x)dx$ converges narrowly to the Dirac measure at 0 and satisfies hypotheses (iii) and (iv) of Theorem II-6.8.*

PROOF. Let $f \in C_b(\mathbf{R})$ and set $f_1(x) = f(x) - f(0)$. Then

$$\int u_\lambda(x) f(x) dx = f(0) \int u_\lambda(x) dx + \int f_1(x) u_\lambda(x) dx.$$

Since the first integral on the right-hand side equals 1, it suffices to show that the second tends to zero. Setting $\lambda^{-1}x = y$, we can write this integral as $\int f_1(\lambda y) u(y) dy$. Fix A so that $\int_{|y|>A} u(y) dy < \frac{\epsilon}{2} \|f_1\|_{C_b}^{-1}$. Then

$$\begin{aligned}
\int f_1(\lambda y) u(y) dy &= \int_{|y|\le A} + \int_{|y|>A}, \\
\left| \int_{|y|>A} \right| &\le \|f_1\|_{C_b} \int_{|y|>A} |u(y)| dy < \frac{\epsilon}{2}, \quad \text{and} \\
\left| \int_{|y|\le A} \right| &\le \max_{|t|\le \lambda A} |f_1(t)| \; \|u\|_{L^1}.
\end{aligned}$$

Since A is fixed, $\lambda A \to 0$ as $\lambda \to 0$. Since $f_1(0) = 0$ and f_1 is continuous, the last expression will be less than $\frac{\epsilon}{2}$ for λ sufficiently small. □

2.4.3 Proposition. *For every $\mu > 0$,*

$$(i) \qquad \exp\left(-\frac{t^2\mu}{2}\right) = \frac{1}{(2\mu\pi)^{1/2}} \int_{\mathbf{R}} \exp\left(-\frac{x^2}{2\mu}\right) e^{itx} dx.$$

PROOF. Cf. IV-4.3.2(ii), where this formula is proved for $\mu = 1$. The general case is obtained by applying 2.4.1(iii).

2.4.4 Proposition. *Set*

$$G_\mu(x) = \frac{1}{(2\mu\pi)^{1/2}} \exp\left(-\frac{x^2}{2\mu}\right).$$

Then, as $\mu \to 0$, $G_\mu(x)dx$ satisfies the conclusions of 2.4.2.

PROOF. It follows from 2.4.3(i), with $t = 0$ and $\mu = 1$, that

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

It now suffices to apply 2.4.2. □

2.4.5 Spectral synthesis theorem. *Let $f \in L^1(\mathbf{R})$, let $\widehat{f}$ be its Fourier transform, and set*

$$(i) \qquad g_\mu(x) = \int_{\mathbf{R}} e^{-itx} \widehat{f}(t) \exp\left(-\frac{t^2\mu}{2}\right) \frac{dt}{2\pi}.$$

If $f \in L^1 \cap L^p$ $(1 \leq p < +\infty)$, then

$$(ii) \qquad \|f - g_\mu\|_{L^p} \to 0 \quad \text{as} \quad \mu \to 0.$$

Remark. We must assume that $f \in L^1$, since otherwise the integral defining the Fourier transform $\widehat{f}$ does not converge. Moreover, since $\|\widehat{f}\|_{L^\infty} \leq \|f\|_{L^1}$, this assumption implies the convergence of the integral defining g_μ.

Proof. By 2.4.2 and 1.9,

$$(iii) \qquad \|f * G_\mu - f\|_{L^p} \to 0.$$

Furthermore, since G_μ is an even function,

$$(f * G_\mu)(x) = \int G_\mu(y - x) f(y) dy.$$

An integral expression for $G_\mu(x)$ is obtained by interchanging t and x, writing μ^{-1} for μ, and multiplying by $\frac{1}{(2\mu\pi)^{1/2}}$ in 2.4.3(i). Substituting this into the integral above yields

$$(f * G_\mu)(x) = \int_{\mathbf{R}} f(y) \left[\int_{\mathbf{R}} \exp\left(-\frac{t^2\mu}{2}\right) \mathrm{e}^{-it(x-y)} \frac{dt}{2\pi}\right] dy.$$

The hypothesis $f \in L^1$ implies the convergence of the double integral

$$\int\int_{\mathbf{R}^2} \exp\left(-\frac{t^2\mu}{2}\right) |f(y)| \, dy \, dt.$$

Hence Fubini's theorem can be applied; reversing the order of integration gives

$$(f * G_\mu)(x) = \int_{\mathbf{R}} \exp\left(-\frac{t^2\mu}{2}\right) \mathrm{e}^{-itx} \left[\int_{\mathbf{R}} \mathrm{e}^{ity} f(y) dy\right] \frac{dt}{2\pi}.$$

Recognizing the quantity in brackets as $\widehat{f}(t)$, we have shown that

$$(iv) \qquad (f * G_\mu)(x) = \int_{\mathbf{R}} \widehat{f}(t) \exp\left(-\frac{t^2\mu}{2}\right) \mathrm{e}^{-itx} \frac{dt}{2\pi}, \quad \forall f \in L^1.$$

Now (iii) and (iv) imply (ii). □

2.4.6 Fourier inversion theorem. *Let $f \in L^1(\mathbf{R})$. Suppose that*

$$(i) \qquad \widehat{f} \in L^1(\mathbf{R}).$$

Then

$$\text{(ii)} \qquad f(x) = \int_{\mathbf{R}} e^{-itx} \widehat{f}(t) \frac{dt}{2\pi} \quad \textit{for almost every } x.$$

(iii) If f is also continuous, equality holds everywhere in (ii).

PROOF. Let g_μ be defined as in 2.4.5(i). Then, as $\mu \to 0$, the integrand in 2.4.5(i) tends everywhere to $\mathrm{e}^{-itx}\widehat{f}(t)$. Furthermore, it is dominated by the function $|\widehat{f}(t)| \in L^1$. By Lebesgue's dominated convergence theorem,

$$g_\mu(x) \to \int \mathrm{e}^{-itx} \widehat{f}(t) \frac{dt}{2\pi} = \varphi(x).$$

Next, since $\|f - g_\mu\|_{L^1} \to 0$, we can extract a subsequence μ_k such that

$$f(x) = \lim g_{\mu_k}(x) \quad \text{almost everywhere.}$$

This implies (ii).

To prove (iii), note that $\varphi(x)$ is continuous by 2.4.0.2. Thus $\varphi(x) - f(x) = u(x)$ is continuous. By the same reasoning as in 2.2.7, $u(x) = 0$ a.e. $\Rightarrow u(x) \equiv 0$. □

In the next section, we will study the space of those functions f to which the Fourier inversion formula applies.

2.4.7 The Wiener algebra $A(\mathbf{R})$

Let

$$A(\mathbf{R}) = \{f \in L^1(\mathbf{R}) : \widehat{f} \in L^1(\mathbf{R})\}.$$

It follows from 2.4.6(ii) that the equivalence class (for equality almost everywhere) of every f in $A(\mathbf{R})$ contains a continuous function. From now on, we will take this function as the representative of f. Thus the Wiener algebra is contained in the Banach space of continuous functions.

The Fourier inversion formula can be applied to f if and only if $f \in A(\mathbf{R})$. We set $\|f\|_{A(\mathbf{R})} = \|f\|_{L^1} + \|\widehat{f}\|_{L^1}$.

(i) $f \in A(\mathbf{R})$ is equivalent to $\widehat{f} \in A(\mathbf{R})$.

PROOF. By the Fourier inversion theorem,

$$f(x) = \int_{\mathbf{R}} \widehat{f}(t) \mathrm{e}^{-itx} \frac{dt}{2\pi}.$$

Set $f(-x) = u(x)$. Then $u(x) = \int \widehat{f}(t)\mathrm{e}^{itx} \frac{dt}{2\pi}$; that is,

$$\text{(ii)} \qquad u = (\widehat{f})^\wedge.$$

Hence

$$(\widehat{f})^\wedge \in L^1 \Leftrightarrow u \in L^1 \Leftrightarrow f \in L^1.$$

(iii) If $f \in A(\mathbf{R})$, then $f \in C_0(\mathbf{R})$ and $\|f\|_{C_0(\mathbf{R})} \leq \|f\|_{A(\mathbf{R})}$.

PROOF. By the inversion formula and 2.4.0.1.

(iv) If $f \in A(\mathbf{R})$, then $f \in L^p$ $\forall p$ $1 \leq p \leq +\infty$.

PROOF. $\displaystyle\int |f|^p dx \leq \|f\|_{C_0}^{p-1} \|f\|_{L^1}$.

(v) If f, $h \in A(\mathbf{R})$, then $f * h \in A(\mathbf{R})$.

PROOF. $\|f * h\|_{L^1} \leq \|f\|_{L^1} \|h\|_{L^1}$ and $(f * h)^\wedge = \widehat{f}\,\widehat{h}$, whence

$$\|(f * h)^\wedge\|_{L^1} = \|\widehat{f}\,\widehat{h}\|_{L^1} \leq \|\widehat{f}\|_{L^\infty} \|\widehat{h}\|_{L^1} \leq \|f\|_{L^1} \|\widehat{h}\|_{L^1}.$$

Thus $f * h \in A(\mathbf{R})$.

(vi) Let f, $h \in A(\mathbf{R})$. Then $(fh)^\wedge = \widehat{f} * \widehat{h}$ and $fh \in A(\mathbf{R})$.

PROOF. By (ii),

$$(fh)(-x) = (\widehat{f} * \widehat{h})^\wedge(x).$$

By (i) and (v), $\widehat{f} * \widehat{h} \in A(\mathbf{R})$. The inversion formula can be applied, and

$$\int (fh)(-x) \mathrm{e}^{-itx} \frac{dx}{2\pi} = (\widehat{f} * \widehat{h})(t).$$

Hence, replacing x by $-x$, we see that $(fh)^\wedge = \widehat{f} * \widehat{h} \in A(\mathbf{R})$; by (i), $fh \in A(\mathbf{R})$.

(vii) $A(\mathbf{R})$ is dense in L^p, $1 \leq p < +\infty$. $A(\mathbf{R})$ is dense in $C_0(\mathbf{R})$.

PROOF. Let L_K^p denote the L^p functions which are zero a.e. outside a compact set. Then L_K^p is dense in L^p. Let $h \in L_K^p$. Set $h_n = h * G_{n^{-1}}$, where G_μ was defined in 2.4.4; then $\|h_n - h\|_{L^p} \to 0$.

We now show that $h_n \in A(\mathbf{R})$. Let K be a compact set such that $h(x) = 0$ a.e. on K^c. By Hölder's inequality,

$$\|h\|_{L^1} \leq \left[\int_K dx\right]^{1/q} \|h\|_{L^p},$$

where p and q are conjugate exponents. Thus $h \in L^1$ and $h_n \in L^1$. Moreover,

$$|\widehat{h}_n(t)| = \left|\widehat{h}(t) \exp\left(-\frac{t^2}{2n}\right)\right| \leq \|h\|_{L^1} \exp\left(-\frac{t^2}{2n}\right),$$

whence $\widehat{h}_n \in L^1$ and $h_n \in A(\mathbf{R})$.

If $h \in C_K(\mathbf{R})$, then $h_n = h * G_{n^{-1}} \in C_0(\mathbf{R})$, $h_n \in A(\mathbf{R})$, and $\|h - h_n\|_{C_0} \to 0$.
□

2.4.8 Theorem on conservation of energy. *Let $f \in L^1 \cap L^2$. Then*

$$\|f\|_{L^2}^2 = (2\pi)^{-1}\|\widehat{f}\|_{L^2}^2.$$

PROOF. Let $f \in L^1 \cap L^2$. Set $f_n = f * G_{n^{-1}}$. Then

$$(i) \qquad f_n \in L^1 \quad \text{and} \quad \widehat{f_n}(t) = \exp\left(-\frac{t^2}{2n}\right)\widehat{f}(t).$$

The Fourier inversion theorem can be applied to f_n, giving

$$f_n(x) = \int_{\mathbf{R}} \widehat{f_n}(t)\mathrm{e}^{-itx}\frac{dt}{2\pi}.$$

Replace $f_n(x)$ by this expression in the scalar product:

$$(f|f_n)_{L^2} = \int_{\mathbf{R}} f(x)\overline{f_n(x)}dx = \int_{\mathbf{R}} f(x)\left[\int_{\mathbf{R}} \overline{\widehat{f_n}(t)}\mathrm{e}^{itx}\frac{dt}{2\pi}\right]dx.$$

Since $f \in L^1$ and $\widehat{f_n} \in L^1$, the double integral converges and, applying Fubini's theorem, we can reverse the order of integration:

$$(f|f_n)_{L^2} = \int_{\mathbf{R}} \overline{\widehat{f_n}(t)}\left[\int_{\mathbf{R}} f(x)\mathrm{e}^{itx}dx\right]\frac{dt}{2\pi} = \int_{\mathbf{R}} \widehat{f_n}(t)\overline{\widehat{f_n}(t)}\frac{dt}{2\pi}.$$

Let $n \to \infty$; then, by 2.4.5, $\|f_n - f\|_{L^2} \to 0$, and the left-hand side thus tends to $\|f\|_{L^2}^2$. Using (i) on the right-hand side, we obtain

$$\lim_{n\to\infty}\int |\widehat{f}(t)|^2 \exp\left(-\frac{t^2}{n}\right)\frac{dt}{2\pi} = \|f\|_{L^2}^2.$$

The sequence $\left\{\exp\left(-\frac{t^2}{n}\right)\right\}$ is increasing. Applying the theorem of Fatou-Beppo Levi shows that $|\widehat{f}(t)|^2$ is integrable and that

$$\int |\widehat{f}(t)|^2\frac{dt}{2\pi} = \|f\|_{L^2}^2.\square$$

2.4.9 Plancherel's extension theorem. *The Fourier transform has an extension*

$$(i) \qquad U : L^2(\mathbf{R}) \to L^2(\mathbf{R}).$$

(ii) $(2\pi)^{-\frac{1}{2}}U$ is an isometric mapping of $L^2(\mathbf{R}) \to L^2(\mathbf{R})$.
(iii) U is a continuous bijection of $L^2(\mathbf{R}) \to L^2(\mathbf{R})$.
(iv) The inverse of U is given by

$$U^{-1}(h) = \frac{1}{2\pi}\overline{U(\overline{h})}.$$

PROOF. Consider the mapping $u : f \mapsto \widehat{f}$, from $V = L^1 \cap L^2$ to L^2. Then, by 2.4.8,

$$\text{(v)} \qquad \|u(v)\|_{L^2}^2 = 2\pi \|v\|_{L^2}^2, \quad \forall v \in V.$$

Hence u is a uniformly continuous mapping into the complete space L^2. It thus has an extension to the closure of $L^1 \cap L^2$ in L^2, which is just L^2. Moreover, 2.4.8 extends by continuity and gives (ii). In particular, U is injective. It remains to prove (iii) and (iv). By 2.4.7(iv),

$$A(\mathbf{R}) \subset L^1(\mathbf{R}) \cap L^2(\mathbf{R}).$$

Hence, by 2.4.7(i),

$$U(L^1 \cap L^2) \supset U(A(\mathbf{R})) = \widehat{A(\mathbf{R})} = A(\mathbf{R}), \quad \text{whence}$$

(vi) $U(L^1 \cap L^2)$ is *dense* in L^2 by 2.4.7(vii).

Next, since $(2\pi)^{-1}U$ is an isometry, the image of L^2 is a complete, hence closed, subspace of L^2. Thus (vi) implies that U is *surjective.* Finally, the inverse mapping of U is, up to a factor of 2π, an isometry. It follows from (v) that it is determined by its restriction to $A(\mathbf{R})$. The restriction is given by the Fourier inversion formula, and can be written as

$$f(x) = \frac{1}{2\pi} \int_{\mathbf{R}} \mathrm{e}^{-itx} \widehat{f}(t) dt = \frac{1}{2\pi} \overline{\int_{\mathbf{R}} \mathrm{e}^{itx} \overline{\widehat{f}(t)} dt} = \frac{1}{2\pi} \overline{U(\overline{\widehat{f}})}.$$

This expression for U^{-1} on a dense set is valid everywhere, since U^{-1} is continuous. □

(vii) REMARK. What is striking in Plancherel's theorem is that it gives an isomorphism of spaces. Thus a problem posed in L^2 is equivalent under the Fourier transform to another problem posed in L^2.

2.5 *Spectral synthesis on* $\mathbf{R}^n$

We now generalize the results of the last section to $\mathbf{R}^n$. Let

$$G_\mu(x) = \frac{1}{(2\mu\pi)^{n/2}} \exp\left(-\frac{1}{2\mu} \|x\|^2 \right),$$

where $\|x\|^2 = (x^1)^2 + \ldots + (x^n)^2$. Then $G_\mu(x) = \prod_{k=1}^n G_\mu(x^k)$.

By (2.4.4), $G_\mu(x^k)dx^k$ converges narrowly in $M^1(\mathbf{R})$ to the Dirac measure at zero. When $\mu \to 0$, $\mu > 0$, we find that

$$G_\mu(x)dx = \otimes G_\mu(x^k)dx^k$$

converges narrowly to the Dirac measure at zero in $M^1(\mathbf{R}^n)$. Moreover, by (2.4.3),

$$G_\mu(x) = \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} \exp\left(-\frac{\mu\|t\|^2}{2}\right) \mathrm{e}^{it.x} dt.$$

Spectral synthesis theorem. *Let $f \in L^1(\mathbf{R}^n)$ and set*

$$g_\mu(x) = f * G_\mu(x) = \int_{\mathbf{R}^n} e^{-it.x} \widehat{f}(t) \exp\left(-\frac{\mu\|t\|^2}{2}\right) \frac{dt}{(2\pi)^n}.$$

Then $\|f - g_\mu\|_{L^1} \to 0$. If, in addition, $f \in L^1 \cap L^p$ $(1 \le p < +\infty)$, then $\|f - g_\mu\|_{L^p} \to 0$.

Fourier inversion theorem. *Let*

$$A(\mathbf{R}^n) = \{f \in L^1(\mathbf{R}^n) : \widehat{f} \in L^1(\mathbf{R}^n)\}.$$

Then $A(\mathbf{R}^n)$ is dense in $L^p(\mathbf{R}^n)$, $1 \le p < +\infty$, and in $C_0(\mathbf{R}^n)$. Furthermore, almost everywhere in x (with equality everywhere if f is continuous),

$$f(x) = \int_{\mathbf{R}^n} \widehat{f}(t) e^{-it.x} \frac{dt}{(2\pi)^n}, \quad \forall f \in A(\mathbf{R}^n).$$

Plancherel's extension theorem. *There exists a bijective mapping U of $L^2(\mathbf{R}^n)$ onto $L^2(\mathbf{R}^n)$ such that*

$$\|U(f)\|_{L^2} = (2\pi)^{n/2}\|f\|_{L^2} \quad \text{and} \quad U(f) = \widehat{f}, \quad \forall f \in L^1 \cap L^2.$$

Moreover,

$$U^{-1}(h) = \frac{1}{(2\pi)^n} \overline{U(\overline{h})}.$$

The proofs of these results are identical to those already given for the case where $n = 1$. We end this section with a new result.

2.6 Parseval's lemma. *Let $f \in A(\mathbf{R}^n)$ and let $\mu \in M^1(\mathbf{R}^n)$. Then*

$$\int_{\mathbf{R}^n} f(x) d\mu(x) = \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} \widehat{f}(t)\widehat{\mu}(-t) dt.$$

PROOF. The Fourier inversion theorem,

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} \widehat{f}(t) \mathrm{e}^{-it.x} dt,$$

can be used to write f as a function of $\widehat{f}$ on the left-hand side of the assertion of the lemma. Since $\widehat{f} \in L^1$, Fubini's theorem can be applied to the resulting double integral. We obtain

$$\int_{\mathbf{R}^n} f(x) d\mu(x) = \int_{\mathbf{R}^n} \widehat{f}(t) dt \left(\int_{\mathbf{R}^n} \mathrm{e}^{-it.x} d\mu(x)\right) \times \frac{1}{(2\pi)^n}. \square$$

Corollary. *Let μ, $\mu' \in M^1(\mathbf{R}^n)$ such that $\widehat{\mu}(t) = \widehat{\mu}'(t)$. Then $\mu = \mu'$.*

PROOF. For all $f \in A(\mathbf{R}^n)$,

$$\int_{\mathbf{R}^n} f(x)d\mu(x) = \int_{\mathbf{R}^n} f(x)d\mu'(x).$$

Since $A(\mathbf{R}^n)$ is dense in $C_0(\mathbf{R}^n)$, $\mu' = \mu$. □

3 Vector Differentiation and Sobolev Spaces

3.1 Differentiation in the vector sense. The spaces W_s^p

The goal of this section is to interpret the notion of derivative in terms of translation operators. The advantage of this point of view is that, since the Fourier transform realizes the spectral analysis of translation operators, the same will be true for differentiation operators.

Given $f \in L^p(\mathbf{R}^n)$ and $a \in \mathbf{R}^n$, we say that *the derivative of f in the direction of a exists in the L^p sense* and equals $D_a f$ if, when $\epsilon \to 0$, $\lim \epsilon^{-1}(\tau_{\epsilon a} f - f)$ *exists in $L^p(\mathbf{R}^n)$* and equals $-D_a f$.

We then have

$$\|D_a f + \epsilon^{-1}(\tau_{\epsilon a} f - f)\|_{L^p} \to 0.$$

Let

$$W_1^p = \{f \in L^p(\mathbf{R}^n) : D_a f \text{ exists in the } L^p \text{ sense for every } a \in \mathbf{R}^n\}.$$

Decomposing $a = a^1 e_1 + \ldots + a^n e_n$ with respect to the canonical basis of $\mathbf{R}^n$, we write $D_a f = \sum a^k D_{e_k} f$ if $f \in W_1^p$. Given an integer $s > 1$, we define

$$W_s^p = \{f \in W_1^p : D_a f \in W_{s-1}^p \quad \forall f \in \mathbf{R}^n\}.$$

If $f \in W_s^p$, $D_{a_1} D_{a_2} \ldots D_{a_s} f$ is defined recursively.

3.1.1 Theorem (Spectral analysis of differentiation operators). *Let $f \in W_1^1$. Then*

$$\widehat{D_a f}(t) = -i(a.t)\widehat{f}(t).$$

PROOF. $D_a f \in L^1$, and hence $\widehat{D_a f}$ is well defined. Since the convergence occurs in L^1, the order of integration in the following expression can be reversed:

$$\widehat{D_a f}(t) = \int \lim_{\epsilon \to 0} \epsilon^{-1}(\tau_{\epsilon a} f - f)(x) e^{ix.t} dx = \lim_{\epsilon \to 0} \epsilon^{-1}(\widehat{\tau_{\epsilon a} f} - \widehat{f})(t).$$

By 1.7.2,

$$-\widehat{D_a f}(t) = \lim_{\epsilon \to 0} \frac{(e^{i\epsilon a.t} - 1)}{\epsilon} \widehat{f}(t) = i(a.t)\widehat{f}(t).\square$$

3.1.2 Corollary. *If $f \in W_s^1$, then*

$$(D_{a_1} D_{a_2} \dots D_{a_s} f)^\wedge(t) = \prod_{k=1}^{s} (-i(a_k.t))\widehat{f}(t).$$

3.1.3 Theorem. *If $f \in W_s^1$, then*

$$\widehat{f}(t) = o(\|t\|^{-s}) \quad as \quad \|t\| \to \infty.$$

PROOF. $D_a^s \in L^1$. By 2.4.0.2, $\widehat{D_a^s f}$ tends to zero at infinity. Hence $|a \cdot t|^s \widehat{f}(t)$ tends to zero at infinity, and this is true for every fixed a. □

3.1.4 Corollary. $W_{n+1}^1 \subset A = \{f \in L^1 : \widehat{f} \in L^1\}$.

PROOF. Since $\widehat{f}(t) = o(\|t\|^{-n-1})$ and $\widehat{f} \in C_0$, it follows that $\widehat{f} \in L^1$. □

3.1.5 Proposition. *Let $\mu \in M^1(\mathbf{R}^n)$ be a finite measure and let $f \in W_1^p$ (where $1 \leq p < +\infty$). Then $\tau_\mu f \in W_1^p$ and*

$$D_a(\tau_\mu f) = \tau_\mu(D_a f).$$

PROOF. $\tau_a \tau_\mu f = \tau_\mu \tau_a f$ and $\epsilon^{-1}(\tau_{\epsilon a} - I)\tau_\mu f = \tau_\mu[\epsilon^{-1}(\tau_{\epsilon a} - I)f]$.

Since τ_μ is a bounded operator on L^p, the convergence of the right-hand side implies the convergence of the left-hand side. □

3.2 The space $\mathcal{D}(\mathbf{R}^n)$

3.2.0 **Definition.** Let $\mathcal{D}(\mathbf{R}^n)$ denote the space of infinitely differentiable functions on $\mathbf{R}^n$ with compact support. We show that $\mathcal{D}(\mathbf{R}^n)$ contains functions that are not identically zero. Let

$$\begin{aligned} f(r) &= \exp\left(\tfrac{-1}{1-r}\right) && \text{if} \quad 0 < r < 1 \\ &= 0 && \text{if} \quad r \leq 1. \end{aligned}$$

Set

$$(i) \qquad \widetilde{F}(x) = f(\|x\|^2), \quad \text{where} \quad \|x\|^2 = (x^1)^2 + \dots + (x^n)^2.$$

Then $\widetilde{F}$ is infinitely differentiable. Since $\widetilde{F} \geq 0$ on $\mathbf{R}^n$ and $\widetilde{F} > 0$ on a nonempty open set, $\int \widetilde{F}(x)dx > 0$. Let $F(x) = \alpha \widetilde{F}(x)$, where the constant α is determined so that $\int F(x)dx = 1$. Then, setting

$$(ii) \qquad F_\lambda(x) = \lambda^{-n} F(\lambda^{-1}x),$$

it follows from 2.4.2 that $F_\lambda(x)dx \to \delta_0$ narrowly.

3.2.1 Proposition. *If $\varphi \in \mathcal{D}$, then $\varphi \in W_s^p$ $(1 \leq p \leq +\infty)$ for every positive integer s. In particular, 3.1.2 holds. Furthermore,*

$$D_a\varphi = \sum a^k \frac{\partial \varphi}{\partial x^k}.$$

PROOF. We use Taylor's formula with integral remainder:

$$[-\epsilon^{-1}(\tau_{\epsilon a} - 1)\varphi + D_a\varphi](x) = \int_0^\epsilon \sum_k a_k \left(\frac{\partial \varphi}{\partial x^k}(x - a\zeta) - \frac{\partial \varphi}{\partial x^k}(x)\right) d\zeta.$$

The right-hand side tends to zero uniformly in x when $\epsilon \to 0$. As its support lies inside a fixed compact set, we obtain convergence in all L^p $(1 \leq p \leq +\infty)$. □

3.2.2 Corollary. *If $f \in L^p$ and $\varphi \in \mathcal{D}$, then $f * \varphi \in W_s^p$ for every integer $s > 0$.*

PROOF. $\epsilon^{-1}(\tau_{\epsilon a} - I)(f * \varphi) = f * (\tau_{\epsilon a} - I)\epsilon^{-1}\varphi$.
The last term on the right-hand side converges in L^1 by 3.2.1 applied to φ, with $p = 1$, $s = 1$. □

3.2.3 Proposition. *Let $\mu \in M^1$ and assume that μ has compact support. Then $(\tau_\mu\varphi) \in \mathcal{D}$ for every $\varphi \in \mathcal{D}$.*

PROOF. Let K_1 be the support of μ and let K_2 be the support of φ. Then the support of $\tau_\mu\varphi$ lies in the compact set $K_1 + K_2$.

Moreover,

$$(\tau_\mu\varphi)(x) = \int_{K_1} \varphi(x - y)d\mu(y).$$

Differentiating with respect to x^1 under the integral sign is legitimate since $\frac{\partial \varphi}{\partial x^1}$ is continuous and the integral is taken on a compact set. Hence

$$\frac{\partial}{\partial x^1}(\tau_\mu\varphi) = \tau_\mu\left(\frac{\partial \varphi}{\partial x^1}\right). \square$$

3.2.4 Proposition. *The space $\mathcal{D}$ is dense in L^p $(1 \leq p < +\infty)$.*

PROOF. Let $f \in L^p$. Using the truncation operator, we see that there exists $\tilde{f} \in L^p$ such that $\tilde{f}$ is zero outside a compact set and

$$\|f - \tilde{f}\|_{L^p} < \epsilon.$$

Set $\tilde{f} * F_\lambda = u_\lambda$. Then, by 3.2.0(ii),

$$\|u_\lambda - \tilde{f}\|_{L^p} \to 0 \quad \text{as} \quad \lambda \to 0.$$

Since $\tilde{f} \in L^p$ and $\tilde{f}$ has compact support, it follows a fortiori that $\tilde{f} \in L^1$. Hence, by 3.2.3, $\tilde{f} * F_\lambda \in \mathcal{D}$. □

3.3 Weak differentiation

3.3.1 **Definition.** We denote by L^1_{loc} the functions which are integrable on every compact set. Given $f \in L^1_{loc}$, the Radon measure $f(x)dx$ is called the measure associated with f.

$f \in L^1_{loc}$ is said to have a derivative in the direction of the vector a *in the weak sense*, or a weak derivative, if there exists $u_a \in L^1_{loc}$ such that

$$\int f D_a \varphi = -\int u_a \varphi, \quad \forall \varphi \in \mathcal{D}.$$

The reader familiar with the distribution theory of Laurent Schwartz will recognize a special case of differentiation in the sense of distributions.

3.3.2 Theorem. *Let $f \in L^p$. Then the following statements are equivalent:*

(i) $f \in W^p_1$.

(ii) For every $a \in \mathbf{R}^n$, $D_a f$ exists in the weak sense and $D_a f \in L^p$.

Proof. (i) $\Rightarrow$ (ii). The identity $\int (\tau_a f) h = \int f(\tau_{-a} h)$ implies

$$(iii) \qquad \int (\epsilon^{-1}(\tau_{\epsilon a} - 1) f) h = \int f(\tau_{-\epsilon a} h - h)\epsilon^{-1} \quad \forall f \in L^p,\ h \in L^q.$$

Writing (iii) with $h = \varphi$, we can pass to the limit on the left-hand side since $\varphi \in \mathcal{D} \subset L^q$, and on the right-hand side since $\varphi \in W^q_1$ by 3.2.1. This yields the formula for integration by parts:

$$\int D_a f \varphi = -\int f(D_a \varphi) \quad \forall f \in W^p_1,\ \varphi \in \mathcal{D}.$$

Hence $u_a = D_a f$, and (iii) follows since $D_a f \in L^p$.

The proof that (ii) $\Rightarrow$ (i) uses the following version of Taylor's formula with integral remainder.

3.3.3 Lemma. *Let $f \in L^1_{loc}$ and suppose that f has a weak derivative in the direction of a, say u_a. Let ρ_ϵ be the Radon measure defined by*

$$\langle g, \rho_\epsilon \rangle = \int_0^\epsilon g(-\zeta a) d\zeta, \quad \forall g \in C_b(\mathbf{R}^n).$$

Then

$$-\epsilon^{-1}[\tau_{\epsilon a} f - f] = \tau_{\rho_\epsilon} u_a.$$

Proof. Let $\varphi \in \mathcal{D}$. Using formula 3.3.2(iii), Taylor's formula with integral remainder for φ, and Fubini's theorem, we have

$$\begin{aligned} \int \epsilon^{-1}[\tau_{\epsilon a} f - f]\varphi &= \int f(\tau_{-\epsilon a}\varphi - \varphi)\epsilon^{-1} \\ &= \int_{\mathbf{R}^n} f(a) dx \int_0^\epsilon \sum_k a^k \frac{\partial \varphi}{\partial x^k}(x + \xi a) d\xi \\ &= \int_0^\epsilon d\xi \int_{\mathbf{R}^n} \sum a^k \frac{\partial \varphi}{\partial x^k}(x + \xi a) f(x) dx. \end{aligned}$$

Since f is weakly differentiable in the direction of a,

$$\begin{aligned}\sum a^k \frac{\partial\varphi}{\partial x^k}(x+\xi a)f(x)dx &= -\int (\tau_{-\xi a}\varphi)(x)u_a(x)dx \\ &= -\int \varphi(x)(\tau_{\xi a}u_a)(x)dx,\end{aligned}$$

whence

$$\int \{\epsilon^{-1}[\tau_{\epsilon a}f - f] + \tau_{\rho_\epsilon}u_a\}(x)\varphi(x)dx = 0, \quad \forall\varphi \in \mathcal{D}.$$

As we saw in 3.2.4, $\mathcal{D}$ is dense in L^q; this implies that the quantity in braces is the zero function of L^p:

$$(*) \qquad -\epsilon^{-1}[\tau_{\epsilon a}f - f] = \tau_{\rho_\epsilon}u_a. \square$$

3.3.4 Proof that (ii) $\Rightarrow$ (i) in Theorem 3.3.2

The result follows from considering the limit of the right-hand side of $(*)$ and using 1.9.1. □

3.3.5 Corollary. *Let $\{e_1, \ldots, e_n\}$ be a basis for $\mathbf{R}^n$ and let $\{f_n\}$ be a sequence of functions in W_1^p such that $\|f_n - f\|_{L^p} \to 0$ and, for all k, $D_{e_k}f_n$ converges in L^p. Then $f \in W_1^p$ and, for any a in $\mathbf{R}^n$,*

$$\|D_a f - D_a f_n\|_{L^p} \to 0.$$

PROOF. It suffices to prove that f is weakly differentiable in the direction of a. The hypotheses allow us to write

$$\int f_n D_{e_k}\varphi = -\int D_{e_k}f_n\varphi, \quad \forall\varphi \in \mathcal{D}.$$

Since f_n and $D_{e_k}f_n$ converge in L^p and since $D_{e_k}\varphi$ and φ are in L^q, we can pass to the limit in this equation, obtaining

$$\int fD_{e_k}\varphi = -\int \varphi \lim(D_{e_k}f_n).$$

That is, f is weakly differentiable in the direction of e_k and its weak derivative is

$$\lim(D_{e_k}f_n) \in L^p.$$

Let $a \in \mathbf{R}^n$, say $a = \sum a^k e_k$. Then $D_a f_n = \sum a^k D_{e_k}f_n$, and hence f has a weak derivative in the direction of a which is equal to $\sum a^k \lim D_{e_k}f_n$. By Theorem 3.3.2, $f \in W_1^p$ and $D_a f = \sum a^k \lim D_{e_k}f_n = \lim D_a f_n$. □

3.3.6 Corollary. *Let W_1^p be given the norm*

$$\|f\|_{W_1^p} = \|f\|_{L^p} + \sum_k \|D_{e_k}f\|_{L^p},$$

where $\{e_1, \ldots, e_n\}$ is the canonical basis of $\mathbf{R}^n$. Then W_1^p is a complete normed vector space and D_a is a continuous mapping from W_1^p into L^p.

PROOF. The only statement that is not obvious is that W_1^p is complete. If $\{f_n\}$ is a Cauchy sequence in the W_1^p norm, then both $\{f_n\}$ and $\{D_{e_k} f_n\}$ are Cauchy sequences in the L^p norm.

Since L^p is complete, f_n converges to some $f \in L^p$. Moreover, $f \in W_1^p$ by 3.3.5. By definition,

$$\|f_n - f\|_{W_1^p} = \|f_n - f\|_{L^p} + \sum_k \|D_{e_k} f_n - D_{e_k} f\|_{L^p}.$$

Since $\|D_{e_k} f_n - D_{e_k} f\|_{L^p} \to 0$ by 3.3.5, f_n converges to f in W_1^p. □

3.4 Action of $\mathcal{D}$ on W_s^p. The space $W_{s,loc}^p$

3.4.1 Proposition. *Let $\varphi \in \mathcal{D}$ and let the operation of multiplication by φ, written m_φ, be defined by $(m_\varphi f)(x) = \varphi(x) f(x)$. Then*

$$m_\varphi : W_s^p \to W_s^p \text{ for every } p \in [1, +\infty] \text{ and for every integer } s.$$

PROOF. We prove the proposition when $s = 1$. First we show that

$$(i) \qquad D_a(\varphi f) = (D_a \varphi) f + \varphi D_a f.$$

This formula is proved by passing to derivatives in the weak sense. Let $\widetilde{D}_a$ denote the weak derivatives. Then

$$\int \widetilde{D}_a(\varphi f)\psi = -\int \varphi f(D_a(\psi)) \quad \forall \psi \in \mathcal{D}.$$

Furthermore, by Leibnitz's formula for continuously differentiable functions, $-\varphi D_a(\psi) = \psi D_a(\varphi) - D_a(\varphi\psi)$, whence

$$\begin{aligned} \int \widetilde{D}_a(\varphi f)\psi &= -\int f D_a(\varphi\psi) + \int f\psi D_a(\varphi) \\ &= \int \widetilde{D}_a(f)\varphi\psi + \int \psi f D_a(\varphi). \end{aligned}$$

Let

$$G = [\widetilde{D}_a(\varphi f) - \varphi \widetilde{D}_a(f) - f D_a(\varphi)].$$

Then G is orthogonal to every $\psi \in \mathcal{D}$. Since $\mathcal{D}$ is dense in L^q if $q < +\infty$, it follows that G is zero. If $p = 1$, the fact that $G = 0$ follows from the density of $\mathcal{D}$ in $C_0(\mathbf{R}^n)$. Thus (i) is proved for weak derivatives:

$$\widetilde{D}_a(f\varphi) = \varphi \widetilde{D}_a f + f D_a(\varphi).$$

Since φ and $D_a\varphi$ are in L^∞, the right-hand side is in L^p if $f \in W_1^p$. Theorem 3.3.2 then gives the result. □

3.4.2 Differentiable partitions of unity

Theorem. *Let U_α be an open cover of an open subset O of $\mathbf{R}^n$. Then there exists a partition of unity φ_n such that*

$$0 \leq \varphi_n \leq 1,$$
$$\varphi_n \in \mathcal{D}(\mathbf{R}^n),$$
$$supp(\varphi_n) \subset U_{\alpha(n)} \subset O,$$

and

$$\sum \varphi_n(x) = 1, \quad \forall x \in O.$$

The series is locally finite; that is, for every compact subset K of O, $supp\,(\varphi_n) \cap K = \emptyset$ except for a finite number of indices.

PROOF. Let

$$K_n = \left\{ x \in O : \operatorname{dist}(x, O^c) \geq \frac{1}{n} \text{ and } \|x\| \leq n \right\}.$$

Then each K_n is a compact set contained in O, and the union of all the K_n equals O. By Theorem II-1.4.1 we can find a partition of unity with continuous functions f_n. We may also assume that U_α is a locally finite cover. Set

$$2\epsilon_n = \operatorname{dist}(\operatorname{supp}(f_n), U^c_{\alpha(n)}).$$

Let $\psi_n = F_{\epsilon_n} * f_n$, where F_λ was defined in 3.2.0(ii). Then

$$\operatorname{supp}(\psi_n) \subset \operatorname{supp}(f_n) + B(0, \epsilon_n) \subset U_{\alpha(n)}.$$

By 3.2.3, $\psi_n \in \mathcal{D}$ since $F_{\epsilon_n} \in \mathcal{D}$.
Next, writing out the integral expression for ψ_n,

$$\int f_n(x-y) F_{\epsilon_n}(y) dy = \psi_n(x),$$

we see that $\psi_n(x) > 0$ whenever $f_n(x) > 0$. Hence

$$\sum \psi_n(x) > 0 \text{ for every } x \in O.$$

Set

$$r(x) = \sum \psi_n(x).$$

Then r^{-1} is an infinitely differentiable function and $\varphi_n = r^{-1}\psi_n$ satisfies the conditions of the theorem. □

3.4.3 The spaces $W^p_{s,loc}$

(i) Let O be an open set in $\mathbf{R}^n$. We denote by $\mathcal{D}(O)$ the infinitely differentiable functions defined in O which have compact support. A function φ in $\mathcal{D}(O)$ can be extended to $\mathbf{R}^n$ by setting $\varphi(x) = 0$ for $x \notin O$.

Writing $\widetilde{\varphi}$ for the extension of φ to $\mathbf{R}^n$, we note that $\widetilde{\varphi}$ is infinitely differentiable: given a point x_0 on the boundary of O, there exists an open neighborhood V of x_0 in $\mathbf{R}^n$ which does not meet the support of φ. Hence $\widetilde{\varphi}$ vanishes identically in V and is therefore infinitely differentiable. Thus

$$\text{(ii)} \qquad \mathcal{D}(O) \simeq \{\varphi \in \mathcal{D}(\mathbf{R}^n) : \operatorname{supp}(\varphi) \subset O\}.$$

We define

$$\begin{aligned} W^p_{s,loc}(O) &= \{f \text{ defined and measurable on } O: \\ &\qquad f\varphi \in W^p_s(\mathbf{R}^n) \text{ for any } \varphi \in \mathcal{D}(O)\}. \end{aligned}$$

(iii) Proposition. *$f \in W^p_{s,loc}(O)$ if and only if for every $x_0 \in O$ there exists an open neighborhood V_{x_0} of x_0 in O such that*

$$\varphi f \in W^p_s(\mathbf{R}^n) \quad \forall \varphi \in \mathcal{D}(\mathbf{R}^n) \text{ with } supp(\varphi) \subset V_{x_0}.$$

PROOF. The forward implication is trivial. The reverse implication is proved by using a partition of unity subordinate to the cover $\{V_{x_0}\}$, where $x_0 \in O$. □

3.5 Sobolev spaces

We now study the spaces W^2_s. Since W^2_s is a subspace of L^2 for every s, Plancherel's theorem allows us to characterize its image under the Fourier transform. The space W^2_s is written H^s and called the Sobolev space of order s. The isomorphism of $L^2(\mathbf{R}^n)$ onto $L^2(\mathbf{R}^n)$ defined by Plancherel's extension of the Fourier transform in 2.5 is denoted by $\mathcal{F}$.

3.5.1 Theorem. *Let $f \in L^2(\mathbf{R}^n)$ and let $h = \mathcal{F}(f)$ be its Fourier-Plancherel transform. Then the following two statements are equivalent:*

$$(i) \qquad f \in H^s.$$

$$(ii) \qquad \int_{\mathbf{R}^n} |h(t)|^2(1 + \|t\|^2)^s dt < +\infty.$$

PROOF. Restricting to the case where $s = 1$, we first show that (i) $\Rightarrow$ (ii). For $f \in H^1$, we have the following extension of Theorem 3.1.1:

$$(iii) \qquad \mathcal{F}(D_{e_k} f) = -it_k\mathcal{F}(f)(t).$$

To prove this, note that $\mathcal{F}(\tau_{\epsilon e_k} f) = \mathrm{e}^{i\epsilon t_k}\mathcal{F}(f)$ and

$$\epsilon^{-1}\mathcal{F}(\tau_{\epsilon e_k} f - f) = \epsilon^{-1}(\mathrm{e}^{i\epsilon t_k} - 1)h(t).$$

Since the left-hand side converges in L^2 to $-\mathcal{F}(D_{e_k}f)$, the right-hand side also converges in L^2. Passing to a subsequence ϵ_k, convergence in L^2 implies convergence a.e.; (iii) follows since the right-hand side converges everywhere to $(it_k)h(t)$. Hence

$$f \in H^s \Rightarrow h \in L^2 \text{ and } (t_k h(t)) \in L^2,$$

and therefore

$$|h(t)|^2(1 + t_1^2 + \ldots + t_n^2) \in L^1.$$

We now prove that (ii) $\Rightarrow$ (i). Let $\varphi \in \mathcal{D}$; then, by Plancherel,

$$\int f\overline{D_{e_k}\varphi} = \frac{1}{(2\pi)^n}\int \mathcal{F}(f)\overline{\mathcal{F}(D_{e_k}\varphi)}.$$

By (iii) (or 3.1.1), $\mathcal{F}(D_{e_k}\varphi)(t) = -it_k\mathcal{F}(\varphi)(t)$, whence

$$\int f\overline{D_{e_k}\varphi} = \frac{1}{(2\pi)^n}\int (it_k)\mathcal{F}(f)(t)\overline{\mathcal{F}(\varphi)(t)}dt.$$

By (ii), $t_k\mathcal{F}(f)(t) \in L^2$. The inverse Plancherel isomorphism $\mathcal{F}^{-1}$ can now be used to show that there exists a function $u_k \in L^2$ such that $\mathcal{F}(u_k)(t) = -it_k(\mathcal{F}f)(t)$. Thus

$$\int fD_{e_k}\varphi = \int u_k\varphi;$$

that is, the weak derivative of f in the direction e_k is the function $u_k \in L^2$. Theorem 3.3.2 shows that $f \in W_1^2 = H^1$. $\square$

3.5.2 Definition of H^s for s not an integer

Let s be a positive real number that is not an integer. Set

$$H^s = \left\{f \in L^2 : \int_{\mathbf{R}^n}(1 + \|t\|^2)^s|(\mathcal{F}f)(t)|^2dt < +\infty\right\}.$$

We define a norm on H^s by

$$(i) \qquad \|f\|_{H^s}^2 = \int_{\mathbf{R}^n}(1 + \|t\|^2)^s|(\mathcal{F}f)(t)|^2dt.$$

For $s = 1$, this norm is different from the W_1^2 norm introduced earlier, but the two are equivalent. The advantage of the present norm is that H^s becomes a *Hilbert space* with scalar product

$$(f_1|f_2)_{H^s} = \int_{\mathbf{R}^n}(h_1\overline{h}_2)(t)(1 + \|t\|^2)^s dt, \quad \text{where} \quad h_k = \mathcal{F}(f_k),\ k = 1, 2.$$

3.5.3 Proposition. *Let $f \in H^s$. Then*

$$(i) \qquad \tau_\mu f \in H^s \textit{ for every measure } \mu \in M^1.$$

PROOF. $\mathcal{F}(\tau_\mu f)(t) = \widehat{\mu}(t)\mathcal{F}(f)(t)$. Hence, since $|\widehat{\mu}(t)| \leq \|\mu\|_{M^1}$, 3.5.2(i) implies that

$$\|\tau_\mu f\|_{H^s} \leq \|\mu\|_{M^1}\|f\|_{H^s} < +\infty. \square$$

3.5.4 Differential characterization of H^s

Proposition. *Let $f \in L^2(\mathbf{R}^n)$ and let $0 < s < 1$. Then the following two statements are equivalent:*

(i) $f \in H^s(\mathbf{R}^n)$.
(ii) $I_s(f) = \int_{\mathbf{R}^n} \|\tau_x f - f\|^2_{L^2} \frac{1}{\|x\|^{n+2s}} dx < +\infty$, *where* $n = dim(E)$.

PROOF. We use the Fourier-Plancherel isomorphism. Let $u = \mathcal{F}(f)$. Then

$$I_s(u) = \int_{\mathbf{R}^n} \frac{dx}{\|x\|^{n+2s}} \int_{\mathbf{R}^n} |e^{-ix.\xi} - 1|^2 |\widehat{u}(\xi)|^2 d\xi.$$

Next, we set

$$\lambda(\xi) = \int_{\mathbf{R}^n} |e^{-ix.\xi} - 1|^2 \frac{dx}{\|x\|^{n+2s}}.$$

This integral is invariant under the mapping $x \to A.x$, where A is an orthogonal matrix. Hence $\lambda({}^tA.\xi) = \lambda(\xi)$; that is, there exists a function $\psi : \mathbf{R}^+ \to \mathbf{R}^+$ such that $\lambda(\xi) = \psi(\|\xi\|)$.

Note that, under the dilation $\xi \mapsto \alpha\xi$ $(\alpha > 0)$,

$$\psi(\alpha\|\xi\|) = \int_{\mathbf{R}^n} |e^{-ix.\alpha\xi} - 1|^2 \frac{dx}{\|x\|^{n+2s}}.$$

Setting $\alpha x = y$ gives

$$\psi(\alpha\|\xi\|) = \int_{\mathbf{R}^n} |e^{-iy.\xi} - 1|^2 \frac{\alpha^{-n} dy}{\|y\|^{n+2s}} \alpha^{n+2s} = \psi(\alpha\|\xi\|) = \alpha^{2s}\psi(\|\xi\|).$$

Setting $\|\xi\| = 1$, this shows that $\psi(\alpha) = \alpha^{2s}\psi(1)$. Hence $\lambda(\xi) = c\|\xi\|^{2s}$, where c is a strictly positive constant. Finally,

$$I_s(u) = c \int_{\mathbf{R}^n} |u(\xi)|^2 \|\xi\|^{2s} d\xi.$$

Since f is assumed to be an L^2 function, $\int |u(\xi)|^2 d\xi < +\infty$. Hence the finiteness of $I_s(u)$ is equivalent to that of

$$\int_{\mathbf{R}^n} |u(\xi)|^2 (1 + \|\xi\|^2)^s d\xi. \square$$

Corollary. *Let $f \in L^2(E)$, where s is a positive real number. Let s be decomposed as $s = p + s'$, with $0 \leq s' < 1$ and p an integer. Then the following statements (iii) and (iv) are equivalent:*

(iii) $f \in H^s(\mathbf{R}^n)$.
(iv) $(D^{m_1}_{e_1} \dots D^{m_n}_{e_n} f) \in H^{s'}$, $\quad \forall m$ *such that* $|m| \leq p$.

REMARK. If $s' = 0$, then $H^{s'} = L^2$ and this is the definition of H^s for integer s given in 3.1.

If $s' > 0$, then $0 < s' < 1$ and membership in $H^{s'}$ is characterized by convergence of the integral (ii).

PROOF. Set $\mathcal{F}f = u$; then (iii) becomes

$$[\xi_1^{m_1}\xi_2^{m_2}\dots\xi_n^{m_n}(1+\|\xi\|)^{s'}]u \in L^2, \quad \forall m \text{ such that } |m| \le p.$$

This is equivalent to

$$(1+\|\xi\|)^s u \in L^2.\square$$

3.5.5 Operator of multiplication by a differentiable function

Proposition. *Let $\varphi \in \mathcal{D}(\mathbf{R}^n)$ and let $f \in H^s$. Then $\varphi f \in H^s$.*

PROOF. The result was proved for integer n in 3.4.1. Using 3.5.4(iii) reduces the proof to the case where $0 < s' < 1$.

We begin by writing

$$(i) \qquad \tau_x(\varphi f) - \varphi f = \varphi(\tau_x f - f) + \tau_x(\varphi f) - \varphi\tau_x(f).$$

Then, since φ is bounded,

$$(ii) \qquad \|\varphi(\tau_x f - f)\|_{L^2} \le \|\varphi\|_{L^\infty}\|\tau_x f - f\|_{L^2}.$$

Set $x - x_0 = y$; then $\int_E |\tau_{x_0}(\varphi f) - \varphi\tau_{x_0} f|^2 dx = \int |\varphi f - (\tau_{-x_0}\varphi)f|^2 dy$. Thus

$$\|\tau_x(\varphi f) - \varphi\tau_{x_0}(f)\|_{L^2}^2 = \|(\tau_{-x_0}\varphi - \varphi)f\|_{L^2}^2 \le \|f\|_{L^2}^2\|\tau_{-x_0}\varphi - \varphi\|_{L^\infty}^2.$$

By the mean value theorem,

$$(iii) \qquad \|\tau_{-x_0}\varphi - \varphi\|_{L^\infty}^2 \le C\|x_0\|^2.$$

Substituting inequalities (ii) and (iii) into (i), we obtain the integral convergence criterion 3.5.4(ii). $\square$

3.5.6 The spaces $H^s_{loc}(O)$

Let O be an open set in $\mathbf{R}^n$. We say that $f \in L^2_{loc}(O)$ if $f\mathbf{1}_K \in L^2(\mathbf{R}^n)$ for every compact subset K of $\mathbf{R}^n$. For $s > 0$, we say that $f \in H^s_{loc}(O)$ if $\varphi f \in H^s(\mathbf{R}^n)$ $\forall\varphi \in \mathcal{D}(O)$. The next proposition follows essentially from 3.5.5.

Proposition. *Let $f \in L^2_{loc}(O)$ and suppose that, for every $x_0 \in O$, there exists a function $\varphi \in \mathcal{D}(O)$ such that $\varphi(x_0) \ne 0$ and $\varphi f \in H^s(\mathbf{R}^n)$. Then $f \in H^s_{loc}(O)$.*

PROOF. Let $v \in \mathcal{D}(O)$ be such that $v \equiv 1$ on a neighborhood of x_0; assume that its support supp (v) is small enough that $\varphi(x) \ne 0$ on supp (v). Multiplying by $\frac{1}{\varphi(x)}v(x)$, we obtain

$\forall x_0 \in O\ \exists\theta \in \mathcal{D}(O)$ such that $\theta\varphi \in H^s(\mathbf{R}^n)$ and $\theta \equiv 1$ on a neighborhood of x_0.

Let U_{x_0} = interior of $\theta^{-1}(1)$. Then, for $x_0 \in O$, the collection $\{U_{x_0}\}$ is an open cover of O. Let $\chi_1, \ldots, \chi_n \in \mathcal{D}(O)$ be a partition of unity subordinate to this cover. Then $\chi_n f = \chi_n \theta f$, where θ corresponds to the open set U containing $\operatorname{supp}(\chi_n)$. By 3.5.5, $\chi_n(\theta f) \in H^s$; that is, $\chi_n f \in H^s$ for every s.

Let $\varphi \in \mathcal{D}(O)$. Then the identity $f = \sum \chi_n f$ gives $\varphi f = \sum \varphi \chi_n f$. This sum is *finite* and all the terms are in H^s; hence $\varphi f \in H^s$. □

3.5.7 Invariance under diffeomorphism

Theorem. *Let O be an open set in $\mathbf{R}^n$ and let g be an infinitely differentiable diffeomorphism from O onto an open set $\widetilde{O}$. If $\widetilde{f} \in H^s_{loc}(\widetilde{O})$, then $(\widetilde{f} \circ g) \in H^s_{loc}(O)$.*

PROOF. We use the criterion in 3.5.6. If s is an integer, it suffices to compute the derivatives of the composite function $\widetilde{\psi} \circ g$ (where $\widetilde{\psi} \in \mathcal{D}(\widetilde{O})$) and to use the characterization of H^s by means of weak derivatives.

By using 3.5.4, we may assume that $0 < s < 1$ and that f and $\widetilde{f}$ have compact support. Then the integral 3.5.4(ii) becomes

$$I_s(f) = \int_{\mathbf{R}^n \times \mathbf{R}^n} |\widetilde{f}(g(x+y)) - \widetilde{f}(g(y))|^2 \frac{dx}{\|x\|^{n+2s}} dy.$$

Consider the mapping of x defined by

$$p_y(x) = g(x+y) - g(y).$$

Then p is a diffeomorphism for fixed y. Let

$$I_s(f) = \int_{\mathbf{R}^n} dy \int_{\mathbf{R}^n} |\widetilde{f}(g(y) + p_y(x)) - \widetilde{f}(g(y))|^2 \frac{dx}{\|x\|^{n+2s}}.$$

Setting $g(y) = \widetilde{y}$ and $p_{\widetilde{y}}(x) = z$ gives $dy = (\det g^{-1})(\widetilde{y}) d\widetilde{y}$ and $dx = \det(p_y^{-1}) dz$. By the change-of-variables formula for multiple integrals,

$$I_s(f) = \int_{\mathbf{R}^n} d\widetilde{y} (\det g^{-1}) \int |\widetilde{f}(\widetilde{y} + z) - \widetilde{f}(\widetilde{y})|^2 (\det p_y^{-1}) \frac{dz}{\|x\|^{n+2s}} < +\infty.$$

Since g is a diffeomorphism and all that matters is its restriction to a compact set, there exists a constant c_1 such that $\|x\| \geq c_1 \|z\|$. Similarly, there exists an *upper bound* c_2 for the functional determinants, and

$$I_s(f) \leq c \int_{\mathbf{R}^n} d\widetilde{y} \int |\widetilde{f}(\widetilde{y} + z) - \widetilde{f}(\widetilde{y})|^2 \frac{dz}{\|z\|^{n+2s}} < +\infty.$$

The integral is finite because $\widetilde{f} \in H^s(\widetilde{O})$. □

3.5.8 Trace theorem. *Let $f \in H^s(\mathbf{R}^n)$, and consider $\mathbf{R}^{n-p} \subset \mathbf{R}^n$. Then the restriction operator*

$$\rho_p : \mathcal{D}(\mathbf{R}^n) \to \mathcal{D}(\mathbf{R}^{n-p})$$

has a continuous operator extension

$$H^s \to H^{s-p/2} \quad \text{if} \quad s > \frac{p}{2}.$$

PROOF. Let $\varphi \in \mathcal{D}(\mathbf{R}^n)$ and write $x = y + z$, where $y \in \mathbf{R}^{n-p}$ and $z \in \mathbf{R}^p$. Then, by the inversion formula,

$$\varphi(y+z) = \int_{\mathbf{R}^n} e^{-i(y+z).\xi} \widehat{\varphi}(\xi) d\xi.$$

Similarly, writing $\xi = \eta + \zeta$,

$$\varphi(y) = \int_{\mathbf{R}^p \times \mathbf{R}^{n-p}} e^{-iy.\eta} \widehat{\varphi}(\eta + \zeta) d\eta d\zeta.$$

Letting $\rho(\varphi)$ denote the operator of restriction to $\mathbf{R}^{n-p}$, we obtain

$$(i) \qquad (\rho(\varphi))^\wedge(\eta) = \int_{\mathbf{R}^p} \widehat{\varphi}(\eta + \zeta) d\zeta.$$

Moreover,

$$\begin{aligned} \|\rho(\varphi)\|^2_{H^{s-\frac{p}{2}}} &= \int |(\rho(\varphi))^\wedge(\eta)|^2 (1 + \|\eta\|^2)^{s-p/2} d\eta \\ &= \int_{\mathbf{R}^{n-p}} \left| \int_{\mathbf{R}^p} \widehat{\varphi}(\eta + \zeta) d\zeta \right|^2 (1 + \|\eta\|^2)^{s-p/2} d\eta. \end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} &\left| \int_{\mathbf{R}^p} \widehat{\varphi}(\eta + \zeta)(1 + \|\eta + \zeta\|^2)^{s/2} (1 + \|\eta + \zeta\|^2)^{-s/2} d\zeta \right|^2 \\ &\le \left(\int_{\mathbf{R}^p} \frac{d\zeta}{(1 + \|\eta + \zeta\|^2)^s} \right) \left(\int_{\mathbf{R}^p} |\widehat{\varphi}(\eta + \zeta)|^2 (1 + \|\eta + \zeta\|^2)^s d\zeta \right). \end{aligned}$$

The first integral on the right-hand side, say $J(\eta)$, converges since $s > p/2$. Moreover,

$$(1 + \|\eta + \zeta\|^2)^s = (1 + \|\eta\|^2 + \|\zeta\|^2)^s$$

and

$$J(\eta) = \int_{\mathbf{R}^p} \frac{d\zeta}{(1 + \|\eta\|^2 + \|\zeta\|^2)^s} = \int_{\|\zeta\| < \|\eta\|} + \int_{\|\zeta\| > \|\eta\|}.$$

The first integral is bounded above by

$$\frac{\text{vol}(\{\|\zeta\| \le \|\eta\|\})}{(1 + \|\eta\|^2)^s} \le C(1 + \|\eta\|^2)^{p/2-s}$$

and the second by

$$\int_{\|\zeta\| > \|\eta\|} \frac{d\zeta}{(1 + \|\zeta\|^2)^s} < (1 + \|\eta\|^2)^{p/2-s},$$

whence $J(\eta) \leq c(1+\|\eta\|^2)^{p/2-s}$ and

$$\begin{aligned}\|\rho(\varphi)\|_{H^{s-p/2}} &\leq c\int_{\mathbf{R}^{n-p}} d\eta(1+\|\eta\|^2)^{p/2-s}(1+\|\eta\|^2)^{s-p/2}\\ &\qquad\times\left[\int_{\mathbf{R}^p}|\widehat{\varphi}(\eta+\zeta)|^2(1+\|\eta+\zeta\|^2)^s d\zeta\right]\\ &\leq c\int_{\mathbf{R}^{n-p}} d\eta\left[\int_{\mathbf{R}^p}|\widehat{\varphi}(\eta+\zeta)|^2(1+\|\eta+\zeta\|^2)^s d\zeta\right] = c\|\varphi\|^2_{H^s}.\end{aligned}$$

Thus

$$(ii) \qquad \|\rho_p(\varphi)\|_{H^{s-p/2}} \leq c\|\varphi\|^2_{H^s} \quad \text{if} \quad s > \frac{p}{2}.$$

The existence of the desired extension follows from the density of $\mathcal{D}(\mathbf{R}^n)$ in $H^s(\mathbf{R}^n)$. □

3.5.9 Corollary (Serge Bernstein). *Let $s > n/2$. Then $H^s(\mathbf{R}^n) \subset C_b(\mathbf{R}^n)$, where $C_b(\mathbf{R}^n)$ denotes the bounded continuous functions.*

PROOF. The inequality 3.5.8(ii), with $p = n$, gives

$$|\rho_n(\varphi)(0)| \leq c\|\varphi\|_{H^s}, \quad \forall\varphi \in \mathcal{D}(\mathbf{R}^n).$$

Since the H^s norm is translation invariant, $|\rho_n(\varphi)(x)| \leq c\|\varphi\|_{H^s}$ for every $x \in \mathbf{R}^n$, whence, taking the sup over x,

$$(i) \qquad \|\rho_n(\varphi)\|_{C_b(\mathbf{R}^n)} \leq c\|\varphi\|_{H^s}.$$

Let $f \in H^s(\mathbf{R}^n)$. There exists a sequence $\varphi_q \in \mathcal{D}(\mathbf{R}^n)$ such that $\|f-\varphi_q\|_{H^s} \to 0$. Then

$$\rho_n(\varphi_q) = u_q \in C_b(\mathbf{R}^n).$$

The u_q converge uniformly by (i); hence

$$\lim u_q(x) = u \in C_b(\mathbf{R}^n).\square$$

3.5.10 Theorem. *Let O be an open set in $\mathbf{R}^n$ and let V be an $(n-p)$-dimensional submanifold of $\mathbf{R}^n$ such that $V \subset O$. If $s > \frac{p}{2}$, then there exists a continuous restriction operator*

$$H^s_{loc}(O) \to H^{s'}_{loc}(V), \quad \textit{where} \quad s' = s - \frac{p}{2}.$$

PROOF. $H^s_{loc}(V)$ is defined via an atlas of charts on V. This definition is independent of the choice of atlas, since passage from one atlas to another is accomplished by local diffeomorphisms.[1] The result follows from Theorem 3.5.7.

Given $v_0 \in V$, there is a local diffeomorphism from a neighborhood U of v_0 to O such that the image of $V \cap U$ is the space $\mathbf{R}^{n-p} \subset \mathbf{R}^n$, and 3.5.9 can be applied. □

---

[1]See, for example, W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry* (New York: McGraw-Hill, 1987).

4 Fourier Transform of Tempered Distributions

Plancherel's theorem, characterizing the image of L^2 under the Fourier transform, played a major role in the last section. Although we hardly considered the spaces W_s^p (s integer, $p \neq 2$), the systematic use of the Plancherel isomorphism enabled us to study the spaces $H^s = W_s^2$. In Section 5, we will study pseudo-differential operators by restricting our attention to their action on the classes H^s, where we will again use the Plancherel isomorphism.

In this section, we characterize the image under the Fourier transform of the space $\mathcal{S}(\mathbf{R}^n)$ of infinitely differentiable functions which, together with all their derivatives, are of rapid decrease. The Fourier transform is an isomorphism from $\mathcal{S}(\mathbf{R}^n)$ onto itself, and $\mathcal{S}(\mathbf{R}^n)$ will be given a topology in which this isomorphism is continuous. The dual of $\mathcal{S}(\mathbf{R}^n)$ is the space of tempered distributions $\mathcal{S}'(\mathbf{R}^n)$ of Laurent Schwartz; the Fourier transform induces, by transposition, an isomorphism from $\mathcal{S}'(\mathbf{R}^n)$ onto itself.

Our study of the Sobolev spaces of negative order will parallel that of $\mathcal{S}(\mathbf{R}^n)$ and $\mathcal{S}'(\mathbf{R}^n)$.

4.1 The space $\mathcal{S}(\mathbf{R}^n)$

(i) Functions of rapid decrease

Definition. A continuous function f on $\mathbf{R}^n$ is said to be of rapid decrease if, for any integer m,

$$(1 + \|x\|^2)^m f(x) \to 0 \quad \text{as} \quad \|x\| \to \infty.$$

The space of functions of rapid decrease is denoted by $C_{0,0}(\mathbf{R}^n)$ and equipped with the following sequence of norms:

$$\|f\|_{m,0} = \max_{x \in \mathbf{R}^n} (1 + \|x\|^2)^n |f(x)|.$$

$C_{0,0}(\mathbf{R}^n)$ is thus a *vector subspace* of $C_0(\mathbf{R}^n)$, the space of continuous functions which vanish at infinity. Moreover,

$$C_{0,0}(\mathbf{R}^n) = \{f \in C_0(\mathbf{R}^n) : \|f\|_{m,0} < +\infty \ \forall m\}.$$

We define

(ii) $\quad \mathcal{S}(\mathbf{R}^n) = \{f \in C_{0,0}(\mathbf{R}^n) : \partial^q f \in C_{0,0}(\mathbf{R}^n), \quad \forall q = (q_1, \ldots, q_n),$

where

$$\partial^q f = \frac{\partial^{|q|} f(x)}{(\partial x^1)^{q_1} \ldots (\partial x^n)^{q_n}},$$

and this derivative is assumed to exist in the elementary sense. In other words, f is infinitely differentiable and all its derivatives are of rapid decrease.

(iii) Norms on $\mathcal{S}(\mathbf{R}^n)$

A countable family of norms is defined on $\mathcal{S}$ by

$$\|f\|_{m,r} = \sup_{|q|\le r} \|\partial^q f\|_{m,0}.$$

These norms can be used to give $\mathcal{S}(\mathbf{R}^n)$ a metrizable topology, with distance defined by

$$\begin{aligned} d(f,0) &= \sum_{r,m} 2^{-(r+m)} \frac{\|f\|_{m,r}}{1+\|f\|_{m,r}}, \\ d(f,f') &= d(f-f',0). \end{aligned}$$

(iv) $\mathcal{D}(\mathbf{R}^n)$ is a dense subset of $\mathcal{S}(\mathbf{R}^n)$

Let φ be an element of $\mathcal{D}(\mathbf{R}^n)$ such that $\varphi = 1$ on a neighborhood of zero. Set $\varphi_n(x) = \varphi(\frac{x}{n})$. If $f \in \mathcal{S}(\mathbf{R}^n)$, then $d(f, f\varphi_n) \to 0$ and $f\varphi_n \in \mathcal{D}$.

A linear functional l on $\mathcal{S}(\mathbf{R}^n)$ is continuous if and only if there exist m, r, and a constant c such that

$$|l(f)| \le c\|f\|_{m,r}.$$

4.2 Isomorphism of $\mathcal{S}(\mathbf{R}^n)$ under the Fourier transform

Theorem (Laurent Schwartz). *Let $f \in \mathcal{S}(\mathbf{R}^n)$. Then*

(i) $\widehat{f} \in L^1$ and the Fourier inversion theorem can be applied:

$$f(x) = \int \widehat{f}(t) e^{-ix.t} \frac{dt}{(2\pi)^n}.$$

(ii) $\widehat{f} \in \mathcal{S}(\mathbf{R}^n)$ and there exist constants $c_{r,s}$ such that

$$\|\widehat{f}\|_{r,s} \le c_{r,s}\|f\|_{r,m+s}, \quad \textit{where} \quad m > n.$$

(iii) The mapping $f \to \widehat{f}$ defines a topological isomorphism of $\mathcal{S}(\mathbf{R}^n)$ onto $\mathcal{S}(\mathbf{R}^n)$.
(iv) $(x^k f)^\wedge(x) = i\frac{\partial}{\partial t_k}\widehat{f}(t)$
(v) $\left(\frac{\partial}{\partial x^k} f\right)^\wedge(t) = -it_k\widehat{f}(t)$
(vi) If $f,\ g \in \mathcal{S}(\mathbf{R}^n)$, then $fg \in \mathcal{S}(\mathbf{R}^n)$ and $(fg)^\wedge = \widehat{f} * \widehat{g}$.
(vii) If $f,\ g \in \mathcal{S}(\mathbf{R}^n)$, then $f * g \in \mathcal{S}(\mathbf{R}^n)$ and $(f*g)^\wedge = \widehat{f}\widehat{g}$.

REMARKS. From now on, whenever there is no possibility of confusion, $\mathcal{S}(\mathbf{R}^n)$ will be abbreviated by $\mathcal{S}$. The Fourier transform on $\mathcal{S}$ has all the right properties: it maps differentiations to multiplications (by $-i$ times the variable of differentiation) and vice versa, and convolutions to products and vice versa.

PROOF. If $f \in \mathcal{S}$,

$$\int f(x)\mathrm{e}^{ix.t}dx = \int \left(\frac{\partial}{\partial x^1} f\right)\left(\frac{i}{t^1}\mathrm{e}^{ix.t}\right) dx.$$

This identity can be checked by an integration by parts on the right-hand side; the variation of the integrated term vanishes because f is of rapid decrease. It follows that $\left(\dfrac{\partial f}{\partial x^1}\right)^{\wedge}(t) = -it_1\widehat{f}(t)$, and (iv) is proved. Moreover,

$$|t_1|\left|\widehat{f}(t)\right| \leq \left\|\frac{\partial f}{\partial x^1}\right\|_{L^1} \leq \left\|\frac{\partial f}{\partial x^1}\right\|_{m,0} \quad \text{since} \quad m > n.$$

(The last inequality uses the fact that $(1 + \|x\|^2)^{-m/2} \in L^1(\mathbf{R}^n)$.)

In general, it follows from repeated integrations by parts that $(\partial^q)^\wedge f(t) = (-i)^{|q|}t^q\widehat{f}(t)$, whence $|t^q\widehat{f}(t)| \leq c\|\partial^q f\|_{m,0}$, and finally

$$(viii) \qquad \|\widehat{f}\|_{r,0} \leq c_m\|f\|_{m,r}.$$

Hence $f \in \mathcal{S}$ implies $\widehat{f} \in C_{0,0} \subset L^1$.

Thus the Fourier inversion formula can be applied, and (i) is proved.

Let ∂_t^q be a derivative of order q in t. It can be computed by differentiating the Fourier integral under the integral sign:

$$\partial_t^q\widehat{f}(t) = \int (\partial_t^q\mathrm{e}^{ix.t})f(x)dx = \int \mathrm{e}^{ix.t}(-i)^{|q|}x^q f(x)dx.$$

Since $x^q f(x) \in \mathcal{S}$, it follows from (viii) that

$$\|\partial_t^q\widehat{f}\| \leq c_m\|x^q f(x)\|_{m,r}.$$

Writing out in detail the norm on the right-hand side gives

$$\|x^q f(x)\|_{m,r} = \sum_{|l|\leq r} \|\partial^l(x^q f(x))\|_{m,0}.$$

By Leibnitz's formula for the derivative of a product,

$$\partial^l(x^q f) = \sum C_r^l(\partial^r x^q)(\partial^{1-r} f)(x).$$

It follows that $\|x^q f(x)\|_{m,r} \leq c_{m,q}\|f\|_{m+q,r}$, whence

$$\|\widehat{f}\|_{r,s} \leq c_{r,s}\|f\|_{m+s,r}.$$

This proves (ii).

To prove (iii), we must show that the mapping $f \to \widehat{f}$ is *surjective*. Let $h \in \mathcal{S}$ be given, and set $h_1(x) = \int h(t) e^{-it.x} \frac{dt}{(2\pi)^n}$. Then

$$(ix) \qquad h_1(x) = \frac{1}{(2\pi)^n} \widehat{h}(-x),$$

and $h_1 \in \mathcal{S}$ by (ii). We now compute its Fourier transform.

$$\widehat{h}_1(\lambda) = \int e^{ix.\lambda} h_1(x) dx = \frac{1}{(2\pi)^n} \int \widehat{h}(-x) e^{ix.\lambda} dx = \frac{1}{(2\pi)^n} \int \widehat{h}(x) e^{-ix.\lambda} dx.$$

By (i), $\widehat{h}_1 = h$. This shows that the Fourier transform is surjective. The inverse transform, given by (ix), is continuous by (ii). Both the isomorphism $\mathcal{S} \to \mathcal{S}$ and its inverse are continuous: it is thus a topological isomorphism.

Applying the Fourier isomorphism to formula (iv), which has already been proved, gives (v).

Since $f, g \in \mathcal{S} \subset L^1$, 1.6.2 can be applied and $(f * g)^\wedge = \widehat{f}\widehat{g}$. It is clear that the product of two functions in $\mathcal{S}$ is in $\mathcal{S}$: if $\widehat{f} \in \mathcal{S}$ and $\widehat{g} \in \mathcal{S}$, then $\widehat{f}\widehat{g} \in \mathcal{S}$. It follows that $f * g \in \mathcal{S}$. This proves the first part of (vii), and the second part follows from (vi) by the Fourier isomorphism. □

4.3 The Fourier transform in spaces of distributions

4.3.1 Notation

Using the notation of Laurent Schwartz, we write $\mathcal{S}'$ for the vector space of continuous linear functionals on $\mathcal{S}$. $\mathcal{S}'$ is called the vector space of *tempered distributions* on $\mathbf{R}^n$. For example, let $\mu \in M(\mathbf{R}^n)$ be such that there exist l and C for which

$$(i) \qquad |\mu|(\{x : \|x\| < R\}) \leq C(\|x\|^2 + 1)^l.$$

Then $\int f(x) d\mu(x)$ converges $\forall f \in \mathcal{S}$ and defines a distribution in $\mathcal{S}'$.

4.3.2 Operations on $\mathcal{S}'$

These are derived by transposition from continuous linear operations on $\mathcal{S}$.

(i) Differentiation is a continuous linear operation on $\mathcal{S}$. Since

$$\left\| \frac{\partial f}{\partial x^1} \right\|_{m,r} \leq \|f\|_{m,r+1},$$

differentiation on $\mathcal{S}'$ can be defined by

$$-\left\langle \frac{\partial}{\partial x^1} f, l \right\rangle = \left\langle f, \frac{\partial l}{\partial x^1} \right\rangle, \quad \forall l \in \mathcal{S}'.$$

The left-hand side clearly defines a continuous form on $\mathcal{S}'$.

(ii) Multiplication by a polynomial P of degree k is a continuous operation on $\mathcal{S}$. Since

$$\|P(x)f(x)\|_{m,r} \leq c\|f\|_{m+k,r}, \quad \text{where } c = c(P),$$

multiplication by a polynomial on $\mathcal{S}'$ can be defined by

$$\langle Pf, l\rangle = \langle f, Pl\rangle.$$

(iii) $\mathcal{S}$ is an algebra: the product of two functions in $\mathcal{S}$ is a function in $\mathcal{S}$. That $\mathcal{S}'$ is an $\mathcal{S}$-module follows from the formula

$$\langle hf, l\rangle = \langle f, hl\rangle, \ \forall f \in \mathcal{S},$$

where l and h are fixed elements of $\mathcal{S}'$ and $\mathcal{S}$, respectively.

4.3.3 The weak topology on $\mathcal{S}'$

Definition. A sequence $l_n \in \mathcal{S}'$ is said to converge weakly to l_0 if

$$\langle f, l_n\rangle \text{ converges to } \langle f, l_0\rangle, \quad \forall f \in \mathcal{S}.$$

Proposition. *The operations defined in 4.3.2 are continuous in the weak topology on $\mathcal{S}'$.*

In particular, if $l_n \to l_0$ weakly, then

$$\frac{\partial}{\partial x^1} l_n \to \frac{\partial}{\partial x^1} l_0.$$

In other words, the differentiation operator is a continuous operator on $\mathcal{S}$ in the topology of weak convergence of sequences.

PROOF. We prove this for differentiation:

$$\left\langle f, \frac{\partial}{\partial x^1} l_n \right\rangle = -\left\langle \frac{\partial f}{\partial x^1}, l_n \right\rangle.$$

Since $\frac{\partial f}{\partial x^1} \in \mathcal{S}$ if $f \in \mathcal{S}$, the right-hand side converges to $\left\langle -\dfrac{\partial f}{\partial x^1}, l_0 \right\rangle$. □

4.3.4 Theorem (Laurent Schwartz). *Let a mapping $\mathcal{F}_{\mathcal{S}'} : \mathcal{S}' \to \mathcal{S}'$ be defined by setting*

$$\langle f, \mathcal{F}_{\mathcal{S}'} l\rangle = \langle \widehat{f}, l\rangle.$$

Then $\mathcal{F}_{\mathcal{S}'}$ is an isomorphism from $\mathcal{S}'$ onto $\mathcal{S}'$, mapping weakly convergent sequences to weakly convergent sequences.

Moreover, $\mathcal{F}_{\mathcal{S}'}$ can be restricted to L^1 and L^2 by means of the inclusions $L^1 \subset \mathcal{S}'$, $L^2 \subset \mathcal{S}'$. The restriction of $\mathcal{F}_{\mathcal{S}'}$ to L^1 gives the Fourier integral; the restriction of $\mathcal{F}_{\mathcal{S}'}$ to L^2 gives the Fourier-Plancherel transform.

Finally, the inverse of $\mathcal{F}_{\mathcal{S}'}$ is given by

$$\mathcal{F}_{\mathcal{S}'}^{-1}(u) = \overline{\mathcal{F}_{\mathcal{S}'}(\overline{u})}, \quad \forall u \in \mathcal{S}'.$$

REMARK. If μ is a positive measure satisfying 4.3.1(i), $\mathcal{F}_{\mathcal{S}'}(\mu)$ is defined even though the integral $\widehat{\mu}(t)$ might diverge for every t.

PROOF. Fixing $l \in \mathcal{S}'$ and setting

$$\varphi(f) = \langle \widehat{f}, l \rangle,$$

we obtain a linear functional on $\mathcal{S}$ which, as the composition of continuous mappings, is itself continuous. Hence there exists $l_1 \in \mathcal{S}'$ such that $\varphi(f) = \langle f, l_1 \rangle$. Let

$$l_1 = \mathcal{F}_{\mathcal{S}'}(l).$$

Since $f \to \widehat{f}$ is an isomorphism of $\mathcal{S}$ onto $\mathcal{S}$, its transpose $\mathcal{F}_{\mathcal{S}'}$ is an isomorphism from $\mathcal{S}'$ onto $\mathcal{S}'$. Moreover, by Parseval's relation (cf. 2.6),

$$\langle \widehat{f}, u \rangle = \langle f, \widehat{u} \rangle, \quad \forall f \in \mathcal{S}, \quad \forall u \in L^1.$$

Hence $\mathcal{F}_{\mathcal{S}'}$ is an extension of the Fourier integral on L^1. The same result holds on L^2.

Finally, the inversion formula for $\mathcal{F}_{\mathcal{S}'}$ is proved by transposing the inversion formula on $\mathcal{S}$. □

4.3.5 Support of a distribution

Let $l \in \mathcal{S}'$. We say that l is *zero on the open set O* if $l(\varphi) = 0$ for any $\varphi \in \mathcal{S}(\mathbf{R}^n)$ such that $\operatorname{supp}(\varphi) \subset O$. Differentiable partitions of unity can be used to show that there exists a largest open set Ω on which l is zero. The complement of Ω is called the *support* of l.

4.3.6 Sobolev scales of distributions

For a fixed positive real number s, let $\mathcal{D}(\mathbf{R}^n)$ be given the H^{-s} norm defined by

$$\|\varphi\|_{H^{-s}} = \sup \int_{\mathbf{R}^n} \varphi f dx, \quad \text{where} \quad f \in H^s, \ \|f\|_{H^s} \leq 1.$$

Since $\mathcal{D}$ is dense in H^s, $\|\varphi\|_{H^{-s}} = 0$ implies that $\varphi = 0$.

Using the notation of Sobolev, we let $H^{-s}(\mathbf{R}^n)$ denote the completion of the space $\mathcal{D}$ with respect to the H^{-s} norm.

Theorem (Sobolev). *The Fourier transform extends from $\mathcal{D}$ to H^{-s} and realizes an isometric isomorphism from H^{-s} onto $L^2(\mathbf{R}^n, \mu_s)$, where $d\mu_s = (1 + \|t\|^2)^{-s} dt$.*

PROOF. If $f \in H^s$, then $f \in L^2$ and the Fourier-Plancherel isomorphism gives

$$\int_{\mathbf{R}^n} \varphi \overline{f} dx = \int_{\mathbf{R}^n} \widehat{\varphi}(t) \overline{\mathcal{F}} f(t) dt.$$

Hence

$$\|\varphi\|_{H^{-s}} = \sup \int \varphi(t) v(t), \quad \text{with} \quad \int |v(t)|^2 (1 + \|t\|^2)^s dt \leq 1.$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}\left|\int \widehat{\varphi}(t)\overline{v(t)}dt\right| &= \left|\int \widehat{\varphi}(t)(1+\|t\|^2)^{-s/2}(1+\|t\|^2)^{s/2}\overline{v}(t)dt\right| \\ &\leq \left[\int \frac{|\widehat{\varphi}(t)|^2}{(1+\|t\|^2)^s}dt\right]^{1/2} \|f\|_{H^s},\end{aligned}$$

whence

$$\|\varphi\|_{H^{-s}} \leq \left[\int \widehat{\varphi}(t)^2 \frac{dt}{(1+\|t\|^2)^s}\right]^{1/2}.$$

Equality occurs when $v(t) = c\widehat{\varphi}(t)(1+\|t\|^2)^{-s}$, with the constant c determined so that $\|v\|_{H^s} = 1$. □

4.3.7 Comparison of the two theories

(i) Proposition. *For every $s > 0$, $H^{-s}(\mathbf{R}^n) \subset \mathcal{S}'(\mathbf{R}^n)$.*

PROOF. $\mathcal{S}(\mathbf{R}^n) \subset H^s(\mathbf{R}^n)$. Moreover,

$$\|f\|_{m,r} \geq \|f\|_{H^s} \quad \text{if} \quad r \geq s,\ m > \frac{n}{2}.$$

Let $\theta \in H^{-s}$. Then θ defines a linear functional on H^s and

$$|\theta(f)| \leq c\|f\|_{H^s} \leq c\|f\|_{m,r} \quad \forall f \in H^r.$$

Hence θ is continuous on H^s if H^s is given the topology induced by that of $\mathcal{S}$. Restricting θ to $\mathcal{S}$ gives a continuous linear functional θ_1 on $\mathcal{S}$ and $\theta \mapsto \theta_1$ defines the desired map $H^{-s} \to \mathcal{S}'$.

This map is *injective*: $\mathcal{D}$ is dense in H^s; a fortiori, so is $\mathcal{S}$; thus a linear functional on H^s that vanishes on $\mathcal{S}$ is identically zero. □

(ii) Proposition. *Let $l \in \mathcal{S}'$ and suppose that l has compact support. Then there exists p such that $l \in H^{-p}(\mathbf{R}^n)$.*

PROOF. There exists a pair of integers m, r such that

$$|l(f)| < c\|f\|_{m,r} \quad \forall f \in \mathcal{S}(\mathbf{R}^n).$$

Let $\varphi \in \mathcal{D}(\mathbf{R}^n)$ such that $\varphi = 1$ on the support of l. Then $l(f\varphi) = l(f)$, whence $|l(f)| \leq c\|\varphi f\|_{m,r}$. But

$$\|\varphi f\|_{m,r} \leq c\|\varphi\|_{m,r}\|f\|_{W_r^\infty}.$$

Moreover, by the corollary to the trace theorem, $\|f\|_{L^\infty} \leq c\|f\|_{H^s}$ if $s > \frac{n}{2}$ and $\|f\|_{W_r^\infty} \leq c\|f\|_{H^{s+r}}$. Hence

$$|l(f)| \leq c\|f\|_{H^{s+r}}.$$

Thus l extends to a continuous linear functional on H^{s+r}, whence $l \in H^{-s-r}$. □

5 Pseudo-differential Operators

The Fourier transform on $\mathbf{R}^n$ diagonalizes linear differential operators with constant coefficients. This property leads to representation theorems for the solution of the homogeneous equation as a limit of sums of complex exponentials, as well as existence theorems for the nonhomogeneous equation. These theorems, due to Leon Ehrenpreis and Bernard Malgrange, use the Fourier transform in $\mathbf{C}^n$ as a fundamental tool.

Complex-analytic methods are needed to prove these theorems, which are naturally formulated in the context of Laurent Schwartz's theory of distributions.

To obtain such general results, we would need not only to study locally convex topologies on spaces of distributions and duality between locally convex spaces, but also to prove minimum modulus theorems for holomorphic functions of several complex variables. All these methods originate in different currents of thought from those we have followed up to now.

We will study differentiable operators with *variable* rather than constant coefficients, and on bounded open subsets of $\mathbf{R}^n$ rather than on all of $\mathbf{R}^n$. In physics, differentiable operators with variable coefficients invariably appear when an inhomogeneous medium is considered.

At first glance, Fourier analysis seems to have no means of obtaining results in this setting. It was thus a striking result when Alberto Calderon, in 1957, introduced an "infinitesimal Fourier transform on the tangent space", which assigns a "symbol" to an operator and thereby embeds differential operators in the wider class of *pseudo-differential operators.* In this class, one introduces an infinitesimal symbolic calculus which consists of multiplying symbols. Calderon's symbolic calculus theorem states that the symbolic calculus corresponds to the composition of operators modulo regularizing operators, i.e. with the gain of one derivative.

The *pseudo-inverse* of a differential operator can be explicitly constructed in integral form.

This section ends with an application of the pseudo-inverse, in the proof of the *elliptic regularity theorem.*

Pseudo-differential operators are a basic tool of the theory of partial differential equations. The spectral pseudo-decomposition they effect, and the integral estimates they entail, make up, to some degree, the extension of Sections 1 to 4 of this chapter.

5.1 Symbol of a differential operator

5.1.0 Notation

In order to distinguish clearly between the variables $x \in \mathbf{R}^n$ and $t \in \mathbf{R}^n$ of the function $f(x)$ and its Fourier transform $\widehat{f}(t)$, we set $\mathbf{R}^n = E$, where E is an n-dimensional vector space over $\mathbf{R}$, and write its dual as $\widehat{E}$. The dual pairing is denoted by

$$\langle x, \xi \rangle \text{ or } x.\xi, \quad \text{where } x \in E,\ \xi \in \widehat{E}.$$

For a fixed choice of volume measure on E, the Fourier transform is written

$$\widehat{f}(\xi) = \int_E f(x) \mathrm{e}^{ix.\xi} dx, \quad \text{where} \quad f \in L^1(E).$$

The volume measure $d\xi$ on $\widehat{E}$ is fixed so that, on $L^1(E) \cap L^2(E)$,

$$\int_E |f(x)|^2 dx = \int_{\widehat{E}} |\widehat{f}(\xi)|^2 d\xi.$$

Similarly, if $\varphi \in A(E)$, the Fourier inversion formula is written

$$\varphi(x) = \int_{\widehat{E}} \widehat{\varphi}(\xi) \mathrm{e}^{-ix.\xi} d\xi. \tag{i}$$

The two measures dx and $d\xi$ are called *associated.* The Fourier-Plancherel transform is an *isometry* of $L^2(E)$ onto $L^2(\widehat{E})$. We observe the convention of choosing a basis for E in such a way that the associated volume element is equal to 1. Under these conditions, we are led to define two bases e_k of E and e^k of $\widehat{E}$ as Fourier-dual if

$$\begin{array}{rcll} \langle e_j, e^k \rangle & = & 2\pi & \text{if} \quad k = j, \\ \langle e_j, e^k \rangle & = & 0 & \text{if} \quad k \neq j. \end{array}$$

Let $\mathcal{E}(E)$ be the vector space of infinitely differentiable functions on E, and let $\mathcal{D}(E)$ be the subspace of $\mathcal{E}(E)$ consisting of functions with compact support. We will consider differential operators of the form

$$L = \sum_{|m| \leq s} a_m(x) \partial^m, \tag{ii}$$

where $m = (m_1, \ldots, m_n)$ denotes a multi-index, that is a system of n non-negative integers. Let $|m| = m_1 + \ldots + m_n$, let $\partial_1 = \partial/\partial x^1, \ldots, \partial_n = \partial/\partial x^n$, and let $\partial^m = \partial_1^{m_1} \ldots \partial_n^{m_n}$. The coefficients $a_m(x)$ will be "sufficiently differentiable" functions of x. If L is not the zero operator, the largest $|m|$ such that $a_m \not\equiv 0$ is called its *order.*

Given $\varphi \in \mathcal{E}(E)$, we define

$$(L\varphi)(x) = \sum_{|m| \leq s} a_m(x) (\partial^m \varphi)(x).$$

If $a_m \in \mathcal{E}(E)$, then L defines a linear operator from $\mathcal{E}(E)$ to $\mathcal{E}(E)$. The *symbol* of the operator L is the function defined on $E \times \widehat{E}$ by

$$\sigma_L(x, \xi) = \mathrm{e}^{ix.\xi} L\varphi_\xi(x), \quad \text{where} \quad \varphi_\xi(x) = \mathrm{e}^{-ix.\xi}. \tag{iii}$$

Writing this out in a basis,

$$(iv) \qquad \sigma_L(x,\xi) = \sum a_m(x)(-i\xi_1)^{m_1}\dots(-i\xi_n)^{m_n}.$$

The symbol is thus a polynomial in ξ for every fixed x. The advantage of (iii) is that it is independent of the choice of basis, while (iv) appears to depend on the choice of basis.

A differential operator can evidently be reconstructed from its symbol; it suffices to write the symbol, in a basis, as a polynomial in ξ, and to substitute $i\partial_k$ for ξ_k in the monomials. This elementary calculation can be replaced by an integral expression, which has the immense advantage of being applicable to functions $\sigma(x,\xi)$ *more general than polynomials in* ξ.

5.1.1 Theorem. *Let L be a differential operator on E with symbol $\sigma_L(x,\xi)$. Then*

$$(i) \qquad (L\varphi)(x) = \int_{\widehat{E}} \sigma_L(x,\xi)\widehat{\varphi}(\xi)e^{-ix.\xi}d\xi, \quad \forall\varphi\in\mathcal{D}(E),$$

where $\widehat{\varphi}(\xi) = \int \varphi(x)e^{ix.\xi}dx$ denotes the Fourier transform of φ.

Proof. By 4.1(iv), $\widehat{\mathcal{D}(E)} \subset \widehat{\mathcal{S}(E)} = \mathcal{S}(\widehat{E})$, whence $\widehat{\varphi}$ is of rapid decrease. Thus $\sigma_L(x,\xi)\widehat{\varphi}(\xi)$ is of rapid decrease and the integral in (i) is convergent. Moreover, by differentiating the inversion formula

$$\varphi(x) = \int_{\widehat{E}} \widehat{\varphi}(\xi)\mathrm{e}^{-ix.\xi}d\xi$$

with respect to ∂_1, we obtain

$$(\partial_1\varphi)(x) = \int_{\widehat{E}} \widehat{\varphi}(\xi)(-i\xi_1)\mathrm{e}^{-ix.\xi}d\xi,$$

and more generally

$$(\partial_1^{m_1}\dots\partial_n^{m_n}\varphi)(x) = \int (-i\xi_1)^{m_1}\dots(-i\xi)^{m_n}\widehat{\varphi}(\xi)\mathrm{e}^{-ix.\xi}d\xi.$$

The theorem follows by multiplying both sides by $a_m(x)$, pulling $a_m(x)$ through the integral sign, and summing over m. □

5.2 Definition of a pseudo-differential operator on $\mathcal{D}(E)$

5.2.1 The class of symbols $\mathcal{C}(\beta,r,0)$

Let β be a real number and let r be a positive integer. We define a class $\mathcal{C}(\beta,r,0)$ of measurable functions q on $E\times\widehat{E}$ which satisfy the following two conditions.

(i) *q has compact support in x*; that is, there exists a compact subset K of E such that

$$q(x, \xi) = 0 \quad \text{if} \quad x \notin K, \quad \text{for any} \quad \xi \in \widehat{E}.$$

Derivatives with respect to x in $E \times \widehat{E}$ are denoted by ∂_x^m. The functions q are required to satisfy the following regularity condition.

(ii) $\|(1 + \|\xi\|)^\beta \partial_x^n q(x, \xi)\|_{L^\infty(E \times \widehat{E})} < +\infty$ for every multi-index n such that $|n| \leq r$.

5.2.2 Example. Let L be the differential operator of order s considered in 5.1.0(ii). If the coefficients of L are in W_r^∞, then

$$\sigma_L(x, \xi) \in \mathcal{C}(-s, r, 0).$$

It is clear from this example that, in the class $\mathcal{C}(\beta, r, 0)$, the integer r corresponds to the *regularity of the coefficients* and the number $-\beta$ to the order of the operator.

5.2.3 Pseudo-differential operators defined on $\mathcal{D}(E)$

With a given symbol $g \in \mathcal{C}(\beta, r, 0)$ and function $\varphi \in \mathcal{D}(E)$, we associate the function

$$(A_g \varphi)(x) = \int_{\widehat{E}} g(x, \xi) \widehat{\varphi}(\xi) \mathrm{e}^{-ix.\xi} d\xi. \tag{i}$$

The integral converges since, for fixed x, $\widehat{\varphi}$ is of rapid decrease in ξ and g is of polynomial growth in ξ. Differentiating under the integral sign with respect to x shows that $A_g \varphi \in W_r^\infty$, and it follows from 5.2.1 that $A_g \varphi$ has compact support. All these observations are trivial; the following theorem is not.

5.3 Extension of pseudo-differential operators to Sobolev spaces

5.3.0 Theorem. *Let $g \in \mathcal{C}(\beta, r + 1, 0)$ and let $n = dim(E)$. Assume that $s \geq -\beta$ satisfies*

$$0 \leq s < r - n. \tag{i}$$

Then there exists a constant c_s such that

$$\|A_g \varphi\|_{H^s} \leq c_s \|\varphi\|_{H^{s+\beta}}, \quad \forall \varphi \in \mathcal{D}(E). \tag{ii}$$

(iii) There exists a unique extension of A_g to a bounded operator A_g^s from $H^s(E)$ to $H^{s+\beta}(E)$.

PROOF. Statement (iii) follows from the density of $\mathcal{D}(E)$ in $H^s(E)$ and from inequality (ii).

Since the H^s norms can be computed in terms of the Fourier transform, (ii) can be expressed as an inequality between Fourier transforms. Since $A_g\varphi$ is a bounded function with compact support, its Fourier transform can be computed. This computation leads to the following lemma.

5.3.1 Lemma.

(i)
$$\widehat{A_g\varphi}(\eta) = \int_{\widehat{E}} K_g(\eta,\xi)\widehat{\varphi}(\xi)d\xi,$$

where

(ii)
$$K_g(\eta,\xi) = \int_E g(x,\xi)e^{ix.(\eta-\xi)}dx.$$

PROOF.

$$(\widehat{A_g\varphi})(\eta) = \int_E e^{ix.\eta}(A_g\varphi)(x)dx = \int_E e^{ix.\eta}dx \int_{\widehat{E}} g(x,\xi)\widehat{\varphi}(\xi)e^{-ix.\xi}d\xi.$$

The double integral $\int_{E\times\widehat{E}} |g(x,\xi)\widehat{\varphi}(\xi)|d\xi dx$ converges: it is bounded above by

$$\left\|(1+\|\xi\|)^\beta g(x,\xi)\right\|_{L^\infty(E\times\widehat{E})} \operatorname{meas}(K') \int_{\widehat{E}} |\widehat{\varphi}(\xi)|(1+\|\xi\|)^{-\beta}d\xi,$$

where K' denotes the *support in x* of the symbol, and the integral on $\widehat{E}$ converges because $\widehat{\varphi}$ is of rapid decrease. Hence Fubini's theorem can be applied to reverse the order of integration:

$$\widehat{A_g\varphi}(\eta) = \int_{\widehat{E}} \widehat{\varphi}(\xi)d\xi \int_E g(x,\xi)e^{ix.(\eta-\xi)}dx.$$

Fubini's theorem guarantees that the integral on E converges for almost every ξ. Since g has compact support in x, it actually converges for every ξ, and there exists a constant c such that

(iii)
$$|K_g(\eta,\xi)| \leq c(1+\|\xi\|)^{-\beta}\operatorname{meas}(K').$$

5.3.2 Estimating the kernel K_g

Lemma. *Suppose that $g \in \mathcal{C}(\beta, r+1, 0)$ and let r' be the integer defined by $r \leq 2r' \leq r+1$. Then*

(i)
$$|K_g(\eta,\xi)| \leq (1+\|\xi\|)^{-\beta}(1+\|\xi-\eta\|^2)^{-r'}.$$

PROOF. Let $\{x^k\}$ be an orthonormal basis with respect to the metric $\|x\|$. In terms of this basis, the Laplace operator on E is defined by

$$\Delta_x = \sum_{k=1}^n (\partial_{x^k})^2.$$

Then

$$\Delta_x e^{ix.\xi} = -\|\xi\|^2 e^{ix.\xi}.$$

Let $A_x^{r'}$ be the differential operator on E with constant coefficients defined by

$$A_x^{r'} = (1-\Delta_x)^{r'} = 1 - r'\Delta + \cdots + (-1)^{r'}\Delta^{r'}.$$

Then $A_x^{r'} e^{ix.(\eta-\xi)} = (1+\|\eta-\xi\|^2)^{r'} e^{ix.(\eta-\xi)}$, whence

$$K_g(\eta,\xi)(1+\|\eta-\xi\|^2)^{r'} = \int_E g(x,\xi) A_x^{r'} e^{ix.(\eta-\xi)}dx.$$

Since $g(x,\xi)$ has compact support, we can integrate by parts and turn derivatives of the exponential into derivatives on g. Thus

$$(1+\|\eta-\xi\|^2)^{r'} K_g(\eta,\xi) = \int (A_x^{r'} g(x,\xi)) e^{ix.(\eta-\xi)}dx,$$

and (i) follows by 5.2.1(ii).

5.3.3 Proof of the extension theorem

(i) Lemma. *Let $f \in L^2(E)$ and let $\mathcal{F}(f)$ denote the Fourier transform of f. Then $f \in H^s$ if and only if $\mathcal{F}(f)(\xi) = (1+\|\xi\|)^{-s}k(\xi)$ with $k \in L^2(\widehat{E})$.*

PROOF. Cf. 3.5.1. □

(ii) Lemma. *Let $\widetilde{K}_g(\eta,\xi) = K(\eta,\xi)\left(\frac{1+\|\eta\|}{1+\|\xi\|}\right)^s (1+\|\xi\|)^{-\beta}$. Let*

$$(G_g f)(\eta) = \int \widetilde{K}_g(\eta,\xi) f(\xi) d\xi.$$

Then 5.3.0(ii) is equivalent to the inequality

$$\|G_g f\|_{L^2(\widehat{E})} \le c\|f\|_{L^2(\widehat{E})}, \quad \forall f \in L^2(\widehat{E}).$$

PROOF. By 5.3.3(i) and 5.3.1(ii). □

(iii) Lemma.

$$|\widetilde{K}_g(\eta,\xi)| \le c(1+\|\xi-\eta\|^2)^{-r'+\frac{s}{2}}.$$

PROOF. This follows from the inequality

$$\left(\frac{1+\|\eta\|}{1+\|\xi\|}\right)^s \le 2^s(1+\|\eta-\xi\|^s),$$

which is proved by considering the following two cases:

(a) $\|\eta\| \le 2\|\xi\|$. Observe that the left-hand side is less than or equal to 2^s.

(b) $\|\eta\| > 2\|\xi\|$. Observe that $1+\|\eta\| \le 1+2\|\xi-\eta\|$. □

(iv) CONCLUSION. To prove 5.3.0(ii), note that $|G_g f|$ is bounded above by replacing $\widetilde{K}_g$ with an upper bound for $|\widetilde{K}_g|$. Using 5.3.3(iii), it must be shown that

$$\left\| \int \frac{|f(\xi)|}{(1+\|\xi-\eta\|^2)^{r'-s/2}} d\xi \right\|_{L^2(E)} \leq c\|f\|_{L^2(E)}.$$

The left-hand side can be written formally as $|f| * u$, where $u(\xi) = (1+\|\xi\|^2)^{-r'+s/2}$.

Next, $-2r'+s \leq s-r < -n$ by 5.3.0(i), whence $u \in L^1$. Finally, using 1.8.2,

$$\|\, |f| * u \,\|_{L^2} \leq \|u\|_{L^1}\|f\|_{L^2}.\square$$

5.4 *Calderon's symbolic pseudo-calculus*

5.4.0 Motivation

The Fourier transform maps a differential operator L^0 with constant coefficients to multiplication by the symbol $\sigma_{L^0}(\xi)$. (The hypothesis of constant coefficients is reflected in the fact that the symbol no longer depends on x.) Thus the composition of constant-coefficient differential operators L^0 and M^0 — that is, the differential operator $Q^0 = L^0 M^0$ — corresponds to the product of symbols $\sigma_{Q^0} = \sigma_{L^0}\sigma_{M^0}$. The differential operators with constant coefficients form a *commutative* algebra for which the Fourier transform makes possible, to some extent, a spectral theory.

The differential operators with variable, but infinitely differentiable, coefficients also form an *algebra*: two such operators can be composed. But this algebra is *no longer commutative.*

For example, consider the differential operators $L = x^1 \frac{\partial}{\partial x^1}$ and $M = \frac{\partial}{\partial x^1}$ on $\mathbf{R}$. Then

$$LM = x^1\left(\frac{\partial}{\partial x^1}\right)^2, \quad ML = \frac{\partial}{\partial x^1} + x^1\left(\frac{\partial}{\partial x^1}\right)^2, \quad \text{and} \quad LM - ML = \frac{\partial}{\partial x^1}.$$

Commutativity has been lost. Nevertheless, the commutator $LM - ML$ is an operator of order 1, while the product is an operator of order 2. One might say that commutativity is preserved, modulo operators of lower order.

5.4.1 Introduction to the classes $\mathcal{C}(\beta, r, 1)$

A subclass $\mathcal{C}(\beta, r, 1)$ of the symbols $\mathcal{C}(\beta, r, 0)$ is defined by imposing the following additional axiom:

(i) $\|(1+\|\xi\|)^{\beta+1}(\partial_x^m \partial_\xi q)(x,\xi)\|_{L^\infty(E\times\widehat{E})} < +\infty, \quad \forall m$ such that $|m| \leq r$.

Similarly, a class $\mathcal{C}(\beta, r, s)$ could be defined by differentiating s times with respect to ξ instead of once, and replacing β by $\beta + s$. These classes would appear in computing multiple commutators; such computations would arise from taking limits that we have held fixed.

Pseudo-products

Let p and q be the symbols of the pseudo-differential operators A_p and A_q. The *pseudo-product* of A_p and A_q is the operator whose symbol is the product of the symbols. This operator is written $A_q \Box A_p$ and, by definition,

$$A_{qp} = A_q \Box A_p.$$

With the formula for the derivative of a product, it is easy to verify that

(ii) if $g \in \mathcal{C}(\beta, r, 0)$ and $h \in \mathcal{C}(\beta', r, 0)$, then $gh \in \mathcal{C}(\beta + \beta', r, 0)$.

The pseudo-product is a commutative operation and therefore cannot correspond to the composition of operators. However, it does give an approximation.

5.4.2 Calderon's commutation theorem. *Let $p \in \mathcal{C}(\beta, 2r+2, 0)$ and let $q \in \mathcal{C}(\beta', r+1, 1)$. Suppose that $r \geq \beta' + 1$. Set*

$$R = A_q A_p - A_q \Box A_p.$$

Then, for s such that $0 \leq s < n - r$,

$$(i) \qquad R : H^s(E) \to H^{s+\beta+\beta'+1}(E)$$

and there exists a constant c_s such that

$$(ii) \qquad \|Rf\|_{H^{s+\beta+\beta'+1}} \leq c_s \|f\|_{H^s}.$$

PROOF. Since $\mathcal{D}(E)$ is dense in H^s, it suffices to prove (ii) when $f \in \mathcal{D}(E)$. As in the extension theorem, we take the Fourier transform of both sides of the inequality. For $f \in \mathcal{D}(E)$, let

$$\widetilde{A}_p f(\eta) = \widehat{A_p f} = \int_{\widehat{E}} K_p(\eta, \xi) \widehat{f}(\xi) d\xi.$$

The kernel K_p was computed in 5.3.1.

The proof of this theorem will require several lemmas.

5.4.3 Lemma.

$$(i) \qquad (\widetilde{A}_q(\widetilde{A}_p f))(\lambda) = \int G_{q,p}(\lambda, \xi) \widehat{f}(\xi) d\xi, \quad \textit{where}$$

$$(ii) \quad G_{q,p}(\lambda, \xi) = \int_{\widehat{E}} d\mu \int\int_{E^2} p(z+h, \xi) q(z, \xi - \mu) e^{i\mu h} e^{iz(\xi - \lambda)} dh\, dz.$$

PROOF. Composing the kernels gives

$$G_{q,p}(\lambda, \xi) = \int_{\widehat{E}} K_q(\lambda, \eta) K_p(\eta, \xi) d\eta.$$

Replacing K_p and K_q by the expressions given in Lemma 5.3.1,

$$G_{q,p} = \int_{\widehat{E}} d\eta \int_{E^2} p(x,\xi)q(z,\eta)e^{ix(\xi-\eta)+iz(\eta-\lambda)}dx\ dz.$$

Setting $x = z + h$ and $z = z$ in E^2, and $\eta = \xi - \mu$ in $\widehat{E}$, we obtain

$$G_{q,p} = \int_{\widehat{E}} d\mu \int_{E^2} p(z+h,\xi)q(z,\xi-\mu)e^{i\mu h+iz(\xi-\lambda)}dz\ dh.\square$$

5.4.4 Lemma. *Let O be a compact subset of E containing the supports of $p(x,.)$ and $q(x,.)$. Then there exists an even function $u \in \mathcal{D}(E)$ such that $u(x_1 - x_2) = 1$ if $x_1,\ x_2 \in O$,*

$$\text{(i)}\quad G_{q,p}(\lambda,\xi) = \int_{\widehat{E}} d\mu \int\int_{E^2} p(z+h,\xi)q(z,\xi-\mu)e^{iz(\xi-\lambda)}u(h)e^{i\mu h}dh\ dz,$$

and

$$\text{(ii)}\qquad 1 = \int_E d\mu \left[\int_E u(h)e^{i\mu h}dh\right].$$

PROOF. Let

$$O_1 = \{y \in E : y = x_1 - x_2,\ x_i \in O\}.$$

Then O_1 is a compact subset of E containing the origin. There exists a function $u \in \mathcal{D}(E)$ equal to 1 on O_1.

The right-hand side of formula 5.4.3(ii) is nonzero if $z+h \in O$ and $z \in O$; that is, if $h \in O_1$. Multiplication by $u(h)$ is multiplication by 1; this proves formula (i).

The second formula is obtained by applying the Fourier inversion formula to $u \in \mathcal{D}(E)$ and noting that, since the origin is in O_1, $u(0) = 1.\square$

REMARK. We must be careful not to write a double integral in (ii), since Fubini's theorem does not apply. Similarly, 5.4.3(ii) cannot be written as a triple integral.

5.4.5 Lemma. $(G_{q,p} - K_{qp})(\lambda,\xi) = I(\lambda,\xi) + J(\lambda,\xi)$, *where*

$$\text{(i)}\qquad I(\lambda,\xi) = \int\int_{E\times\widehat{E}} \widehat{u}(\mu)p(z,\xi)[q(z,\xi-\mu) - q(z,\xi)]e^{iz(\xi-\lambda)}dzd\mu$$

$$\text{(ii)}\qquad J(\lambda,\xi) = \int_{\widehat{E}} d\mu \int_E q(z,\xi-\mu)e^{iz(\xi-\lambda)}dz\left[\int_E \dots dh\right],\quad \textit{and}$$

$$\left[\int_E \dots dh\right] = \int_E (p(z+h,\xi) - p(z,\xi))e^{i\mu h}u(h)dh.$$

PROOF. Formulas (i) and (ii) of Lemma 5.4.4 and 5.4.3(i) imply that

$$\begin{aligned}
G_{p,q} - K_{pq} &= \textstyle\int\int_{E\times\widehat{E}} e^{iz(\xi-\lambda)}dzd\mu \int e^{i\mu h}[\ \ ]u(h)dh, \quad \text{where}\\
[\ \ ] &= p(z+h,\xi)q(z,\xi-\mu) - p(z,\xi)q(z,\xi) = [\ \ ]_1 + [\ \ ]_2 \quad \text{with}\\
[\ \ ]_1 &= p(z,\xi)(q(z,\xi-\mu) - q(z,\xi)) \quad \text{and}\\
[\ \ ]_2 &= q(z,\xi-\mu)(p(z+h,\xi) - p(z,\xi)).
\end{aligned}$$

Note that the first term no longer contains h; hence the integration in h affects only $e^{i\mu h}u(h)$, which, since u is even, gives $\widehat{u}(\mu)$. Thus we have

$$I(\lambda,\xi) = \int_{\widehat{E}} d\mu \int_E \widehat{u}(\mu)[\quad]_1 \exp((iz(\xi-\lambda))dz.$$

Since $\widehat{u} \in L^1$, Fubini can now be applied to obtain (i). Integrating the expression $[\quad]_2$ and applying Fubini to the integral $\int\int_{E^2}$ yield (ii). □

5.4.6 Lemma. *Set*

$$g(z,\xi) = \int_{\widehat{E}} (q(z,\xi-\mu) - q(z,\xi))\widehat{u}(\mu)d\mu,$$

and let

$$l(z,\xi) = p(z,\xi)g(z,\xi).$$

Then $I = K_l$.

PROOF. Integrate 5.4.3(i) with respect to μ, then use Lemma 5.3.1. □

5.4.7 Estimating the integral I

We use the extension theorem 5.3.0 to show that

$$(i) \qquad g \in \mathcal{C}(\beta'+1, r+1, 0).$$

5.4.1(ii) will then imply that $pg \in \mathcal{C}(\beta+\beta'+1, r+1, 0)$.

We first use Taylor's formula with integral remainder on $\widehat{E}$ to obtain

$$g(z,\xi) = \int_{\widehat{E}} \left(\sum\nolimits_k q_k(z,\xi,\mu)\mu_k\right) \widehat{u}(\mu)d\mu, \quad \text{where}$$
$$q_k(z,\xi,\mu) = -\int_0^1 (\partial_{\xi_k} q(z,\xi-t\mu))dt.$$

Differentiating with respect to z gives

$$\partial_z^m(g(z,\xi)) = \sum_k \int_{\widehat{E}} (\partial_z^m q_k(z,\xi,\mu))\widehat{u}(\mu)\mu_k d\mu,$$

whence, by 5.4.1(i),

$$(ii) \qquad |\partial_z^m(g(z,\xi))| < c\int_E \int_0^1 (1+\|\xi-t\mu\|)^{-\beta'-1}\|\mu\|\ \widehat{u}(\mu)dtd\mu.$$

Let $v(\mu) = \|\mu\|\ |\widehat{u}(\mu)|$. Then v is of rapid decrease. Set

$$F(t,\xi) = \int (1+\|\xi-t\mu\|)^{-\beta'-1} v(\mu)d\mu = \int_A + \int_{A^c},$$

where $A = \{\mu : \|\mu\| > \frac{\|\xi\|}{2t}\}$. For any integer m,

$$\int_A \leq \int \mathbf{1}_A v \leq c_m \left(1 + \frac{\|\xi\|}{t}\right)^{-m}.$$

Moreover,

$$\int_{A^c} \leq c(1 + \|\xi\|)^{-\beta'-1}\|v\|_{L^1}.$$

Hence, taking $m \geq \beta' + 1$,

$$(iii) \qquad F(t,\xi) < C_1(1 + \|\xi\|)^{-\beta'-1},$$

where the constant C_1 is independent of t. Integrating with respect to t gives (i). □

5.4.8 Estimating the integral J

We now use Taylor's formula with integral remainder on E. Set

$$\varphi_k(z,h,\xi) = \int_0^1 (\partial_{x^k} p)(z + th, \xi)dt.$$

Then $p(z+h,\xi) - p(z,\xi) = \sum h^k \varphi_k$. Writing $ih_k = \partial_{\mu k} \mathrm{e}^{i\mu.h}$ gives

$$-\int_E (p(z+h,\xi) - p(z,\xi))\mathrm{e}^{ih.\mu}u(h)dh = i\int_E \sum \varphi_k(\partial_{\mu k}\mathrm{e}^{ih.\mu})u(h)dh.$$

Since μ appears only in the exponential terms and we can differentiate under the integral sign,

$$\int_E (p(z+h,\xi) - p(z,\xi))\mathrm{e}^{ih.\mu}u(\mu)d\mu = \sum \partial_{\mu_k}\psi_k(z,\mu,\xi),$$

where

$$(i) \qquad -\psi_k(z,\mu,\xi) = i\int_E \varphi_k(z,h,\xi)\mathrm{e}^{ih.\mu}u(h)dh.$$

Since φ_k and u are sufficiently differentiable in h and u has compact support, it follows that ψ_k, which can be regarded as a Fourier transform in h, vanishes at infinity together with its first derivative. Substituting into 5.4.5(ii) and reversing the order of integration, we can thus integrate by parts on $\widehat{E}$ with respect to μ_k, and we obtain

$$J(x,\xi) = \sum_k \int_E \mathrm{e}^{iz(\xi-\lambda)}dz \int_{\widehat{E}} (\partial_{\mu_k} q)(z, \xi - \mu)\psi_k(z,\mu,\xi)d\mu.$$

Let

$$g_k(z,\xi) = \int_{\widehat{E}} (\partial_{\mu_k} q)(z, \xi - \mu)\psi_k(z,\mu,\xi)d\mu.$$

Then $J = \sum_k K_{g_k}$. We now show that

$$(ii) \qquad g_k \in \mathcal{C}(\beta + \beta' + 1, r + 1, 0)$$

by finding an upper bound for

$$U_k(z, \xi) = \sum_{m_1, m_2} \int_{\widehat{E}} (\partial_z^{m_1} \partial_{\mu_k} q)(z, \xi - \mu) \partial_z^{m_2} \psi_k(z, \mu, \xi) d\mu,$$

with $|m_1| + |m_2| \leq r + 1$. Since $q \in \mathcal{C}(\beta', r, 1)$,

$$(iii) \qquad |U_k(z, \xi)| \leq c \sum_{m_2} \int_{\widehat{E}} (1 + \|\xi - \mu\|)^{-\beta' - 1} |\partial_z^{m_2} \psi_k(z, \mu, \xi)| d\mu.$$

(iv) Lemma. *There exists a constant c, independent of ξ, such that*

$$(1 + \|\xi\|)^{\beta} |\partial_z^{m_2} \psi_k(z, \mu, \xi)| < \frac{c}{(1 + \|\mu\|^2)^{r'}},$$

where r' is the integer such that $r \leq 2r' \leq r + 1$.

PROOF. Using (i),

$$\int_E ((-\Delta_h + 1)^{r'} [u(h) \partial_z^{m_2} (\varphi_k(z, h, \xi))]) e^{ih.\mu} dh = (1 + \|\mu\|^2)^{r'} \partial_z^{m_2} \psi_k(z, \mu, h).$$

The inequality follows, with

$$c = \operatorname{meas}(K) \|(-\Delta_h + 1)^{r'} [u(h) \partial_z^{m_2} \varphi_k(z, \mu, h)\|_{L^\infty(E \times \widehat{E})}.$$

Here K is the support of p in x.

The following lemma, 5.4.9, together with (iii), (iv), and the hypothesis that $r > \beta' + 1$, imply that

$$|U_k(z, \xi)| < c(1 + \|\xi\|)^{-\beta} (1 + \|\xi\|)^{-\beta' - 1}.$$

That is, (ii) is proved, and with it the commutation theorem 5.4.2. □

5.4.9 Lemma. *Let r be a positive number and let*

$$h_r(\eta) = (1 + \|\eta\|)^{-r}, \quad \textit{where} \quad \eta \in \widehat{E}, \ \dim \widehat{E} = n.$$

Then, if $r > n$ and $s \geq 0$,

$$h_r * h_s \leq c(r, s) h_t, \quad \textit{where} \quad t = \inf(r, s).$$

PROOF. Let

$$(h_r * h_s)(\eta) = \int \frac{1}{(1 + \|\lambda\|)^r} \frac{1}{\|\lambda - \eta\| + 1)^s} d\lambda = \int_{A^c(\eta)} + \int_{A(\eta)},$$

where $A(\eta) = \{\lambda : \frac{1}{2}\|\eta\| \leq \|\lambda\| \leq 2\|\eta\|\}$. Then

$$\int_{A^c(\eta)} \leq \max_{\lambda \in A^c(\eta)} \left(\frac{1}{\|\lambda - \eta\| + 1)^s}\right) \|h_r\|_{L^1} \leq \frac{c}{(1 + \|\eta\|)^s} \quad \text{and}$$

$$\int_{A(\eta)} \leq \max_{\lambda \in A(\eta)} \frac{1}{(1 + \|\lambda\|)^r} \|\mathbf{1}_{A(\eta)} \tau_\eta(h_s)\|_{L^1},$$

where τ denotes the translation operator.

(i) If $s > n$, then $\|\mathbf{1}_{A(\eta)} \tau_\eta(h_s)\|_{L^1} \leq \|h_s\|_{L^1} < +\infty$. Hence

$$\int_{A(\eta)} \leq c \frac{1}{(1 + \|\lambda\|)^r}$$

and

$$h_r * h_s \leq c(h_r + h_s) \leq ch_t.$$

(ii) If $s \leq n$, then $\|\mathbf{1}_{A(\eta)} \tau_\eta(h_s)\|_{L^1} \leq \int_{\|\eta\| < 2\|\xi\|} h_s = c(1 + \|\xi\|)^{n-s}$ and

$$h_r * h_s \leq ch_s + c(1 + \|\xi\|)^{n-s-r} \leq c(h_s + h_{s+r-n}).$$

The conclusion follows by noting that $s + r - n > s$, whence $h_r * h_s \leq ch_s$. Since $s \leq n < r$, the lemma is proved. □

5.5 *Elliptic regularity*

5.5.0 **Definition.** Let L be a differential operator defined on an open subset O of R^n:

$$L = \sum_{|m| \leq d} a_m(x) \partial^m.$$

Let $\sigma_L(x, \xi)$ be its symbol. L is said to be an *elliptic operator* if, for every compact subset K of O, there exist two constants c_1, c_2 depending on K such that

$$(i) \qquad |\sigma_L(x, \xi)| \geq c_1 \|\xi\|^d \text{ if } x \in K \text{ and if } \|\xi\| > c_2.$$

(ii) Example. Consider the Cauchy-Riemann operator on $\mathbf{R}^2$,

$$L_0 = \frac{\partial}{\partial x^1} + i \frac{\partial}{\partial x^2} \quad (\text{where } i = \sqrt{-1}).$$

Then

$$\sigma_{L_0}(\xi) = -i\xi_1 + \xi_2 \quad \text{and} \quad |\sigma_{L_0}(\xi)| = \|\xi\|.$$

(iii) Example. On $\mathbf{R}^n$, consider the operator

$$L_1 = -\sum a^{kj}(x) \partial_k \partial_j + C^k(x) \partial_{x^k} + q(x),$$

where the matrix a_{ij} is symmetric and positive definite. Then

$$\sigma_{L_1}(x, \xi) = \sum a^{kj}(x) \xi_k \xi_j - iC^k(x) \xi_k + q(x).$$

For $\|\xi\|$ sufficiently large, the quadratic form dominates the first-order terms and L_1 is elliptic.

5.5.1 Theorem. *Let L be an elliptic operator of order d defined on the open set O. Suppose that the coefficients of L are functions in W^{∞}_{2r+2}, where $r > d+1$, $r > n$. Let $f \in H^d_{loc}(O)$; then Lf is well defined and $Lf \in L^2_{loc}(O)$. Under these hypotheses, the following two statements are equivalent:*

$$(i) \qquad Lf \in H^s_{loc}(O), \quad \textit{where} \quad 0 \leq s < n - r.$$
$$f \in H^{s+d}_{loc}(O), \quad \textit{where} \quad 0 \leq s < n - r. (ii)$$

PROOF. It is trivial that (ii) $\Rightarrow$ (i).

In order to prove that (i) $\Rightarrow$ (ii), we construct a *local pseudo-inverse* of L. Here *pseudo-inverse* means an inverse in the sense of Calderon's symbolic pseudo-calculus, and *local* means on a compact subset of O. Let O_1 be an open set such that $\overline{O}_1 \subset O$. Let φ and ψ be elements of $\mathcal{D}(O)$ such that $\varphi = 1$ on $\overline{O}_1$ and $\psi = 1$ on the support of φ. Let $L_1 = \varphi L$, $u = Lf$, $u_1 = \varphi u$, and $f_1 = \psi f$. Then $u_1 \in H^s(E)$, $f_1 \in H^d(E)$, and $\varphi L(\psi f) = \varphi L(f)$ since $\psi = 1$ on the support of φ. Hence

$$(i) \qquad L_1 f_1 = u_1.$$

Let $\sigma_{L_1}(x,\xi)$ be the symbol of L_1. Then σ_{L_1} has compact support in x (since its support is contained in the support of φ). Let $\theta \in \mathcal{D}(\widehat{E})$ be equal to 1 if $\|\xi\| \leq c_2(\overline{O}_1)$. Set

$$g(x,\xi) = \varphi(x)(1-\theta(\xi))[\sigma_L(x,\xi)]^{-1}.$$

Then it follows from 5.5.0(i) that $g \in \mathcal{C}(d,r,1)$.

Moreover, let $g\sigma_{L_1} = p$, where $p(x,\xi) = \varphi^2(x)(1-\theta(\xi))$. Multiplying the two sides of (i) by A_g gives

$$A_g L_1 f_1 = A_g u_1 = v, \quad \text{where} \quad v \in H^{s+d}.$$

Set $\widetilde{\theta}(x) = \widehat{\theta}(-x)$. Then

$$(A_p f_1)(x) = \varphi^2(x)[f_1(x) - (\widetilde{\theta} * f_1)(x)].$$

By the commutation theorem (5.4.2),

$$A_g L_1 = A_p + R, \quad \text{where} \quad R : H^{s'} \to H^{s'+1}.$$

Since $\varphi^2(x)(\widetilde{\theta} * f_1)(x) \in \mathcal{D}(E)$, it follows that

$$(ii) \qquad \varphi^2 f_1 + R f_1 = W, \quad \text{with} \quad W \in H^{s+d}(E).$$

$Rf_1 \in H^{d+1}$ since $f_1 \in H^d$; thus $\varphi^2 f_1 \in H^{d+1}$ if $s > 1$. As this is true for every $\varphi \in \mathcal{D}(O)$,

$$f_1 \in H^{d+1}_{loc}(O), \quad \text{or} \quad f\psi \in H^{d+1}_{loc}(O).$$

This last relation holds for every ψ; hence $f \in H^{d+1}_{loc}(O)$, and we have *gained a degree of differentiability.* Working backwards, we conclude that $f_1 \in H^{d+1}$ and therefore $Rf \in H^{d+2}$.

Substituting into (ii) gives

$$\varphi^2 f = W - Rf, \quad \text{with} \quad W \in H^{s+d} \quad \text{and} \quad Rf \in H^{d+2}.$$

Hence, if $s \geq 2$, $f \in H^{d+2}(O_1)$; as this is true for all O_1, we conclude that $f \in H^{d+2}_{loc}(O)$.

Substituting again into (ii), we find that $\varphi f_n \in H^{d+3}(E)$ for $s \geq 3$, and hence that $f \in H^{d+3}_{loc}(O_1)$. As this is true for all O_1, it follows that $f \in H^{d+3}_{loc}(O)$. Hence

$$(\psi f) \in H^{d+3}(E) \quad \text{and} \quad R(\psi f) \in H^{d+4}(E).$$

Substituting a third time into (ii) gives, as before,

$$\varphi^2 f \in H^{d+4}(E) \quad \text{if} \quad s \geq 4.$$

We continue this procedure until forced to stop, when $d + j > s + d$. The last possible step gives

$$\varphi^2 f \in H^{s+d}(E), \quad \text{whence} \quad f \in H^{s+d}_{loc}(O).\square$$

REMARK. With appropriate hypotheses on the differentiability of symbols, it is possible to let pseudo-differential operators act on Sobolev spaces of negative order and obtain the following improvement of the elliptic regularity theorem (5.5.1). *Let L be an elliptic operator of order d with infinitely differentiable coefficients, and let s be a real number. Then $Lf \in H^s_{loc}$ implies that $f \in H^{s+d}_{loc}$.*

IV

Hilbert Space Methods and Limit Theorems in Probability Theory

1 Foundations of Probability Theory

1.1 Introductory remarks on the mathematical representation of a physical system

Before we introduce the notion of probability, it seems advisable to describe the type of mathematical model used to represent a physical system.

Representations can be given from two distinct points of view:

- the point of view of essences, or
- the point of view of phenomena.

The point of view of essences, generally that of the pure mathematician, consists of thinking that the physical system can be perfectly known. The space of all possible states is introduced, and a state is a point in the *space of states*. This point of view is, for instance, that of rational mechanics: the state of a system of n physical points is completely determined by a point in $\mathbf{R}^{6n}$ (position and velocity of each of the particles).

The point of view of phenomena, generally that of the experimental physicist, consists of observing a few facts which occur in a physical system so complex that the physicist, at the outset, concedes that he will never understand its basic structure. For example, the physicist can use thermodynamics to analyze the phenomena of a gas without having to determine the state of all its molecules.

The mathematical model corresponding to a phenomenological representation is based on a logical calculus. The physicist introduces *the set $\mathcal{B}$ of all events* that he will be in a position to observe in studying the physical system. $\mathcal{B}$ is given the structure of the logical calculus, in which

- $A_1 + A_2$ denotes the occurrence of the event A_1 *or* the event A_2;
- $A_1.A_2$ denotes the occurrence of *both* the event A_1 *and* the event A_2; and
- $\emptyset$ denotes the impossible event and $\mathbf{1}$ the sure event.

The set $\mathcal{B}$ of all events thus forms an *abstract Boolean algebra.* (See I-1 for the definition of Boolean algebras of sets.)

The phenomenological point of view, initially of more modest scope than the point of view of essences, is much more adaptable to describing gains in knowledge. Indeed, a physical system described twenty years ago by a Boolean algebra $\mathcal{B}_0$ of events can be described today, after a more detailed analysis, by a Boolean algebra $\mathcal{B}_1$. All the events that appeared twenty years ago in $\mathcal{B}_0$ will appear in $\mathcal{B}_1$. Thus there is an injective mapping

$$\mathcal{B}_0 \to \mathcal{B}_1,$$

which commutes with the operations of the logical calculus and permits $\mathcal{B}_0$ to be identified with a subalgebra of $\mathcal{B}_1$. Progress in understanding the system is described by a sequence of Boolean algebras,

$$\mathcal{B}_0 \to \mathcal{B}_1 \to \mathcal{B}_2 \to \mathcal{B}_3 \to \dots,$$

where the arrows are injective homomorphisms of Boolean algebras. This sequence will give progressively more detailed representations of the physical system, although it may never arrive at a final representation that would correspond to complete understanding, beyond the reach of the experimenter.

1.2 Axiomatic definition of abstract Boolean algebras

A *Boolean algebra* is a set $\mathcal{B}$ together with two commutative and associative operations, written

$$A \cup A' \quad \text{and} \quad A \cap A'.$$

Each of the two operations is assumed to be distributive with respect to the other; that is,

$$\begin{aligned} A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \quad \text{and} \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C). \end{aligned}$$

We assume further that there exist two elements $\emptyset$ and $\mathbf{1}$ in $\mathcal{B}$ such that

$$A \cup \emptyset = A, \ A \cap \emptyset = \emptyset, \ A \cup \mathbf{1} = \mathbf{1}, \quad \text{and} \quad A \cap \mathbf{1} = A,$$

and that there exists a mapping $A \to A^c$ of $\mathcal{B}$ into $\mathcal{B}$ such that

$$A \cup A^c = \mathbf{1}, \quad A \cap A^c = \emptyset, \quad \text{and} \quad (A^c)^c = A.$$

Using the commutativity and associativity of $\cup$ and $\cap$ and the distributivity of each of these relations with respect to the other, it is easy to verify that

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

Finally, $\mathbf{1}^c = \emptyset$ and $\emptyset^c = \mathbf{1}$.

Associated order relation

Given a Boolean algebra $\mathcal{B}$ and $A, B \in \mathcal{B}$, we say that A *implies* B, and write $A \leq B$, if $A \cap B = A$.

It is easily verified that $\leq$ is an order relation on $\mathcal{B}$. With respect to this ordering, $\mathbf{1}$ is the largest element and $\emptyset$ the smallest element; that is, for any $A \in \mathcal{B}$, $\emptyset \leq A \leq \mathbf{1}$.

Using the commutativity of $\cup$ and $\cap$, we note that $A \cup B$ and $A \cap B$ are, respectively, an upper and a lower bound of A and B. In fact, $A \cup B$ is the *least upper bound of A and B* and $A \cap B$ is the *greatest lower bound of A and B*. Let us show this, for example, for $A \cup B$. Let C be an element of $\mathcal{B}$ such that $A \leq C$ and $B \leq C$; then, by definition of the order relation,

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) = A \cup B, \quad \text{and} \quad A \cup B \leq C.$$

1.3 Representation of a Boolean algebra

How to pass from the point of view of essences to that of phenomena is clear.

If Ω is the space of states of the physical system being studied, we associate with an event A of this system the following subset of Ω:

$$A' = \{\omega \in \Omega : \text{the event } A \text{ is satisfied by } \omega\}.$$

The operations of the logical calculus correspond to taking unions and intersections in the set $\mathcal{P}(\Omega)$ of subsets of Ω. With these two operations, $\mathcal{P}(\Omega)$ is a Boolean algebra. The following statement summarizes our observations.

1.3.1 The data of a phenomenological representation of a physical system of which the space of states Ω is known are equivalent to the data of a Boolean subalgebra of $\mathcal{P}(\Omega)$.

The converse, that every abstract Boolean algebra can be represented as a subalgebra of $\mathcal{P}(\Omega)$, is proved in the following fundamental theorem.

1.3.2 Stone's theorem. *Let $\mathcal{B}$ be an abstract Boolean algebra. Then there exist a compact space Ω and a representation identifying $\mathcal{B}$ with a Boolean subalgebra of $\mathcal{P}(\Omega)$ of subsets that are both open and closed in Ω.*

PROOF. The proof of Stone's theorem is clear when $\operatorname{card}(\mathcal{B}) < \infty$. In this case, we define *atomic events* as those that are minimal in $\mathcal{B}$ with respect to the relation $\leq$; then Ω is the set of atomic events.

In the general case, we introduce the notion of a filter on $\mathcal{B}$. A *filter* $\mathcal{F}$ is a nonempty subset of $\mathcal{B}$ such that

$$A_1,\ A_2 \in \mathcal{F} \quad \text{implies} \quad A_1 \cap A_2 \in \mathcal{F};$$
$$A_1 \in \mathcal{F},\ A_1 \leq A_2 \quad \text{implies} \quad A_2 \in \mathcal{F};$$

and

$$\emptyset \notin \mathcal{F}.$$

The inclusion relation on the set of subsets of $\mathcal{B}$ defines an order relation on the set of filters:

$$\mathcal{F}_1 \geq \mathcal{F}_2 \quad \text{if} \quad A \in \mathcal{F}_2 \text{ implies } A \in \mathcal{F}_1. \quad (\mathcal{F}_1 \text{ is then called } \textit{finer} \text{ than } \mathcal{F}_2).$$

An *ultrafilter* is a filter $\mathcal{U}$ of $\mathcal{B}$ such that $\mathcal{F} = \mathcal{U}$ for every filter $\mathcal{F}$ such that $\mathcal{F} \geq \mathcal{U}$. Zorn's lemma shows that, given a filter $\mathcal{F}_0$, there always exists an ultrafilter $\mathcal{U}$ finer than $\mathcal{F}_0$.[1]

1.3.3 Lemma. *Let $\mathcal{F}$ be a filter on $\mathcal{B}$ and let $A_0 \in \mathcal{B}$. Suppose that $A \cap A_0 \neq \emptyset$ for any $A \in \mathcal{F}$. Set*

$$\mathcal{F}_{A_0} = \{Z \in \mathcal{B} : Z \textit{ contains a set of the form } A \cap A_0 \textit{ with } A \in \mathcal{F}\}.$$

Then $\mathcal{F}_{A_0}$ is a filter.

PROOF. Clear.

1.3.4 Lemma. *A necessary and sufficient condition that a filter $\mathcal{U}$ be an ultrafilter on $\mathcal{B}$ is that, for any $A_0 \in \mathcal{B}$, either $A_0 \in \mathcal{U}$ or $A_0^c \in \mathcal{U}$.*

PROOF. Suppose that $\mathcal{U}$ is an ultrafilter. If $A_0 \notin \mathcal{U}$, then it is impossible that $A \cap A_0 \neq \emptyset$ for every $A \in \mathcal{U}$. Otherwise 1.3.3 would imply that $\mathcal{U}_{A_0}$ is an ultrafilter, necessarily finer than $\mathcal{U}$ since $A_0 \in \mathcal{U}_{A_0}$; but this is a contradiction. Hence, if $A_0 \notin \mathcal{U}$ and $A_0^c \notin \mathcal{U}$, there must exist X, $Y \in \mathcal{U}$ such that

$$A_0 \cap X = \emptyset \quad \text{and} \quad A_0^c \cap Y = \emptyset.$$

From this it would follow that $X \cap Y = \emptyset$, a contradiction.

Conversely, let $\mathcal{F}$ be a finer filter than $\mathcal{U}$. Let $A_0 \in \mathcal{F}$. It is impossible that $A_0^c \in \mathcal{U}$, since this would imply $A_0^c \in \mathcal{F}$, a contradiction. Hence $A_0 \in \mathcal{U}$ and $\mathcal{F} = \mathcal{U}$. $\square$

---

[1] See Bourbaki, *General Topology*, I.6.4.

1.3.5 PROOF OF STONE'S THEOREM. Let Ω be the space of ultrafilters on the Boolean algebra $\mathcal{B}$.

Let a mapping φ from $\mathcal{B}$ into $\mathcal{P}(\Omega)$ be defined by setting

$$\varphi(A) = \{\mathcal{U} \in \Omega : A \in \mathcal{U}\}, \quad A \in \mathcal{B}.$$

If $A_1 \geq A_2$ and $A_2 \in \mathcal{U}$, then $A_1 \in \mathcal{U}$; hence φ is compatible with the order relations, and is thus a Boolean algebra homomorphism. Let us show that φ is injective. Suppose that $A \neq B$; then either $A \cap B^c \neq \emptyset$ or $A^c \cap B \neq \emptyset$. Suppose, for example, that $A \cap B^c \neq \emptyset$, and consider the filter

$$\mathcal{F} = \{X \in \mathcal{B} : X \geq A \cap B^c\}.$$

Let $\mathcal{U}$ be a finer ultrafilter than $\mathcal{F}$. Then $\mathcal{U} \in \varphi(A)$ and $\mathcal{U} \notin \varphi(B)$.

To endow Ω with a topology, consider $\Omega_1 = 2^{\mathcal{B}}$, the product of infinitely many sets of two elements with the factors indexed by the set $\mathcal{B}$. Then Ω_1 is the product of compact spaces and hence is compact. Let a mapping $\Phi : \Omega \to \Omega_1$ be defined by setting

$$\Phi(\mathcal{U}) = \{\mathbf{1}_{\mathcal{U}}(A)\}_{A \in \mathcal{B}},$$

where $\mathbf{1}_{\mathcal{U}}(A) = 1$ if $A \in \mathcal{U}$ and is zero otherwise. Φ is clearly injective; thus Ω can be identified with a subset of Ω_1. We now prove that

1.3.6 $\Phi(\Omega)$ *is a closed subset of* Ω_1.

PROOF OF 1.3.6. Let Ω_1 be identified with the set of functions f defined on $\mathcal{B}$ and with values in $\{0, 1\}$. We will need the following lemma.

1.3.7 Lemma. $f \in \Phi(\Omega)$ if and only if the following conditions are satisfied for any A, A', A'', $A''' \in \mathcal{B}$:

$$\begin{aligned} f(\emptyset) &= 0, \\ f(A) &\leq f(A') \quad \text{if} \quad A \leq A,' \\ f(A'' \cap A''') &= \inf(f(A''), f(A''')), \\ f(A) + f(A^c) &= 1. \end{aligned}$$

PROOF. The first three conditions simply restate the fact that $\Phi(\mathcal{U})$ is a filter, and the fourth that $\mathcal{U}$ is an ultrafilter.□

Now let

$$L_A = \{f \in \Omega_1 : f(A) + f(A^c) = 1\}.$$

Then L_A is a *closed subset* of Ω_1, and $\cap_{A \in \mathcal{B}} L_A$ is a closed subset of Ω_1. Proceeding similarly with the other conditions of 1.3.7 completes the proof of 1.3.6. □

With the topology induced by Ω_1, $\Phi(\Omega)$ is compact; pulling back this topology makes Ω a compact space.

Fix $A_0 \in \mathcal{B}$ and define $f_0(\mathcal{U}) = \mathbf{1}_{\mathcal{U}}(A_0)$. Then

$$\varphi(A_0) = \{\mathcal{U} \in \Omega : f_0(\mathcal{U}) = 1\}.$$

Since f_0 is continuous, $\varphi(A_0)$ is a closed subset of Ω. But $(\varphi(A_0))^c = \varphi(A_0^c)$ is also closed, so $\varphi(A_0)$ is an open and closed subset of Ω. This completes the proof of Stone's theorem. □

1.4 Probability spaces

1.4.1 Definitions

A *probability space* is a measure space $(X, \mathcal{A}, \mu)$ for which the measure μ has total mass 1: $\mu(X) = 1$.

Following the usual practice in this field, we denote X by Ω and μ by P. Thus a probability space is written in the form $(\Omega, \mathcal{A}, P)$.

A measurable set $A \in \mathcal{A}$ is sometimes called an *event*. The measure of the measurable set A is called the probability of A and written $P(A)$. Clearly $0 \leq P(A) \leq 1$.

P is called the *probability measure.*

A property that is true a.e. on Ω is called an *almost sure* (or a.s.) property.

1.4.2 Transporting a probability measure

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $(Y, \mathcal{B})$ be a measurable space. Let Φ be a measurable mapping from Ω to Y:

$$\Phi \in \mathcal{M}((\Omega, \mathcal{A}); (Y, \mathcal{B})).$$

Then a probability measure P_1 is defined on $(Y, \mathcal{B})$ by setting

1.4.3
$$P_1(B) = P(\Phi^{-1}(B)).$$

Axioms I-1.0.1 to 1.0.3 are easily verified. Moreover, $P_1(Y) = P(\Omega) = 1$. P_1 is written

1.4.4
$$P_1 = \Phi_*(P)$$

and called the *direct image*, or simply the image, of the probability measure P under the mapping Φ. ($\Phi_* P$ is sometimes called the measure induced by Φ on Y.)

1.4.5 Proposition. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, let $(Y, \mathcal{B})$ be a measurable space, and let Φ, $\Phi' \in \mathcal{M}((\Omega, \mathcal{A})); (Y, \mathcal{B}))$. If $\Phi(\omega) = \Phi'(\omega)$ a.s., then $\Phi_* P = \Phi'_* P$.*

PROOF. Let $A_0 = \{\omega \in \Omega : \Phi(\omega) \neq \Phi'(\omega)\}$.

Then $P(A_0) = 0$ and $P(A) = P(A \cap A_0^c)\ \ \forall A \in \mathcal{A}$.

In particular, $P(\Phi^{-1}(B)) = P(\Phi^{-1}(B) \cap A_0^c)$ for any $B \in \mathcal{B}$. If $\omega \in \Phi^{-1}(B) \cap A_0$, then $\Phi'(\omega) = \Phi(\omega) \in B$, whence $\Phi^{-1}(B) \cap A_0^c \subset (\Phi')^{-1}(B)$, or

$$P(\Phi^{-1}(B)) \leq P((\Phi')^{-1}(B)).$$

Since the argument is symmetric in Φ and Φ', the opposite inequality also holds. □

1.4.6 Corollary. *The direct image $\Phi_* P$ depends only on the equivalence class of Φ in $M_P((\Omega, \mathcal{A}); (Y, \mathcal{B}))$.*

1.4.7 REMARK. In Chapter I, we never found it necessary to change the measure space, which was fixed once and for all. In probability theory, however, two operations will play a fundamental role:

(i) *transporting a probability by a measurable mapping; and*
(ii) *restricting a probability to a sub-σ-algebra .*

1.5 Morphisms of probability spaces

1.5.1 **Definition.** Let $(\Omega, \mathcal{A}, P)$ and $(\Omega', \mathcal{A}', P')$ be probability spaces and let

$$\Phi \in M_P((\Omega, \mathcal{A}); (\Omega', \mathcal{A}')).$$

If $\Phi_* P = P'$, Φ is called a *morphism of probability spaces* and is said to *preserve* probabilities.

1.5.2 The inverse image operation

Let $\Phi \in \mathcal{M}((\Omega, \mathcal{A}); (\Omega', \mathcal{A}'))$ and let $(Y, \mathcal{B})$ be a measurable space. With

$$u' \in \mathcal{M}((\Omega', \mathcal{A}'); (Y, \mathcal{B}))$$

we associate $\Phi^* u'$, its *inverse image* under Φ, defined by

$$(\Phi^* u')(\omega) = (u' \circ \Phi)(\omega).$$

Then

$$(\Phi^* u') \in \mathcal{M}((\Omega, \mathcal{A}); (Y, \mathcal{B})).$$

($\Phi^* u'$ is sometimes called the pullback of u'.)

If we also assume that $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ are equipped with probability measures P and P' and that Φ is a morphism of probability spaces, then

(i) *The equivalence class of* $(\Phi^* u')$ *in* $M_p((\Omega, \mathcal{A}); (Y, \mathcal{B}))$ *depends only on the class of* u' *in* $M_{P'}((\Omega', \mathcal{A}'); (Y, \mathcal{B}))$.

Let u', $u'_1 \in \mathcal{M}((\Omega', \mathcal{A}'); (Y, \mathcal{B}))$ and set

(ii) $A = \{\omega : (\Phi^* u')(\omega) \neq (\Phi^* u'_1)(\omega)\}$ and $A' = \{\omega' : u'(\omega') \neq u'_1(\omega')\}$. Then $A = \Phi^{-1}(A')$.

$P(A) = P(A') = 0$ since $P' = \Phi_* P$.

By abuse of language, Φ^* will denote the inverse image mapping induced by Φ between the spaces M_P and $M_{P'}$.

(iii) *Let* Φ, $\Phi_1 \in \mathcal{M}((\Omega, \mathcal{A}); (\Omega', \mathcal{A}'))$ *and suppose that* $\Phi = \Phi_1$ *a.s. Then* Φ^* *and* Φ_1^* *define the same mapping from* $M_{P'}((\Omega', \mathcal{A}'); (Y, \mathcal{B}))$ *to* $M_P((\Omega, \mathcal{A});$ $(Y, \mathcal{B}))$.

If not, there would exist $u' \in M_{P'}$ such that

$$A = \{\omega : \Phi^* u' \neq \Phi_1^* u'\} \quad \text{and} \quad P(A) > 0.$$

Let

$$A_1 = \{\omega \in \Omega : \Phi(\omega) \neq \Phi_1(\omega)\}.$$

Then $A \subset A_1$ and $P(A_1) = 0$. But this implies that $P(A) = 0$, a contradiction. □

(iv) *Functoriality. Let* $\Phi_3 = \Phi_2 \circ \Phi_1$. *Then* $(\Phi_3)_* = (\Phi_2)_* \circ (\Phi_1)_*$ *and* $\Phi_3^* = \Phi_1^* \circ \Phi_2^*$.

The proof is trivial. It suffices to recall that the composition of inverse images occurs in the opposite order to that of mappings.

1.5.3 Injectivity proposition. *Let Φ be a morphism of the probability space $(\Omega, \mathcal{A}, P)$ into $(\Omega', \mathcal{A}', P')$ and let $(Y, \mathcal{B})$ be an arbitrary measure space. Then Φ^* defines an injective mapping*

$$M_{P'}((\Omega', \mathcal{A}'); (Y, \mathcal{B})) \to M_P((\Omega, \mathcal{A}); (Y, \mathcal{B})).$$

PROOF. Let u', $u_1' \in M_{P'}((\Omega, \mathcal{A}'); (Y, \mathcal{B}))$. Define $u = \Phi^* u'$, $u_1 = \Phi^* u_1'$,

$$A = \{\omega : u \neq u_1\}, \quad \text{and} \quad A' = \{\omega' : u' \neq u_1'\}.$$

Then $\Phi^{-1}(A') = \Phi(A)$ by 1.5.2(ii), whence $P'(A') > 0 \Rightarrow P(A) > 0$. □

1.5.4 Dynkin's theorem (Measurability and functional dependence). *Let $(\Omega, \mathcal{A}, P)$ and $(\Omega', \mathcal{A}', P')$ be two probability spaces, let Φ be a morphism from the first to the second, and let $\mathcal{B} = \Phi^{-1}(\mathcal{A}')$. Then $u \in L_P^0(\Omega, \mathcal{A})$ can be written in the form*

$$(i) \qquad u = u' \circ \Phi, \quad \textit{with} \quad u' \in L_{P'}^0(\Omega', \mathcal{A}')$$

if and only if the class of u contains a $\mathcal{B}$-measurable function.

PROOF. The forward implication is clear. Conversely, suppose that u is $\mathcal{B}$-measurable. Then, by I-6.4.2, there exists a sequence $\{f_n\}$ of simple $\mathcal{B}$-measurable functions that converges pointwise to u. If $B \in \mathcal{B}$, then there exists $A' \in \mathcal{A}'$ such that $B = \Phi^{-1}(A')$; hence $\mathbf{1}_B = \Phi^* \mathbf{1}_A$.

This implies that every simple $\mathcal{B}$-measurable function satisfies (i). Hence $f_n = u_n' \circ \Phi$, with $u_n' \in L_{P'}^0(\Omega', \mathcal{A}')$.

(ii) We show that u_n' converges a.s. on Ω'.

If not, there would exist $\epsilon > 0$ and $A' \in \mathcal{A}'$, with $P'(A') > 0$, such that

$$\sup_{m,n>p} |u_n'(\omega) - u_m'(\omega)| > \epsilon, \quad \forall p \; \forall \omega \in A'.$$

Then u_n would satisfy the same inequality on $\Phi^{-1}(A')$. But this would contradict the a.s. convergence of f_n, since $P(\Phi^{-1}(A')) = P'(A') > 0$.

Thus (ii) is proved. Let $u' = \lim u'_n \in L^0_{P'}(\Omega', \mathcal{A}')$; then $u = \lim f_n = u' \circ \Phi$. □

1.5.5 Corollary. *Let Φ be a probability space morphism from $(\Omega, \mathcal{A}, P)$ to $(\Omega', \mathcal{A}', P')$ and let $\mathcal{B} = \Phi^{-1}(\mathcal{A}')$. Using Φ^*, one can identify $L^0_{P'}(\Omega', \mathcal{A}')$ with the subalgebra of $L^0_P(\Omega, \mathcal{A})$ consisting of the $\mathcal{B}$-measurable functions.*

PROOF. By 1.5.3 and 1.5.4.

1.6 Random variables and distributions of random variables

1.6.1 **Definition.** Given a probability space $(\Omega, \mathcal{A}, P)$, a *random variable* X is a class of measurable functions, that is an element of $L^0_P(\Omega, \mathcal{A})$. We will often write simply r.v.

1.6.2 **Definition.** The *distribution* of the random variable X is the direct image of P under X.

Thus X_*P is a Borel measure on $\mathbf{R}$ of total mass 1. Hence, by II-3.1,

(i) $\quad$ *(X_*P) defines a Radon measure of total mass* 1.

1.6.3 **Definition.** Given a finite set $X_1, \ldots, X_k$ of r.v. defined on the probability space $(\Omega, \mathcal{A}, P)$, their *joint distribution* is the direct image of P under the mapping $\Phi : \omega \to \mathbf{R}^k$ defined by the $X_p(\omega)$, $1 \le p \le k$.

It follows from I-2.4.2 and I-2.3.5 that $\Phi \in M((\Omega, \mathcal{A}); (\mathbf{R}^k, \mathcal{B}_{\mathbf{R}_k}))$.

Hence Φ_*P is a finite Borel measure on $\mathbf{R}^k$ and, by II-3.1,

(ii) $\quad$ *Φ_*P defines a Radon measure on $\mathbf{R}^k$ of total mass* 1.

1.6.4 Let p_1 be the projection of an element of $\mathbf{R}^k$ onto its first component, let μ be the joint distribution of $X_1, \ldots, X_k$, and let μ_1 be the distribution of X_1. Then $\mu_1 = (p_1)_*\mu$.

This follows from functoriality, 1.5.2(iv).

1.7 Mathematical expectation and distributions

1.7.0 Notation for expectations

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $X \in L^1_P(\Omega, \mathcal{A})$. Then the *mathematical expectation of* X is written $\mathbf{E}(X)$ and defined by

$$\mathbf{E}(X) = \int X(\omega) dP(\omega).$$

The reader should note that the measure P, and the probability space Ω itself, are implicit in the notation $\mathbf{E}$.

In this notation, the L^q norm is written

$$[\mathbf{E}(|Y|^q)]^{1/q} = \|Y\|_{L^q}.$$

1.7.1 Change of variables

Let $(\Omega, \mathcal{A}, P)$ and $(\Omega', \mathcal{A}', P')$ be probability spaces, let Φ be a morphism from the first space to the second, and let $\Phi^* : L^0(\Omega', \mathcal{A}') \to L^0(\Omega, \mathcal{A})$ be as defined in 1.5.3.

Proposition. *Let $u' \in L^1_{P'}(\Omega', \mathcal{A}')$. Then $u = (\Phi^* u') \in L^1_P(\Omega, \mathcal{A})$ and $\mathbf{E}(u) = \mathbf{E}(u')$.*

PROOF. Suppose that u' is a simple function, say $u' = \sum \alpha_k \mathbf{1}_{A'_k}$. Then

$$u = \sum \alpha_k \mathbf{1}_{A_k}, \quad \text{where} \quad A_k = \Phi^{-1}(A'_k).$$

By 1.4.3, $P(A_k) = P'(A'_k)$, whence $\mathbf{E}(u) = \mathbf{E}(u')$.

Let $v' \in L^1_{P'}$; then there exists a sequence $\{u'_n\}$ of simple functions such that

$$\mathbf{E}(|v' - u'_n|) = \|v' - u'_n\|_{L^1_{P'}} \to 0.$$

Let $u_n = \Phi^* u'_n$. Then

$$\|u_n - u_m\|_{L^1_P} = \mathbf{E}(|u_n - u_m|) = \mathbf{E}(|u'_n - u'_m|) \to 0 \quad \text{as} \quad m, n \to +\infty.$$

Thus $\{u_n\}$ is a Cauchy sequence in L^1_P. Let v be its limit; then $v \in L^1_P$. There exists a subsequence $\{u'_n : n \in \sigma\}$ of $\{u'_n\}$ that converges a.e. on Ω'. Similarly, there exists a subsequence $\{u_n : n \in \tau\}$ of $\{u_n : n \in \sigma\}$ that converges a.e. on Ω to v. Then the relation $u_n = u'_n \circ \Phi$ passes to the limit, and $v = v' \circ \Phi$. Moreover, since

$$\mathbf{E}(v) = \lim \mathbf{E}(u_n) \quad \text{and} \quad \mathbf{E}(v') = \lim \mathbf{E}(u'_n),$$

the fact that $\mathbf{E}(u_n) = \mathbf{E}(u'_n)$ implies that

$$\mathbf{E}(v) = \mathbf{E}(v'). \square$$

1.7.2 Computing expectations by means of distributions

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $X_1, \dots, X_k$ be a finite set of r.v. defined on Ω. Let μ be the Radon measure on $\mathbf{R}^k$ that is the distribution of $X_1, \dots, X_k$.

Proposition. *Let $\varphi \in L^1_\mu$ and let $Y(\omega) = \varphi(X_1(\omega), \dots, X_k(\omega))$. Then $Y \in L^1_P$ and $\mathbf{E}(Y) = \int_{\mathbf{R}^k} \varphi \, d\mu$.*

PROOF. By 1.7.1.

1.8 Various notions of convergence in probability theory

This section consists of two subsections. In the first, we introduce the vocabulary used in probability theory to study concepts that are already familiar. In the second, we study the new concept of convergence in distribution.

1.8.1 Vocabulary of probability theory

Let $\{X_n\}$ be a sequence of r.v. defined on the probability space $(\Omega, \mathcal{A}, P)$, and let Y be another r.v. defined on the same probability space.

Definitions

(i) *X_n converges to Y almost surely (abbreviated a.s.) if $X_n(\omega)$ converges a.e. to $Y(\omega)$.*

(ii) *X_n converges to Y in mean if*

$$\|X_n - Y\|_{L^1} \to 0, \quad \text{or} \quad \mathbf{E}(|X_n - Y|) \to 0.$$

(iii) *X_n converges to Y in mean square if*

$$\|X_n - Y\|_{L^2} \to 0, \quad \text{or} \quad \mathbf{E}(|X_n - Y|^2) \to 0.$$

(iv) *X_n converges to Y in probability if X_n converges to Y in measure.*

(v) The relations among these different kinds of convergence were studied in Chapter I.

1.8.2 Convergence in distribution

Let $(\Omega_n, \mathcal{A}_n, P_n)$ be a sequence of probability spaces and let $(\Omega', \mathcal{A}', P')$ be another probability space.

Let $X_n \in L^0(\Omega_n, \mathcal{A}_n, P_n)$ and $Y \in L^0(\Omega', \mathcal{A}', P')$ be given. We say that *the sequence of distributions of X_n converges to the distribution of Y* if, writing

$$(i) \qquad (X_n)_* P_n = \mu_n \quad \text{and} \quad Y_* P' = \nu$$

for the respective distributions,

(ii) *μ_n converges narrowly to ν.*

A sequence μ_n such that

(iii) *μ_n converges narrowly*

is commonly, though rather ambiguously, described by saying that

(iv) *the r.v. X_n converge in distribution.*

1.8.3 Criterion for convergence in distribution

Theorem. *The r.v. X_n converge in distribution to the distribution of Y if and only if*

$$(i) \qquad \lim \mathbf{E}(\varphi(X_n)) = \mathbf{E}(\varphi(Y)), \quad \forall \varphi \in C_K(\mathbf{R}).$$

PROOF. By 1.7.2, in the notation of 1.8.2(i),

$$\mathbf{E}(\varphi(X_n)) = \int_{\mathbf{R}} \varphi d\mu_n \quad \text{and} \quad \mathbf{E}(\varphi(Y)) = \int_{\mathbf{R}} \varphi d\nu.$$

Thus

$$\int \varphi d\mu_n \to \int \varphi d\nu, \quad \forall \varphi \in C_K(\mathbf{R}).$$

That is,

(ii) *μ_n converges vaguely to ν.*

By 1.6.2(i),

$$\mu_n(\mathbf{R}) = 1 \quad \text{and} \quad \nu(\mathbf{R}) = 1;$$

hence $\lim \mu_n(\mathbf{R}) = \nu(\mathbf{R})$, and II-6.8 shows that (ii) is equivalent to narrow convergence. □

1.8.4 Extension to r.v. with values in $\mathbf{R}^m$

An ordered m-tuple of r.v. $X^1, \ldots, X^m$ is called an r.v. with values in $\mathbf{R}^m$, or an $\mathbf{R}^m$-valued r.v. Such a r.v. is sometimes denoted by $\vec{X} \in M_P((\Omega, \mathcal{A}); (\mathbf{R}^m, \mathcal{B}_{\mathbf{R}^m}))$.

Given a r.v. $\vec{X}$ with values in $\mathbf{R}^m$, its *distribution* is the joint distribution of the X^k considered in 1.6.3; it is thus a Radon measure on $\mathbf{R}^m$.

A sequence of r.v. with values in $\mathbf{R}^m$, say $\vec{X}_1, \ldots, \vec{X}_n, \ldots$, is said to converge in distribution to $\vec{X}_0$ if the sequence of distributions converges narrowly to that of $\vec{X}_0$. We have the following propositions.

(i) *The sequence of r.v. $\vec{X}_n$ with values in $\mathbf{R}^m$ converges to the distribution of $\vec{X}_0$ if and only if*

$$\lim \mathbf{E}(\varphi(\vec{X}_n)) = \mathbf{E}(\varphi(\vec{X}_0)), \quad \forall \varphi \in C_K(\mathbf{R}^m).$$

In this criterion, a compactly supported φ can be replaced by a bounded continuous ψ. The next statement results from letting ψ be a function that depends only on the first coordinate of $\mathbf{R}^m$ and applying 1.8.3.

(ii) *If $\vec{X}_n$ converges in distribution to $\vec{X}_0$, then each component X_n^k converges in distribution to X_0^k.*

The converse of this statement is false.

1.8.5 Comparison of convergence in distribution with other types of convergence

Proposition.

(i) A.s. convergence implies convergence in distribution.
(ii) Convergence in probability implies convergence in distribution.
(iii) Convergence in L^p implies convergence in distribution.

PROOF. Let the probability space Ω be fixed and let $X_n, Y \in L^0(\Omega, \mathcal{A}, P)$. Assume first that X_n converges a.s. to Y. Then $\forall \varphi \in C_K(\mathbf{R})$, $\varphi(X_n(\omega))$ converges a.s. to $\varphi(Y(\omega))$. Since φ is bounded, Lebesgue's dominated convergence theorem can be applied to show that $\mathbf{E}(\varphi(X_n)) \to \mathbf{E}(\varphi(Y))$. This, with 1.8.3, gives (i).

Assume now that X_n converges in probability to Y. By I-5.2.7, every subsequence $\{X_n\}_{n\in\sigma}$ itself contains a subsequence $\{X_n\}_{n\in\sigma'}$ such that $\{X_n\}_{n\in\sigma'}$ converges a.s. Hence, if $\varphi \in C_K(\mathbf{R})$, it follows from (i) that

$$\lim_{n\in\sigma'} \mathbf{E}(\varphi(X_n)) = \mathbf{E}(\varphi(Y)).$$

Let $\beta_n = \mathbf{E}(\varphi(X_n))$ and let $\gamma = \mathbf{E}(\varphi(Y))$. Then every subsequence $\{\beta_n\}_{n\in\sigma}$ of $\{\beta_n\}$ contains a subsequence $\{\beta_n\}_{n\in\sigma'}$ that converges to γ. This implies that $\lim \beta_n = \gamma$, and (ii) now follows from 1.8.3.

Finally, by I-9.3.1, convergence in L^p implies convergence in probability; thus (iii) follows from (ii). □

2 Conditional Expectation

2.0 Phenomenological meaning

We now resume the discussion of the principles of probability theory begun in 1.1.

From the phenomenological point of view, the set of all measurements an experimenter can possibly make on a physical system is represented by a Boolean algebra $\mathcal{B}$. The physicist is interested in exhibiting the "laws of nature" in the context of $\mathcal{B}$; given certain measurements, he would like to *predict* the values of others.

There are two kinds of predictions. The first involves a functional dependence. For example, in Ohm's law (that $V = RI$), the measurement of two quantities completely determines the third. The second involves a "correlation" without necessity; for example, a substantial drop in barometric pressure makes it "likely" that a cyclone is approaching.

The experimenter represents the known *information* about the physical system by a *subalgebra* $\mathcal{B}'$ of $\mathcal{B}$. Given a physical quantity X, he asks himself the following questions.

(a) Is X determined by the information $\mathcal{B}'$? That is, in terms of 1.5.4, is X measurable with respect to the σ-algebra generated by $\mathcal{B}'$?

(b) If not, the experimenter will try to extract from the information $\mathcal{B}'$ all it implies about X. What is the most likely value of X? Does he risk making a major error by taking this most likely value as the value of X? And so on.

Passing to σ-algebras generated by Boolean algebras allows the problem to be posed as follows:

Given a probability space $(\Omega, \mathcal{A}, P)$, a sub-σ-algebra $\mathcal{A}'$ of $\mathcal{A}$, and $X \in L^0(\Omega, \mathcal{A}, P)$, can X be *approximated* by $Y \in L^0(\Omega, \mathcal{A}', P)$? (We abuse language by writing P for the restriction of P to $\mathcal{A}'$.)

In the next section, we will try to solve this problem by using an approximation that minimizes the L^2 norm, i.e. an orthogonal projection on L^2.

2.1 Conditional expectation as a projection operator on L^2

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. $L^p(\Omega, \mathcal{B}, P)$ is abbreviated as $L^p(\mathcal{B})$, and so on.

2.1.1 Lemma. *Let $1 \leq p \leq +\infty$. Then $L^p(\mathcal{B})$ can be identified with a closed vector subspace of $L^p(\mathcal{A})$.*

Proof. A $\mathcal{B}$-measurable function is $\mathcal{A}$-measurable: $\mathcal{L}^0(\mathcal{B}) \subset \mathcal{L}^0(\Omega, \mathcal{A})$. The same holds for simple functions: $\mathcal{E}(\mathcal{B}) \subset \mathcal{E}(\mathcal{A})$. Since the probability measure on $\mathcal{B}$ is the restriction of that on $\mathcal{A}$, the integral on the integrable simple functions $\mathbf{E}^1(\mathcal{B})$ is given by restriction of the integral defined on $\mathbf{E}^1(\mathcal{A})$. Endowing $\mathbf{E}^1(\mathcal{B})$ with the norm $\| \;\; \|_{L^p}$, we obtain an isometric mapping from $\mathbf{E}^1(\mathcal{B})$ to $\mathbf{E}^1(\mathcal{A})$.

Since $\mathbf{E}^1(\mathcal{B})$ is dense in $L^p(\mathcal{B})$ and $L^p(\Omega, \mathcal{A}, P)$ is complete, this isometry extends to an isometry

$$L^p(\mathcal{B}) \to L^p(\mathcal{A}).$$

The image of a complete space under an isometry is complete; hence the image of $L^p(\mathcal{B})$ is complete and, in particular, closed in $L^p(\mathcal{A})$.

2.1.2 **Definition.** $\mathbf{E}^{\mathcal{B}}$ denotes the orthogonal projection operator from $L^2(\mathcal{A})$ onto $L^2(\mathcal{B})$. Given $f \in L^2(\mathcal{A})$, $\mathbf{E}^{\mathcal{B}}(f)$ is called the *conditional expectation* of f given $\mathcal{B}$.

2.1.3 Theorem (Properties of the conditional expectation).

(i) $$\mathbf{E}^{\mathcal{B}}(f) \in L^2(\mathcal{B})$$

(ii) $$\|\mathbf{E}^{\mathcal{B}}(f)\|_{L^2} \leq \|f\|_{L^2}.$$

Let $\mathcal{B}$ and $\mathcal{C}$ be sub-σ-algebras of $\mathcal{A}$ such that $\mathcal{B} \supset \mathcal{C}$. Then

(iii) $$\mathbf{E}^{\mathcal{C}}\mathbf{E}^{\mathcal{B}} = \mathbf{E}^{\mathcal{C}} \quad \textit{and}$$

(iv) $$\mathbf{E}(\mathbf{E}^{\mathcal{B}}) = \mathbf{E}.$$

(v) Let $\varphi \in L^\infty(\mathcal{B})$. Then $\mathbf{E}^{\mathcal{B}}(\varphi f) = \varphi \mathbf{E}^{\mathcal{B}}(f)$, $\forall f \in L^2(\mathcal{A})$.

PROOF. Properties (i) and (ii) follow from properties of the orthogonal projection.

The inclusion between σ-algebras $\mathcal{B} \supset \mathcal{C}$ implies, for functions, that $L^2(\mathcal{B}) \supset L^2(\mathcal{C})$.

Let $f \in L^2(\mathcal{A})$ be decomposed as

$$f = u + v, \text{ with } v \in (L^2(\mathcal{B}))^\perp \text{ and } u = \mathbf{E}^{\mathcal{B}}(f).$$

Then $u = w + h$, with $h \in (L^2(\mathcal{C}))^\perp$ and $w = \mathbf{E}^{\mathcal{C}}(u)$. Substituting this into the last line gives

$$f = w + (h + v). \tag{vi}$$

By definition, $w \in L^2(\mathcal{C})$, and since

$$L^2(\mathcal{B}) \supset L^2(\mathcal{C}) \Rightarrow (L^2(\mathcal{B}))^\perp \subset (L^2(\mathcal{C}))^\perp,$$

$v \in (L^2(\mathcal{C}))^\perp$. Hence $h + v \in (L^2(\mathcal{C}))^\perp$. The decomposition (vi) implies that $w = \mathbf{E}^{\mathcal{C}}(f)$. Thus (iii) is proved.

Let $\mathcal{A}_0$ denote the *coarse σ-algebra* containing only the two sets Ω and $\emptyset$. A function φ is $\mathcal{A}_0$-measurable if and only if it is constant. $(L^2(\mathcal{A}_0))^\perp$ consists of the functions with zero expectation. Any function $f \in L^2$ can be written as

$$f = \mathbf{E}(f)\mathbf{1}_\Omega + h, \quad \text{where} \quad \mathbf{E}(h) = 0,$$

and thus

$$\mathbf{E}^{\mathcal{A}_0}(f) = \mathbf{E}(f)\mathbf{1}_\Omega. \tag{vii}$$

By abuse of language, we identify the conditional expectation relative to $\mathcal{A}_0$ with the expectation. Then (iv) becomes a special case of (iii).

It remains to prove (v). Let M_φ denote the bounded operator defined on $L^2(\mathcal{A})$ by multiplication by φ. Thus $M_\varphi : f \mapsto \varphi f$. Since $\varphi \in L^0(\mathcal{B})$ and $L^0(\mathcal{B})$ is an algebra,

$$M_\varphi(L^2(\mathcal{B})) \subset L^2(\mathcal{B}).$$

Note that M_φ is a *hermitian operator*; that is,

$$(M_\varphi(f)|g)_{L^2} = (f|M_\varphi(g))_{L^2}.$$

This is just a restatement of the fact that

$$\mathbf{E}((\varphi f)g) = \mathbf{E}(f(\varphi g)), \quad \forall f, g \in L^2(\mathcal{A}).$$

Since $L^2(\mathcal{B})$ is invariant under the hermitian operator M_φ, *its orthogonal complement is invariant under* M_φ. Thus, if $f = u+v$ with $u \in L^2(\mathcal{B})$ and $v \in (L^2(\mathcal{B}))^\perp$, then

$$M_\varphi f = M_\varphi u + M_\varphi v, \text{ where } (M_\varphi u) \in L^2(\mathcal{B}),\ (M_\varphi v) \in L^2(\mathcal{B})^\perp.$$

That is, $\mathbf{E}^{\mathcal{B}}(M_\varphi f) = M_\varphi \mathbf{E}^{\mathcal{B}}(f)$, and (v) is proved. □

2.2 Conditional expectation and positivity

2.2.1 Proposition. *Let $f \in L^2(\mathcal{A})$, $f \geq 0$, and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Then $\mathbf{E}^{\mathcal{B}}(f) \geq 0$.*

PROOF. Let $B \in \mathcal{B}$. Then, by (v),

$$\mathbf{E}(\mathbf{E}^{\mathcal{B}}(f)\mathbf{1}_B) = \mathbf{E}(\mathbf{E}^{\mathcal{B}}(f\mathbf{1}_B)) = \mathbf{E}(f\mathbf{1}_B) \geq 0,$$

where the second equality follows from (iv). Setting $v = \mathbf{E}^{\mathcal{B}} f$, we have just shown that

$$(i) \qquad \mathbf{E}(v\mathbf{1}_B) \geq 0, \quad \forall B \in \mathcal{B}.$$

Let $B_n = \{\omega : v(\omega) < -n^{-1}\}$. Since $v \in L^0(\mathcal{B})$, $B_n \in \mathcal{B}$; it follows from (i) that

$$\mathbf{E}(v\mathbf{1}_{B_n}) \geq 0.$$

Moreover, $\mathbf{E}(v\mathbf{1}_{B_n}) \leq -n^{-1}P(B_n)$. Hence $P(B_n) = 0$ for all n, and thus $P(\cup B_n) = \lim P(B_n) = 0$. □

2.2.2 Corollary. *Let $f,\ g \in L^2(\mathcal{A})$. Then*

$$(i) \qquad f \geq g \Rightarrow \mathbf{E}^{\mathcal{B}}(f) \geq \mathbf{E}^{\mathcal{B}}(g)$$

and

$$(ii) \qquad |\mathbf{E}^{\mathcal{B}}(f)| \leq \mathbf{E}^{\mathcal{B}}(|f|).$$

PROOF. Since $f - g \geq 0$, we have $\mathbf{E}^{\mathcal{B}}(f-g) \geq 0$. Furthermore, $-|f| \leq f \leq |f|$ implies (ii). □

2.3 Extension of conditional expectation to L^1

Theorem. *The operator $\mathbf{E}^{\mathcal{B}}$ defined on $L^2(\mathcal{A})$ in 2.1 has a continuous extension $\mathcal{E}^{\mathcal{B}}$, defined on $L^1(\mathcal{A})$ and with values in $L^1(\mathcal{B})$. This extension has the following properties:*

(i) $\mathcal{E}^{\mathcal{B}}(f) = f$ *for every* $f \in L^1(\mathcal{B})$.

(ii) $\|\mathcal{E}^{\mathcal{B}}(f)\|_{L^1} \leq \|f\|_{L^1}$.
(iii) If $\mathcal{B} \supset \mathcal{C}$, *then* $\mathcal{E}^{\mathcal{C}}\mathcal{E}^{\mathcal{B}} = \mathcal{E}^{\mathcal{C}}$; *in particular,* $\mathbf{E}\mathcal{E}^{\mathcal{B}} = \mathbf{E}$.
(iv) If $\varphi \in L^\infty(\mathcal{B})$, *then* $\mathcal{E}^{\mathcal{B}}(\varphi f) = \varphi\mathcal{E}^{\mathcal{B}}(f)$.

PROOF. Let $f \in L^2$. Then $|\mathbf{E}^{\mathcal{B}} f| \leq \mathbf{E}^{\mathcal{B}}(|f|)$ by 2.2.2(ii), and hence $\mathbf{E}(|\mathbf{E}^{\mathcal{B}} f|) \leq \mathbf{E}(\mathbf{E}^{\mathcal{B}}(|f|))$. It follows from 2.1.3(iv) that

$$\mathbf{E}(\mathbf{E}^{\mathcal{B}}(|f|)) = \mathbf{E}(|f|) = \|f\|_{L^1}.$$

That is,

$$(v) \qquad \|\mathbf{E}^{\mathcal{B}} f\|_{L^1} \leq \|f\|_{L^1}, \quad \forall f \in L^2(\mathcal{A}).$$

Thus $\mathbf{E}^{\mathcal{B}}$ is a bounded operator when $L^2(\mathcal{A})$ is equipped with the L^1 norm. Since $L^1(\mathcal{B})$ is complete and $L^2(\mathcal{A})$ is *dense* in $L^1(\mathcal{A})$, $\mathbf{E}^{\mathcal{B}}$ can be extended to an operator from $L^1(\mathcal{A})$ to $L^1(\mathcal{B})$. This extension is denoted by $\mathcal{E}^{\mathcal{B}}$.

Since $\mathbf{E}^{\mathcal{B}}(f) = f$ if $f \in L^2(\mathcal{B})$ and since $L^2(\mathcal{B})$ is dense in $L^1(\mathcal{B})$, the operator extended by continuity has the same property; this implies (i). Assertion (ii) follows from (v).

(iii) and (iv) are obtained from 2.1.3(iii), (iv) and (v), which we extend by continuity. □

(vi) ABUSE OF LANGUAGE. From now on we use the same notation, namely $\mathbf{E}^{\mathcal{B}}$, for both $\mathcal{E}^{\mathcal{B}}$ and $\mathbf{E}^{\mathcal{B}}$.

2.4 Calculating $\mathbf{E}^{\mathcal{B}}$ *when* $\mathcal{B}$ *is a finite* σ*-algebra*

Let $\mathcal{B}$ be a finite sub-σ-algebra of $\mathcal{A}$ and let $e_1, \ldots, e_n$ be the atoms of $\mathcal{B}$ with strictly positive probability.

2.4.1 Proposition. $\mathbf{E}^{\mathcal{B}}(f) = \sum \alpha_k \mathbf{1}_{e_k}$, *where* $\alpha_k = \frac{1}{P(e_k)}\mathbf{E}(f\mathbf{1}_{e_k})$.

PROOF. Since $\mathbf{E}^{\mathcal{B}}(f) \in L^0(\mathcal{B})$, it suffices to check that $f - \mathbf{E}^{\mathcal{B}}(f)$ is orthogonal to $L^0(\mathcal{B})$. Since the $\mathbf{1}_{e_s}$ form a basis for $L^0(\mathcal{B})$, it suffices to show that

$$\mathbf{E}((f - \mathbf{E}^{\mathcal{B}}(f))\mathbf{1}_{e_s}) = 0.$$

But

$$\mathbf{E}(f\mathbf{1}_{e_s}) = \alpha_s\mathbf{E}(\mathbf{1}_{e_s}) = \alpha_s(P(e_s)).\square$$

2.4.2 **Definition.** Let a measure μ_k be defined on $\mathcal{A}$ by setting

$$\mu_k(A) = \frac{1}{P(e_k)}P(A \cap e_k).$$

Note that $\mu_k(\Omega) = 1$.

μ_k is called the *conditional probability given the atom* e_k. With the notation of 2.4.1,

$$\alpha_k = \int f d\mu_k.$$

2.4.3 Proposition. *Let $\mathcal{B}$ be a finite σ-algebra of $\mathcal{A}$, let φ be a convex function, and let $f \in L^1(\mathcal{A})$, $f \geq 0$. Then*

$$\varphi(\mathbf{E}^{\mathcal{B}}(f)) \leq \mathbf{E}^{\mathcal{B}}(\varphi(f)).$$

PROOF. Retaining the notation of 2.4.1 and letting μ_k denote the conditional probabilities, we have

$$\begin{aligned} \mathbf{E}^{\mathcal{B}}(\varphi(f)) &= \sum \beta_k \mathbf{1}_{e_k}, \quad \text{where} \quad \beta_k = \int \varphi(f) d\mu_k, \quad \text{and} \\ \varphi(\mathbf{E}^{\mathcal{B}}(f)) &= \sum \varphi(\alpha_k) \mathbf{1}_{e_k}, \quad \text{where} \quad \alpha_k = \int f d\mu_k. \end{aligned}$$

Since μ_k has total mass 1, Jensen's inequality (I-9.2.2) can be applied, and shows that $\varphi(\alpha_k) \leq \beta_k$. □

2.5 Approximation by finite σ-algebras

2.5.1 Proposition. *Let $f_1, \ldots, f_n \in L^1(\mathcal{A})$. Then there exists an increasing sequence $\mathcal{B}_1 \subset \ldots \mathcal{B}_k \subset \ldots \subset \mathcal{B}$ of finite σ-algebras such that*

$$\|\mathbf{E}^{\mathcal{B}_k} f_j - \mathbf{E}^{\mathcal{B}} f_j\|_{L^1} \to 0 \quad \text{as } k \to \infty, \quad j = 1, 2, \ldots, n.$$

PROOF. We first consider the case where $n = 1$, and write f for f_1. Let $u = \mathbf{E}^{\mathcal{B}}(f)$; then $u \in L^1(\mathcal{B})$, and hence u is the limit in L^1 of a sequence $\{u_k\}$ of simple functions in $L^1(\mathcal{B})$. Let $\mathcal{B}_k$ be the σ-algebra generated by the u_s, $s \leq k$; then $\mathcal{B}_k \subset \mathcal{B}$ and u_k is $\mathcal{B}_k$-measurable. $\mathbf{E}^{\mathcal{B}_k}(u_k) = u_k$ and $\|\mathbf{E}^{\mathcal{B}_k}(u_k - u)\| \leq \|u_k - u\|_{L^1}$, whence

$$\begin{aligned} \|\mathbf{E}^{\mathcal{B}_k} u - u\|_{L^1} &\leq \|\mathbf{E}^{\mathcal{B}_k}(u) - \mathbf{E}^{\mathcal{B}_k}(u_k)\|_{L^1} + \|\mathbf{E}^{\mathcal{B}_k}(u_k) - u\|_{L^1} \\ &\leq \|\mathbf{E}^{\mathcal{B}_k}(u - u_k)\|_{L^1} + \|u_k - u\|_{L^1} \leq 2\|u - u_k\|_{L^1}. \end{aligned}$$

Moreover, $\mathbf{E}^{\mathcal{B}_k} u = \mathbf{E}^{\mathcal{B}_k}(\mathbf{E}^{\mathcal{B}} f) = \mathbf{E}^{\mathcal{B}_k}(f)$ by 2.3(iii). Thus

$$\|\mathbf{E}^{\mathcal{B}_k} f - \mathbf{E}^{\mathcal{B}} f\|_{L^1} \leq 2\|u - f\| \to 0.$$

This ends the proof for $n = 1$.

The general case is treated by induction on n. Letting $\{\mathcal{B}'_k\}$ be a sequence of finite σ-algebras adapted to $f_2, \ldots, f_n$, we take $\mathcal{B}''_k$ to be the σ-algebra generated by $\mathcal{B}'_k$ and $\mathcal{B}_k$. □

2.5.2 Corollary (Jensen's inequality). *Let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$, let $f \in L^1(\mathcal{A})$, $f \geq 0$, and let φ be a nonnegative convex function such that* $\mathbf{E}([\varphi(f)]) < +\infty$. *Then*

$$\varphi(\mathbf{E}^{\mathcal{B}}(f)) \leq \mathbf{E}^{\mathcal{B}}(\varphi(f)).$$

PROOF. By 2.5.1, there exists a sequence of finite σ-algebras $\mathcal{B}_k$ such that

$$\|\mathbf{E}^{\mathcal{B}_k} f - f\|_{L^1} \to 0 \quad \text{and} \quad \|\mathbf{E}^{\mathcal{B}_k}(\varphi(f)) - \varphi(f)\|_{L^1} \to 0.$$

By 2.4.3, $\varphi(\mathbf{E}^{\mathcal{B}_k}(f)) \leq \mathbf{E}^{\mathcal{B}_k}(\varphi(f))$, or $\mathbf{E}^{\mathcal{B}_k}(\varphi(f)) - \varphi(\mathbf{E}^{\mathcal{B}_k}(f)) \geq 0$. Since L^1 convergence preserves positivity,

$$\mathbf{E}^{\mathcal{B}}(\varphi(f)) - \varphi(\mathbf{E}^{\mathcal{B}}(f)) \geq 0.\square$$

2.6 Conditional expectation and L^p spaces

Let $1 < p \leq +\infty$. Then, since $L^p(\mathcal{A}) \subset L^1(\mathcal{A})$, the conditional expectation operator $\mathbf{E}^{\mathcal{B}}$ is defined on $L^p(\mathcal{A})$.

2.6.1 Proposition. *Let $1 \leq p \leq +\infty$. If $f \in L^p(\mathcal{A})$, then $\mathbf{E}^{\mathcal{B}}(f) \in L^p(\mathcal{B})$ and*

$$(i) \qquad \|\mathbf{E}^{\mathcal{B}}(f)\|_{L^p} \leq \|f\|_{L^p}.$$

Let p and q be conjugate exponents. Then

$$(ii) \qquad \mathbf{E}^{\mathcal{B}}(fg) = g\mathbf{E}^{\mathcal{B}}(f), \quad \forall f \in L^p(\mathcal{A}),\ g \in L^q(\mathcal{B}).$$

(iii)

$$\mathbf{E}((\mathbf{E}^{\mathcal{B}}g)(\mathbf{E}^{\mathcal{B}}f)) = \mathbf{E}(g\mathbf{E}^{\mathcal{B}}(f)) = \mathbf{E}(f\mathbf{E}^{\mathcal{B}}(g)), \quad \forall f \in L^p(\mathcal{A}),\ g \in L^q(\mathcal{A}).$$

PROOF. If $1 \leq p < +\infty$, the function $\varphi(t) = t^p$, $t \geq 0$, is convex. Hence (i) follows (except when $p = \infty$) from 2.5.2 (Jensen's inequality).

It remains to prove (i) if $p = \infty$. Given $f \in L^\infty$, we can find a sequence $\mathcal{B}_k$ of finite sub-σ-algebras such that

$$\|\mathbf{E}^{\mathcal{B}_k} f - \mathbf{E}^{\mathcal{B}} f\|_{L^1} \to 0.$$

Using the expressions given in 2.4.1 and 2.4.2,

$$\|\mathbf{E}^{\mathcal{B}_k} f\|_{L^\infty} = \sup \left| \int f d\mu_k \right| \leq \|f\|_{L^\infty}.$$

Let v_{k_s} be a subsequence of $v_k = \mathbf{E}^{\mathcal{B}_k} f$ such that $v_{k_s} \to \mathbf{E}^{\mathcal{B}} f$ a.s. Then, since

$$|v_{k_s}(\omega)| \leq \|f\|_{L^\infty},$$

(i) holds for $p = \infty$.

Note that (ii) holds for bounded functions by 2.3(iv). Using the truncation operator (I-6.7), we can find sequences $f_n \in L^\infty(\mathcal{A})$ and $g_n \in L^\infty(\mathcal{B})$ such that $\|f_n - f\|_{L^p} \to 0$ (whence, by (i), $\|\mathbf{E}^{\mathcal{B}} f_n - \mathbf{E}^{\mathcal{B}} f\|_{L^p} \to 0$) and $\|g_n - g\|_{L^q} \to 0$.

Hence, by 2.3(iv), $\mathbf{E}^{\mathcal{B}}(f_n g_n) = g_n \mathbf{E}^{\mathcal{B}}(f_n)$.

Since $\|f_n g_n - fg\|_{L^1} \to 0$ by Hölder's inequality, $\|\mathbf{E}^{\mathcal{B}}(f_n g_n) - \mathbf{E}^{\mathcal{B}}(fg)\|_{L^1} \to 0$. Similarly, $g_n \mathbf{E}^{\mathcal{B}} f_n$ converges to $g\mathbf{E}^{\mathcal{B}} f$ in L^1, and (ii) follows.

When f, $g \in L^2(\mathcal{A})$, we consider the scalar product

$$(f|g)_{L^2} = \mathbf{E}(fg).$$

By the properties of the orthogonal projection,

$$(\mathbf{E}^{\mathcal{B}} f|g) = (f|\mathbf{E}^{\mathcal{B}} g) = (\mathbf{E}^{\mathcal{B}} f|\mathbf{E}^{\mathcal{B}} g).$$

Since $L^\infty \subset L^2$, this proves (iii) for the special case where f, $g \in L^\infty(\mathcal{A})$. The general case is proved by using the truncation operator as above. □

3 Independence and Orthogonality

3.0 Independence of two sub-σ-algebras

3.0.1 **Definition.** Let $\mathcal{B}$ and $\mathcal{C}$ be two sub-σ-algebras of the probability space $(\Omega, \mathcal{A}, P)$. $\mathcal{B}$ and $\mathcal{C}$ are said to be *independent* (relative to P) if $L^2(\mathcal{B})$ and $L^2(\mathcal{C})$ are orthogonal on the constant functions; that is, if

$$f \in L^2(\mathcal{B}),\ g \in L^2(\mathcal{C}), \text{ and } \mathbf{E}(f) = \mathbf{E}(g) = 0 \text{ imply } \mathbf{E}(fg) = 0.$$

REMARKS.

(i) The notion of *independence* involves the L^2 norm, and thus the probability measure P. To be precise, we should speak of independence relative to P. Since we have considered P as given once and for all, by abuse of language we say simply *independent.*

(ii) Since both $L^2(\mathcal{B})$ and $L^2(\mathcal{C})$ contain the function $\mathbf{1}_\Omega$, they can never be orthogonal; independence corresponds to the strongest notion of orthogonality that can be expected.

(iii) Consider the codimension-1 subspace $\mathcal{H}$ composed of functions orthogonal to the constant functions:

$$\mathcal{H} = \{f \in L^2(\mathcal{A}) : \mathbf{E}(f) = 0\}.$$

The relation $\mathbf{E} = \mathbf{E}\mathbf{E}^{\mathcal{B}}$ implies $\mathbf{E}^{\mathcal{B}}(\mathcal{H}) \subset \mathcal{H}$. Moreover, 3.0.1 can be written as

$$\mathcal{H} \cap L^2(\mathcal{B}) \text{ is orthogonal to } \mathcal{H} \cap L^2(\mathcal{C}).$$

(iv) It follows from 3.0.1 that $L^2(\mathcal{B}) \cap L^2(\mathcal{C})$ reduces to the constant functions.

Since $L^2(\mathcal{B}) \cap L^2(\mathcal{C}) = L^2(\mathcal{B} \cap \mathcal{C})$, where $\mathcal{B} \cap \mathcal{C}$ is the σ-algebra of those functions in Ω that belong to both $\mathcal{B}$ and $\mathcal{C}$, we conclude that if $\mathcal{B}$ and $\mathcal{C}$ are *independent*, then $\mathcal{B} \cap \mathcal{C}$ reduces to the sets of probability zero and their complements. Up to sets of probability zero, $\mathcal{B} \cap \mathcal{C}$ is thus equivalent to the coarse σ-algebra.

3.0.2 Mutual independence of n sub-σ-algebras

Let $\mathcal{B}_1, \ldots, \mathcal{B}_n$ be n sub-σ-algebras of $\mathcal{A}$, let H be a subset of $[0,1]$, and let $\mathcal{B}_H$ be the σ-algebra generated by $\{\mathcal{B}_i : i \in H\}$. Then $\mathcal{B}_1, \ldots, \mathcal{B}_n$ are said to be *mutually independent* if

$$\mathcal{B}_H \text{ and } \mathcal{B}_{H^c} \text{ are independent } \sigma\text{-algebras for every } H \in \mathcal{P}([0,1]).$$

3.1 Independence of random variables and of σ-algebras

(i) Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\mathcal{B}$ and $\mathcal{C}$ be two sub-σ-algebras that are *independent* on this space. Let $\mathcal{B}'$ and $\mathcal{C}'$ be two other sub-σ-algebras such that $\mathcal{B}' \subset \mathcal{B}$ and $\mathcal{C}' \subset \mathcal{C}$. Then

$$\mathcal{B}' \text{ and } \mathcal{C}' \text{ are } \textit{independent.}$$

Indeed, $L^2(\mathcal{B}) \cap \mathcal{H} \supset L^2(\mathcal{B}') \cap \mathcal{H}$ and $L^2(\mathcal{C}) \cap \mathcal{H} \supset L^2(\mathcal{C}') \cap \mathcal{H}$. Hence the orthogonality of the first pair of subspaces implies the orthogonality of the second pair.

(ii) Let $X_1, \ldots, X_n$ be n random variables and let $\mathcal{B}_k = X_k^{-1}(\mathcal{B}_{\mathbf{R}})$. Then $X_1, \ldots, X_n$ are said to be *mutually independent* if the $\mathcal{B}_k$ are mutually independent σ-algebras.

(iii) Let $\mathcal{D}_1, \ldots, \mathcal{D}_n$ be mutually independent sub-σ-algebras of the probability space $(\Omega, \mathcal{A}, P)$. Let X_k be a $\mathcal{D}_k$-measurable r.v. defined on $(\Omega, \mathcal{A}, P)$. Then the r.v. X_k are *independent.*

This follows from (ii) and the fact that $X_k^{-1}(\mathcal{B}_{\mathbf{R}}) \subset \mathcal{D}_k$.

(iv) *Stability of independence under a change of variables.* Let $X_1, \ldots, X_n$ be independent r.v., let $\varphi_1, \ldots, \varphi_n$ be Borel functions from $\mathbf{R}$ to $\mathbf{R}$, and let $Y_k = \varphi_k(X_k)$. Then the Y_k are mutually independent r.v.

$Y_k^{-1}(\mathcal{B}_{\mathbf{R}}) \subset X_k^{-1}(\varphi_k^{-1}(\mathcal{B}_{\mathbf{R}})) \subset X_k^{-1}(\mathcal{B}_{\mathbf{R}})$, where the second inclusion holds since φ is Borel. (i) now implies the result. □

3.2 Expectation of a product of independent r.v.

3.2.1 Theorem. *Let $\mathcal{B}$ and $\mathcal{C}$ be two sub-σ-algebras of the probability space $(\Omega, \mathcal{A}, P)$. Then the following statements are equivalent:*

(i) $\mathcal{B}$ *and* $\mathcal{C}$ *are independent.*
(ii) $\mathbf{E}(fg) = \mathbf{E}(f)\mathbf{E}(g) \quad \forall f \in L^2(\mathcal{B}),\ g \in L^2(\mathcal{C})$.

PROOF. Decompose f and g as $f = u + \mathbf{E}(f)\mathbf{1}_\Omega$ and $g = v + \mathbf{E}(g)\mathbf{1}_\Omega$. Then $u \in \mathcal{H} \cap L^2(\mathcal{B})$ and $v \in \mathcal{H} \cap L^2(\mathcal{C})$. Moreover,

(iii) $\mathbf{E}(fg) = \mathbf{E}(uv) + \mathbf{E}(f)\mathbf{E}(g)$,

since $\mathbf{E}(u\ \mathbf{1}_\Omega) = 0$ and $\mathbf{E}(v\ \mathbf{1}_\Omega) = 0$ if $u,\ v \in \mathcal{H}$. In view of (iii), (ii) is equivalent to

$$\mathbf{E}(uv) = 0, \quad \forall u \in L^2(\mathcal{B}) \cap \mathcal{H},\ v \in L^2(\mathcal{C}) \cap \mathcal{H};$$

that is, to the orthogonality of $L^2(\mathcal{B}) \cap \mathcal{H}$ and $L^2(\mathcal{C}) \cap \mathcal{H}$.□

3.2.2 Proposition. *Let* $\mathcal{B}_1, \ldots, \mathcal{B}_n$ *be mutually independent sub-*σ*-algebras of the probability space* $(\Omega, \mathcal{A}, P)$*. If* $f_i \in L^\infty(\mathcal{B}_i)$*,* $i = 1, \ldots, n$*, then*

$$\mathbf{E}\left(\prod_{i=1}^n f_i\right) = \prod_{i=1}^n \mathbf{E}(f_i).$$

REMARK. The converse will be proved in 3.6.1.

PROOF. We proceed by induction on n. Assume that the theorem has been proved for $q < n$ and let $\prod_{i=2}^n f_i = h$.

Let $\mathcal{B}_H$ denote the σ-algebra generated by $\{f_i^{-1}(\mathcal{B}_\mathbf{R}) : 2 \leq i \leq n\}$. Then $h \in L^\infty(\mathcal{B}_H)$ and, since $f_1^{-1}(\mathcal{B}_\mathbf{R})$ and $\mathcal{B}_H$ are independent by 3.0.2, it follows from 3.2.1 that

$$\mathbf{E}(hf_1) = \mathbf{E}(h)\mathbf{E}(f_1).$$

We conclude by using the induction hypothesis $\mathbf{E}(h) = \prod_{i=2}^n \mathbf{E}(f_i)$. □

3.2.3 Corollary. *Let* $f_1, \ldots, f_n \in L^1(\Omega, \mathcal{A}, P)$ *and let* $h = \prod_{i=1}^n f_i$*. If the* f_i *are independent, then*

$$h \in L^1(\Omega, \mathcal{A}, P) \quad \textit{and} \quad \mathbf{E}(h) = \prod_{i=1}^n \mathbf{E}(f_i).$$

PROOF. We first prove the corollary under the hypothesis $f_i \geq 0$, $i = 1, \ldots, n$.

Let T_q be the truncation operator. By 3.1(iv), the $T_q(f_i)$ are independent; by 3.2.2,

$$\mathbf{E}\left(\prod_i T_q(f_i)\right) = \prod_i \mathbf{E}(T_q(f_i)) \leq \prod_i \mathbf{E}(f_i) = M.$$

Set $u_q = \prod_i T_q(f_i)$. Then $\{u_q\}$ is an increasing sequence and $\mathbf{E}(u_q) \leq M$; hence Fatou-Beppo Levi implies that $\lim u_q = h \in L^1$ and $\lim \mathbf{E}(u_q) = \mathbf{E}(h)$.

The general case is reduced to this special case by writing

$$f_i = f_i^0 - f_i^1, \quad \text{where} \quad f_i^0 = f_i^+ = \sup(f_i, 0) \quad \text{and} \quad f_i^1 = f_i^- = \sup(-f_i, 0),$$

and expanding the product

$$\prod f_i = \sum_\alpha (-1)^{\alpha_1+\ldots+\alpha_n} \prod_i f_i^{\alpha_i}.$$

Since the $f_i^{\alpha_i}$ are nonnegative,

$$\mathbf{E}\left(\prod_i f_i^{\alpha_i}\right) = \prod \mathbf{E}(f_i^{\alpha_i}).$$

As the sum of 2^n functions in L^1, $\prod f_i$ is in L^1. □

3.3 Conditional expectation and independence

3.3.1 Theorem. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\mathcal{B}$ and $\mathcal{C}$ be two sub-σ-algebras. Then the following two statements are equivalent:*

(i) The σ-algebras $\mathcal{B}$ and $\mathcal{C}$ are independent.
(ii) $\mathbf{E}^{\mathcal{B}}(f) = \mathbf{E}(f) \quad \forall f \in L^1(\mathcal{C})$.

REMARK. The roles of $\mathcal{B}$ and $\mathcal{C}$ can be interchanged for a different formulation of (ii).

This statement can be given the following concrete interpretation. If $\mathcal{B}$ and $\mathcal{C}$ are independent, then "knowledge of the events in the σ-algebra $\mathcal{B}$" in no way improves the "mean value" of a $\mathcal{C}$-measurable r.v.

PROOF. (i) $\Rightarrow$ (ii). Assume that $f \in L^2(\mathcal{C})$. Set

$$\tilde{f} = f - \mathbf{E}(f)\mathbf{1}_\Omega.$$

Then $\tilde{f} \in \mathcal{H}$ and

$$\mathbf{E}^{\mathcal{C}}(\tilde{f}) = \mathbf{E}^{\mathcal{C}}(f) - \mathbf{E}(f)\mathbf{1}_\Omega = f - \mathbf{E}(f)\mathbf{1}_\Omega = \tilde{f},$$

whence $\tilde{f} \in L^2(\mathcal{C}) \cap \mathcal{H}$.

By (i), $\tilde{f}$ is orthogonal to $L^2(\mathcal{B})$; thus $\mathbf{E}^{\mathcal{B}}(\tilde{f}) = 0$, or

$$\mathbf{E}^{\mathcal{B}}(f) = \mathbf{E}(f)\mathbf{1}_\Omega,$$

implying (ii).

When $f \in L^1$, we use the truncation operator and pass to the limit.
(ii) $\Rightarrow$ (i). Let $f \in L^2(\mathcal{C}) \cap \mathcal{H}$. Then, by (ii), $\mathbf{E}^{\mathcal{B}}(f) = 0$. That is, every f in $L^2(\mathcal{C}) \cap \mathcal{H}$ is orthogonal to $L^2(\mathcal{B})$, and it follows that $\mathcal{B}$ and $\mathcal{C}$ are independent. □

3.3.2 Corollary. *Let $\mathcal{B}$ and $\mathcal{C}$ be independent sub-σ-algebras of the probability space $(\Omega, \mathcal{A}, P)$. Then*

$$\mathbf{E}^{\mathcal{B}}\mathbf{E}^{\mathcal{C}} = \mathbf{E}.$$

PROOF. Let $f \in L^1(\Omega, \mathcal{A}, P)$. Then $\mathbf{E}^{\mathcal{C}} f \in L^1(\mathcal{C})$. Set $u = \mathbf{E}^{\mathcal{C}}(f)$; then

$$\mathbf{E}(u) = \mathbf{E}(\mathbf{E}^{\mathcal{C}}(f)) = \mathbf{E}(f).$$

Since $u \in L^1(\mathcal{C})$, it follows from 3.3.1 that $\mathbf{E}^{\mathcal{B}}(u) = \mathbf{E}(u) = \mathbf{E}(f)$. □

3.4 Independence and distributions (case of two random variables)

3.4.1 Theorem. *Let X_1 and X_2 be two r.v. defined on the probability space $(\Omega, \mathcal{A}, P)$. Let μ_1 and μ_2 denote the distributions of X_1 and X_2, respectively, and let μ denote their joint distribution. Then the following statements are equivalent:*

(i) X_1 and X_2 are independent r.v.
(ii) For all bounded Borel functions φ_1, φ_2 defined on $\mathbf{R}$,

$$\mathbf{E}(\varphi_1(X_1)\varphi_2(X_2)) = \mathbf{E}(\varphi_1(X_1))\mathbf{E}(\varphi_2(X_2)).$$

(iii) $\mu = \mu_1 \otimes \mu_2$.

PROOF. (i) $\Leftrightarrow$ (ii). Let $\mathcal{B}_i = X_i^{-1}(\mathcal{B}_{\mathbf{R}})$. Then the independence of the σ-algebras $\mathcal{B}_1$ and $\mathcal{B}_2$ is equivalent to that of the r.v. X_1 and X_2.

Let $f_i \in L^2(\mathcal{B}_i)$. By 1.5.4, the functional dependence theorem, there exist Borel functions $\psi_i : \mathbf{R} \to \mathbf{R}$ such that

$$\psi_i(X_i) = f_i \quad (i = 1, 2).$$

Hence (i) is equivalent, by 3.2.1, to

$$\mathbf{E}(\psi_1(X_1))\mathbf{E}(\psi_2(X_2)) = \mathbf{E}(\psi_1(X_1)\psi_2(X_2))$$

for all Borel functions ψ_i such that $\psi_i(X_i) \in L^2$.

Using the truncation operator shows that this last condition is equivalent to the more restrictive condition that ψ_i be a bounded Borel function; that is, to (ii).

(ii) $\Rightarrow$ (iii). Let C, $D \in \mathcal{B}_{\mathbf{R}}$. Set $\varphi = \mathbf{1}_C$ and $\psi = \mathbf{1}_D$. Then, computing the expectations by means of the distributions,

$$\int_{\mathbf{R}^2} \mathbf{1}_{C \times D} d\mu = \mathbf{E}(\mathbf{1}_C(X_1)\mathbf{1}_D(X_2)).$$

But

$$\int_{\mathbf{R}} \mathbf{1}_C d\mu_1 = \mathbf{E}(\mathbf{1}_C(X_1)),$$

whence, using (ii),

$$\mu(C \times D) = \mu_1(C)\mu_2(D).$$

Since μ is a Borel measure on $\mathbf{R}^2$ and $\mathcal{B}_{\mathbf{R}^2} = \mathcal{B}_{\mathbf{R}} \otimes \mathcal{B}_{\mathbf{R}}$, this last relation shows by I-8 that $\mu = \mu_1 \otimes \mu_2$.

(iii) $\Rightarrow$ (ii). Again using the distributions to compute the expectations, we have

$$\mathbf{E}(\varphi_1(X_1)\varphi_2(X_2)) = \int_{\mathbf{R}^2} \varphi_1(\xi_1)\varphi_2(\xi_2)d\mu_1(\xi_1) \otimes d\mu_2(\xi_2).$$

By Fubini's theorem, this is equal to

$$\int_{\mathbf{R}} \varphi_1(\xi_1)d\mu_1(\xi_1) \left[\int \varphi_2(\xi_2)d\mu_2(\xi_2)\right],$$

and (ii) is proved. □

3.5 A function space on the σ-algebra generated by two σ-algebras

3.5.1 Theorem. *Let $\mathcal{B}$ and $\mathcal{C}$ be two sub-σ-algebras of the probability space $(\Omega, \mathcal{A}, P)$ and let $\mathcal{D}$ denote the σ-algebra they generate. Let V be the vector subspace of $L^\infty(\mathcal{A})$ defined by*

$$V = \left\{ h \in L^\infty(\mathcal{A}) : h = \sum_{i=1}^{n} f_i g_i, \ \textit{with } f_i \in L^\infty(\mathcal{B}), \ g_i \in L^\infty(\mathcal{C}) \right\}.$$

Then $L^2(\mathcal{D}) \supset V$ and V is dense in $L^2(\mathcal{D})$.

PROOF. We prove the theorem in the special case that there exist two mappings $u : \Omega \to \mathbf{R}^n$ and $v : \Omega \to \mathbf{R}^p$ such that

$$(i) \qquad u^{-1}(\mathcal{B}_{\mathbf{R}^n}) = \mathcal{B} \quad \text{and} \quad v^{-1}(\mathcal{B}_{\mathbf{R}^p}) = \mathcal{C}.$$

Let $w : \Omega \to \mathbf{R}^{n+p}$ be defined by $w(\omega) = (u(\omega), v(\omega))$. Then $w^{-1}(\mathcal{B}_{\mathbf{R}^{n+p}})$ is a σ-algebra containing $\mathcal{B}$ and $\mathcal{C}$.

Moreover, by I-2.4.2,

$$\mathcal{B}_{\mathbf{R}^{n+p}} = \mathcal{B}_{\mathbf{R}^n} \otimes \mathcal{B}_{\mathbf{R}^p}.$$

Hence $\mathcal{B}_{\mathbf{R}^{n+p}}$ is generated by the rectangles $R = X \times Y$, with $X \in \mathcal{B}_{\mathbf{R}^n}$, $Y \in \mathcal{B}_{\mathbf{R}^p}$. We have

$$\begin{aligned} w^{-1}(R) &= \{\omega : u(\omega) \in X \text{ and } v(\omega) \in Y\} \\ &= u^{-1}(X) \cap v^{-1}(Y). \end{aligned}$$

That is, $w^{-1}(R) \in \mathcal{D}$. With the hypothesis (i), we have thus shown that

$$(ii) \qquad \mathcal{D} = w^{-1}(\mathcal{B}_{\mathbf{R}^{n+p}}).$$

Let ρ be the distribution of w and let w^* be the inverse image mapping. Then it follows from 1.7.2 that

$$(iii) \qquad w^* : L^2(\mathbf{R}^{n+p}, \rho) \to L^2(\mathcal{D}),$$

and the mapping is a surjective isometry.

The continuous functions with compact support, $C_K(\mathbf{R}^{n+p})$, are dense in $L^2(\mathbf{R}^{n+p}; \rho)$. (See II-3.)

Let $\varphi \in C_K(\mathbf{R}^{n+p})$. Then, by the Stone-Weierstrass theorem, there exists a sequence of polynomials P_r converging uniformly to φ on a compact set $K_1 \times K_2$ which contains the support of φ. Let $g_r = P_r \mathbf{1}_{K_1 \times K_2}$. Then

$$(iv) \qquad \|g_r - \varphi\|_{L^2(\rho)} \to 0.$$

We now show that

$$(v) \qquad w^*(g_r) \in V.$$

This follows since the polynomial P_r is the sum of monomials of the form

$$(u^1)^{m_1} \dots (u^n)^{m_n} (v_1)^{q_1} \dots (v^p)^{q_p}$$

and, setting

$$f = \mathbf{1}_{K_1} (u^1)^{m_1} \dots (u^n)^{m_n} \quad \text{and} \quad g = \mathbf{1}_{K_2} (v^1)^{q_1} \dots (v^p)^{q_p},$$

we can write $w^*(g_r)$ as a linear combination of functions of the form fg. Thus (v) holds.

Since w^* is an isometry, $w^*(C_K(\mathbf{R}^{n+p}))$ is dense in $L^2(\mathcal{D})$, and to the convergence of g_r to φ in $L^2(\rho)$ there corresponds a convergence in $L^2(\mathcal{D})$.

(vi) REMARK. To prove the theorem without the hypothesis (i), we would consider finite systems of $\mathcal{B}$-measurable functions $u_1, \dots, u_n$ and $\mathcal{C}$-measurable functions $v_1, \dots, v_q$. Then $\mathcal{B}$ could be viewed as the σ-algebra generated by all the $u^{-1}(\mathcal{B}_{\mathbf{R}^n})$, and similarly for $\mathcal{C}$. We would then "pass to the limit". This passage to the limit will be carried out in detail for closely related cases in Section 6 of this chapter.

3.5.2 Corollary. *Let $\mathcal{B}_1, \dots, \mathcal{B}_n$ be a finite collection of sub-σ-algebras of the probability space $(\Omega, \mathcal{A}, P)$ and let $\mathcal{D}$ be the σ-algebra generated by $\mathcal{B}_1, \dots, \mathcal{B}_n$. Set*

$$W_n = \left\{ h \in L^\infty(\mathcal{A}) : h = \sum_{p=1}^{q} f_p^1 f_p^2 \dots f_p^n, \ \text{with } f_p^i \in L^\infty(\mathcal{B}_i) \right\}.$$

Then $W_n \subset L^2(\mathcal{D})$ and W_n is dense in $L^2(\mathcal{D})$.

PROOF. We proceed by induction on n. Let $\mathcal{C}$ be the σ-algebra generated by $\mathcal{B}_2, \dots, \mathcal{B}_n$. Then, by the induction hypothesis,

$$(i) \qquad W_{n-1} \text{ is dense in } L^2(\mathcal{C}).$$

The σ-algebra generated by $\mathcal{B}_1, \ldots, \mathcal{B}_n$ equals the σ-algebra generated by $\mathcal{B}_1$ and $\mathcal{C}$. Let

$$V = \{h : h = \sum f_i g_i, \text{ with } f_i \in L^\infty(\mathcal{B}_1), \; g_i \in L^\infty(\mathcal{C})\}.$$

Then, by 3.5.2, V is dense in $L^2(\mathcal{D})$. Let

$$(ii) \qquad V' = \{h : h = \sum f_i g_i, \text{ where } f_i \in L^\infty(\mathcal{B}_1), \; g_i \in L^2(\mathcal{C})\}.$$

Then $V' \subset L^2(\mathcal{D})$, and V' is dense in $L^2(\mathcal{D})$ since $V' \supset V$. By (i), each g_i can be approximated by elements of W_{n-1}. Hence there exists a sequence $k_i^s \in W_{n-1}$ such that $\|k_i^s - g_i\|_{L^2} \to 0$, and

$$\|\sum f_i g_i - \sum f_i k_i^s\|_{L^2} \leq \sum \|f_i\|_{L^\infty} \|g_i - k_i^s\|_{L^2}.$$

The right-hand side tends to zero, and we conclude by noting that $\sum f_i k_i^s \in W_n$. □

3.6 Independence and distributions (case of n random variables)

Theorem. *Let $\mathcal{B}_1, \ldots, \mathcal{B}_n$ be n sub-σ-algebras of the probability space $(\Omega, \mathcal{A}, P)$. Then the following statements are equivalent:*

(i) $\mathcal{B}_1, \ldots, \mathcal{B}_n$ are mutually independent.
(ii) $\mathbf{E}\left(\prod_{i=1}^n f_i\right) = \prod_{i=1}^n \mathbf{E}(f_i)$ for any $f_i \in L^\infty(\mathcal{B}_i)$.

Proof. Recall that the direction (i) ⇒ (ii) was proved in 3.2.2. We now prove that (ii) ⇒ (i). Let H be a subset of $\{1, \ldots, n\}$, let H' be the complement of H, and let $\mathcal{C}$ and $\mathcal{C}'$ denote the σ-algebras generated by $\{B_i : i \in H\}$ and $\{B_j : j \in H'\}$, respectively. We must prove the independence of $\mathcal{C}$ and $\mathcal{C}'$. By 3.2.1, this will follow from the identity

(iii) $\mathbf{E}(gg') = \mathbf{E}(g)\mathbf{E}(g') \quad \forall g \in L^2(\mathcal{C}), \; g' \in L^2(\mathcal{C}')$.

Using (ii), the function space constructed in 3.5.2 on the σ-algebra generated by $\mathcal{C}$ and $\mathcal{C}'$, and bilinearity, it suffices to calculate

$$\mathbf{E}\left(\prod_{i \in H} f^i \prod_{j \in H'} f^j\right) = \prod_{i \in H} \mathbf{E}(f^i) \prod_{j \in H'} \mathbf{E}(f^j).$$

Using (ii) again and setting $f^i = \mathbf{1}$ if $i \in H'$, we find that the first term on the right-hand side is $\mathbf{E}(g)$, and similarly the second is $\mathbf{E}(g')$. This proves (iii). □

3.6.2 Theorem. *Let $X_1, \ldots, X_n$ be n r.v. on the probability space $(\Omega, \mathcal{A}, P)$. Then the following statements are equivalent:*

(i) The r.v. X_k are mutually independent.
(ii) For all bounded Borel functions φ_k on $\mathbf{R}$,

$$\mathbf{E}\left(\prod_{k=1}^{n} \varphi_k(X_k)\right) = \prod_{k=1}^{n} \mathbf{E}(\varphi_k(X_k)).$$

(iii) Let μ denote the joint distribution of $X_1, \dots, X_n$ and let μ_i denote the distribution of X_i. Then

$$\mu(A_1 \times A_2 \times \dots \times A_n) = \prod_{i=1}^{n} \mu_i(A_i)$$

for any $A_i \in \mathcal{B}_{\mathbf{R}}$. In other words, $\mu = \otimes \mu_i$.

PROOF. The theorem was proved in 3.4.1 for $n = 2$. The general case is proved in the same way, with Theorem 3.2.1 replaced by Theorem 3.6.1. □

4 Characteristic Functions and Theorems on Convergence in Distribution

4.1 The characteristic function of a random variable

Let $(\Omega, \mathcal{A}, P)$ be a probability space on which the $\mathbf{R}^n$-valued r.v.

$$X = (X^1, \dots, X^n)$$

is defined. The *characteristic function* of the r.v. X is the function defined on $\mathbf{R}^n = \{(t_1, \dots, t_n)\}$ by

$$\varphi_X(t_1, t_2, \dots, t_n) = \mathbf{E}(\exp[i(t_1 X^1 + t_2 X^2 + \dots + t_n X^n)]),$$

where $i = \sqrt{-1}$. Since the imaginary exponential is a function with modulus 1, the expectation of the right-hand side exists for every $t \in \mathbf{R}^n$.

4.1.1 Determining the distribution from its characteristic function

Proposition. *Let $(\Omega, \mathcal{A}, P)$ and $(\Omega', \mathcal{A}', P')$ be probability spaces and let X and X' be $\mathbf{R}^n$-valued r.v. Then statements (i) and (ii) are equivalent.*

(i) $\varphi_X(t) = \varphi_{X'}(t), \quad \forall t \in \mathbf{R}^n$.
(ii) X and X' have the same distribution.

PROOF. Let μ and μ' be the distributions of X and X'.

Calculating the expectations by means of the distributions, we obtain

$$\varphi_X(t) = \int_{\mathbf{R}^n} e^{it.x} d\mu(x) \quad \text{and} \quad \varphi_{X'}(t) = \int_{\mathbf{R}^n} e^{it.x} d\mu'(x).$$

That is,

$$(iii) \qquad \varphi_X(t) = \widehat{\mu}(t),$$

where $\widehat{\mu}$ denotes the Fourier transform of μ.

But it was shown in III-2.6 that two measures with the same Fourier transform coincide. □

4.1.2 Convergence in distribution and characteristic functions

Theorem (Paul Lévy). *Let $\{X_p\}$ be a sequence of $\mathbf{R}^n$-valued r.v. defined on different probability spaces. Then the following statements are equivalent:*

(i) $\{X_p\}$ converges in distribution.
(ii) The functions $\varphi_{X_p}(t)$ converge uniformly on compact sets.

Moreover, if (ii) holds, let

$$\psi(t) = \lim \varphi_{X_p}(t).$$

Then there exists a positive Radon measure ν of total mass 1 on $\mathbf{R}^n$ such that $\widehat{\nu}(t) = \psi(t)$ and the distributions of the X_p converge to ν.

PROOF. We first prove that (ii) $\Rightarrow$ (i). Let μ_p denote the distribution of X_p.

(iii) Vague convergence of the μ_p.

Consider the linear functionals l_p on $C_0(\mathbf{R}^n)$ defined by

$$l_p(\mu) = \int u d\mu_p.$$

Then $|l_p(u)| \leq \|u\|_{C_0}$. Moreover, by Parseval's lemma (III-2.6),

$$l_p(f) = \frac{1}{(2\pi)^n} \int \widehat{f}(t)\varphi_{X_p}(-t)dt, \quad \forall f \in A(\mathbf{R}^n).$$

Since $\widehat{f} \in L^1$, we can apply the dominated convergence theorem to obtain

$$\lim(l_p(f)) = \frac{1}{(2\pi)^n} \int \widehat{f}(t)\psi(-t)dt.$$

Since $A(\mathbf{R}^n)$ is dense in $C_0(\mathbf{R}^n)$ (cf. III-2.5), II-6.8(iii) can be applied to show that there exists $\nu \in M_+^1(\mathbf{R}^n)$ such that the μ_p converge weakly to ν; that is,

$$\lim \int u d\mu_p = \int u d\nu, \quad \forall u \in C_0(\mathbf{R}^n).$$

Furthermore,

$$\int \widehat{f}(t)\widehat{\nu}(-t)dt = \int \widehat{f}(t)\psi(-t)dt.$$

Since $(A(\mathbf{R}^n))^\wedge = A(\mathbf{R}^n)$ is dense in $C_0(\mathbf{R}^n)$, it follows that $\widehat{\nu}(t) = \psi(t)$.

(iv) Narrow convergence of the μ_p.

Only the pointwise convergence of $\varphi_{X_p}(t)$ was used to prove (iii). We must now exploit uniform convergence. Let

$$G_\lambda(x) = \frac{1}{(2\pi\lambda)^{n/2}} \exp\left(-\frac{\|x\|^2}{2\lambda}\right).$$

Then

$$(G_\lambda)^\wedge(t) = \exp\left(-\frac{\lambda}{2}\|t\|^2\right).$$

Consider the following integral of the *nonnegative function* $(1-\widehat{G}_\lambda)$ with respect to the *positive measure* $d\mu_p$:

$$I_{\lambda,p} = \int (1 - \widehat{G}_\lambda(x))d\mu_p(x).$$

Writing Parseval's relation and taking into account that $(\widehat{G}_\lambda)^\wedge(t) = G_\lambda(-t) = G_\lambda(t)$ and that $\widehat{\mu}_p(0) = 1$, we obtain

$$I_{\lambda,p} = 1 - \int \widehat{\mu}_p(t) G_\lambda(t)dt.$$

Since $\int G_\lambda = 1$, this can be written

$$I_{\lambda,p} = \int (1 - \widehat{\mu}_p(t))G_\lambda(t)dt = \int_{\|t\|<\eta} + \int_{\|t\|>\eta},$$

where η is determined by first fixing q such that $|\widehat{\mu}_p(t) - \widehat{\mu}_q(t)| < \epsilon$ if $p \geq q$ and $\|t\| \leq 1$, then choosing $\eta < 1$ such that $|\widehat{\mu}_q(t) - \widehat{\mu}_q(0)| < \epsilon$ if $\|t\| \leq \eta$. Then

$$(v) \qquad |\widehat{\mu}_p(t) - 1| < 3\epsilon \quad \text{if} \quad \|t\| < \eta,$$

whence

$$0 < I_{\lambda,p} < 3\epsilon \int_{\mathbf{R}^n} G_\lambda + 2\int_{\|t\|>\eta} G_\lambda.$$

The first integral equals 1; the second tends to zero as λ tends to zero, for *fixed* η. Hence there exists λ_0 such that $|I_{\lambda_0,p}| < 4\epsilon$ for every $p \geq q$. Let $h \in C_b(\mathbf{R}^n)$ and set $u = h\widehat{G}_{\lambda_0}$; then $u \in C_0(\mathbf{R}^n)$ and, by (iii),

$$\int u d\mu_p \to \int u d\nu.$$

Since $1 - \widehat{G}_{\lambda_0}$ is nonnegative,

$$\left|\int (u-h)d\mu_p\right| \le \|h\|_{C_b}\|1-\widehat{G}_{\lambda_0}\|_{L^1_\mu} = \|h\|_{C_b}\int (1-\widehat{G}_{\lambda_0})\, d\mu_p \le 4\epsilon\|h\|_{C_b}.$$

Moreover, $\widehat{\nu}(t) = \psi(t)$, the limit of the $\widehat{\mu}_p(t)$, satisfies (v). Similarly,

$$\left|\int (u-v)d\nu\right| \le 4\epsilon\|h\|_{C_b},$$

and finally

$$\lim \int h d\mu_p = \int h d\nu, \quad \forall h \in C_b(\mathbf{R}^n).$$

In particular, taking $h = 1$ shows that $\nu(\mathbf{R}^n) = 1$; that is, ν is a probability measure and the X_p converge in distribution to the distribution ν. This proves (i).

PROOF OF (i) $\Rightarrow$ (ii). By the definition of narrow convergence,

$$\int \mathrm{e}^{it.x} d\mu_p(x) \to \int \mathrm{e}^{it.x} d\nu(x)$$

for every fixed t. We must now prove uniform convergence in t. By II-6.8((ii) $\Rightarrow$ (iv)), given $\epsilon > 0$ there exists M such that $\mu_p([-M, M]^c) < \epsilon$ for p sufficiently large. Then

$$\varphi_{X_p}(t) = \int_{-M}^{+M} \mathrm{e}^{it.x} d\mu_p(x) + \theta\epsilon, \quad \text{where} \quad |\theta| < \epsilon.$$

Differentiating with respect to t under the integral sign shows that the first partial derivatives of φ_{X_p} are bounded by M. Hence the $\varphi_{X_p}(t)$ are equicontinuous functions, and the result follows by Ascoli's theorem that pointwise convergence on a compact set implies uniform convergence.[2] □

4.1.3 Differentiability of characteristic functions

Proposition. *Let X be a r.v. with values in $\mathbf{R}^n$. Suppose that*

$$\mathbf{E}(\|X\|^p_{\mathbf{R}^n}) < \infty, \quad \textit{where} \quad p \ge 1.$$

Then φ_X is r times continuously differentiable in t for $r \le p$ and

$$(i) \quad \left(\left(\frac{\partial}{\partial t_1}\right)^{r_1} \cdots \left(\frac{\partial}{\partial t_n}\right)^{r_n} \varphi_X\right)(t) = \mathbf{E}[(iX^1)^{r_1} \ldots (iX^n)^{r_n} e^{it.X}].$$

PROOF. Using the criterion for differentiation under the integral sign (I-7), we have

$$\frac{\partial}{\partial t_1}\mathbf{E}(\exp[i(t_1X^1 + \ldots + t_nX^n)]) = \mathbf{E}(iX^1 \exp[i(t_1X^1 + \ldots + t_nX^n)]).$$

[2]See Bourbaki, *General Topology*, X.2.4.

The result follows by noting that

$$\|X\| \in L^p \Rightarrow (X^1)^{r_1}(X^2)^{r_2}\dots(X^n)^{r_n} \in L^1 \quad \text{if} \quad r_1 + r_2 + \dots + r_n \le p.\square$$

4.1.4 Taylor series expansion of a characteristic function at the origin

Proposition. *Let X be a r.v. with values in $\mathbf{R}^n$ and suppose that $\|X\|_{\mathbf{R}^n} \in L^2$. Then*

$$(i) \qquad \varphi_X(t) = 1 + i\mathbf{E}(X).t - \frac{1}{2}q_X(t) + o(\|t\|^2),$$

where $q_X(t) = \sum a^{k,j} t_k t_j$ and $a^{k,j} = \mathbf{E}(X^k X^j)$.

The matrix $a^{k,j}$ is symmetric and nonnegative; that is,

$$(ii) \qquad q_X(t) \ge 0 \quad \textit{for every} \quad t \in \mathbf{R}^n.$$

PROOF. Since $\|X\|_{\mathbf{R}^n} \in L^2$, 4.1.3 implies that φ_X is twice continuously differentiable. The derivatives at the origin can be computed using 4.1.3(i), and (i) follows by using Taylor's formula with remainder.

Moreover,

$$q_X(t) = \sum_{k,j} t_k t_j \mathbf{E}(X^k X^j) = \mathbf{E}\left[\left(\sum_s t_s X^s\right)^2\right] \ge 0.\square$$

4.1.5 **Definitions.** X is said to be *centered* if $\mathbf{E}(X) = 0$.

If X is not centered, a centered r.v. is obtained by setting $Y = X - \mathbf{E}(X)\mathbf{1}_\Omega$. The quadratic form $q_Y(t)$ associated with the centered variable is called the *covariance* of X and written $\sigma_X(t)$.

4.2 Characteristic function of a sum of independent r.v.

4.2.1 Proposition. *Let $X_1, \dots, X_p$ be mutually independent $\mathbf{R}^n$-valued r.v. on the probability space $(\Omega, \mathcal{A}, P)$. Let*

$$(i) \qquad \varphi_{X_k}(t) = \mathbf{E}(e^{it.X_k})$$

be their characteristic functions, and set

$$(ii) \qquad S = X_1 + \dots + X_p.$$

Then

$$(iii) \qquad \varphi_S(t) = \prod_{k=1}^{p} \varphi_{X_k}(t).$$

PROOF. $\varphi_S(t) = \mathbf{E}(\mathrm{e}^{it.X_1}\mathrm{e}^{it.X_2}\dots\mathrm{e}^{it.X_p})$.

Using 3.2.2 with $f_k = \mathrm{e}^{it.X_k}$, we obtain

$$\varphi_S(t) = \prod_k \mathbf{E}(\mathrm{e}^{it.X_k}) = \prod_k \varphi_{X_k}(t). \square$$

4.2.2 Corollary. *With the notation of 4.2.1, let $\mu_1, \dots, \mu_p$ be the distributions of $X_1, \dots, X_p$ and let ν be the distribution of S. Then*

$$\nu = \mu_1 * \mu_2 * \cdots \mu_p.$$

PROOF. Using 4.1.1(iii), we may write 4.2.1(iii) in the form

$$\widehat{\nu} = \widehat{\mu}_1(t)\widehat{\mu}_2(t) \dots \widehat{\mu}_p(t).$$

By III-1.4.4, the convolution product of measures corresponds to the product of the Fourier transforms. □

4.2.3 Proposition. *Let $X_1, \dots, X_p$ be independent $\mathbf{R}^n$-valued r.v. Suppose that $\|X_k\|_{\mathbf{R}^n} \in L^2$, $1 \leq k \leq p$, and let $S = X_1 + \cdots X_p$. Then the covariance forms are related by*

$$\sigma_S(t) = \sum_{k=1}^{p} \sigma_{X_k}(t).$$

PROOF. Setting $\tilde{X}_k = X_k - \mathbf{E}(X_k)\mathbf{1}_\Omega$, we can reduce the proof to the case where the X_k are centered; then S is centered. We must verify the identity $q_S(t) = \sum_{k=1}^{p} q_{X_k}(t)$, or

$$\mathbf{E}\left(\left(\sum_j t_j S^j\right)^2\right) = \mathbf{E}\left(\left(\sum_k t_k \sum_j X_j^k\right)^2\right) = \sum_{k,l} t_k t_l \mathbf{E}\left(\sum_{j,m} X_j^k X_m^l\right).$$

But, for $j \neq m$, X_j^k and X_m^l are independent r.v. by 3.1(iv). Hence, by 3.2.1,

$$\mathbf{E}(X_j^k X_m^l) = \mathbf{E}(X_j^k)\mathbf{E}(X_m^l) = 0.$$

Thus

$$q_S(t) = \sum_{k,l} t_k t_l \sum_j \mathbf{E}(X_j^k X_j^l) = \sum_j q_{X_j}(t). \square$$

4.3 Laplace's theorem and Gaussian distributions

4.3.1 Laplace's theorem. *Let $X_1, X_2, \ldots, X_p, \ldots$ be a sequence of independent $\mathbf{R}^n$-valued r.v. defined on the probability space $(\Omega, \mathcal{A}, P)$.*

Suppose that the X_p all have the same distribution, that $\|X_1\|_{\mathbf{R}^n} \in L^2$, and that $\mathbf{E}(X_1) = 0$. Set

$$G_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n).$$

Then the sequence of r.v. G_n converges in distribution to a r.v. G with characteristic function

$$\varphi_G(t) = \exp\left(-\frac{1}{2}q_{X_1}(t)\right).$$

PROOF. Since $G_n = \frac{1}{\sqrt{n}} S_n$,

$$\varphi_{G_n}(t) = \mathbf{E}\left(\exp\left(it \cdot \frac{S_n}{\sqrt{n}}\right)\right) = \mathbf{E}\left(\exp\left(\frac{it}{\sqrt{n}} \cdot S_n\right)\right) = \varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right).$$

By 4.2.1, $\varphi_{S_n}(t) = (\varphi_{X_1}(t))^n$. Hence

$$\varphi_G(t) = \left[\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n,$$

or

$$\varphi_{G_n}(t) = \exp\left(n \log\left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)\right).$$

(Note that $\varphi_{X_1}(0) = 1$. By continuity, there exists $\epsilon > 0$ such that, for $|t| < \epsilon$, $|\varphi_{X_1}(t)| \neq 0$ and $-\frac{\pi}{2} < \arg \varphi_{X_1}(t) < \frac{\pi}{2}$. Thus $\log \varphi_{X_1}(t)$ is well defined for $|t| < \epsilon$.) Furthermore,

$$\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{1}{2n}q_{X_1}(t) + o\left(\frac{1}{n}\right)$$

uniformly in t, when t ranges over a compact subset of $\mathbf{R}^n$. Hence

$$\log \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = -\frac{1}{2n}q_{X_1}(t) + o\left(\frac{1}{n}\right),$$

and $\varphi_{G_n}(t) \to \exp(-\frac{1}{2}q_{X_1}(t))$ uniformly on compact sets. 4.1.2 implies the result. □

4.3.2 Gaussian distributions

With the next few results, we make Laplace's theorem more explicit by computing the distribution of G.

(i) Lemma. $\int_{\mathbf{R}} \exp\left(-\frac{t^2}{2}\right) dt = (2\pi)^{1/2}$.

PROOF. We prove this well-known result by using a trick from real variables. Let $I = \int_{\mathbf{R}} \exp\left(-\frac{t^2}{2}\right) dt$. Then, by Fubini's theorem,

$$I^2 = \int_{\mathbf{R}^2} \exp\left(-\frac{u^2}{2}\right) \exp\left(-\frac{v^2}{2}\right) dudv.$$

Passing to polar coordinates, let $u = r\sin\theta$ and $v = r\cos\theta$, with $r > 0$ and $0 \leq \theta < 2\pi$.

This change of coordinates defines a diffeomorphism of $\mathbf{R}^2$ with Jacobian matrix

$$J = \begin{pmatrix} \sin\theta & \cos\theta \\ r\cos\theta & -r\sin\theta \end{pmatrix}.$$

Since $|\det \mathrm{J}| = r\ dr\ d\theta$,

$$I^2 = \int_0^{2\pi} \int_0^{+\infty} \exp\left(-\frac{r^2}{2}\right) r\ dr\ d\theta = 2\pi \int_0^{+\infty} \exp\left(-\frac{r^2}{2}\right) r\ dr.$$

The last integral can be computed by setting $r^2 = w$. Thus

$$\int_0^{+\infty} \exp\left(-\frac{r^2}{2}\right) r\ dr = \int_0^{+\infty} \exp(-w)dw = 1.\square$$

(ii) Lemma. $\frac{1}{(2\pi)^{1/2}} \int_{\mathbf{R}} \exp\left(itx - \frac{t^2}{2}\right) dt = \exp\left(-\frac{x^2}{2}\right)$.

PROOF. Let τ be an auxiliary parameter defined by $p_\tau(t) = \frac{1}{(2\pi\tau)^{1/2}} \exp\left(-\frac{t^2}{2\tau}\right)$. It is straightforward to verify that

$$(i) \qquad \frac{\partial p_\tau}{\partial \tau} = \frac{1}{2}\frac{\partial^2 p_\tau}{\partial t^2}.$$

Note that $x \mapsto p_\tau(x)$ is an element of the space $\mathcal{S}(\mathbf{R})$. By III-4.2, differentiation with respect to x is mapped to multiplication by $-it$ of the Fourier transform:

$$\widehat{p}_\tau(t) = \int_{\mathbf{R}} p_\tau(x) \exp(itx) dx.$$

By I-7.8.4, we can differentiate under the integral sign; thus (i) can be written

$$(ii) \qquad \frac{\partial}{\partial \tau}\widehat{p}_\tau(t) = -\frac{t^2}{2}\widehat{p}_\tau(t).$$

Note that, as $\tau \to 0$, $p_\tau(t)dt$ converges narrowly to the Dirac measure at zero. Hence $\widehat{p}_\tau(t) \to 1$ for each fixed t as $\tau \to 0$. The differential equation (ii) thus gives $\widehat{p}_\tau(t) = \exp\left(-\frac{t^2}{2}\tau\right)$; the lemma follows by setting $\tau = 1$. $\square$

(iii) Lemma. *Let* $Q(t) = (t_1)^2 + \cdots (t_n)^2$. *Then*

$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbf{R}^n} \exp\left[-\frac{1}{2}Q(x) + it.x\right] dx = \exp\left[-\frac{1}{2}Q(t)\right].$$

PROOF. Since $\exp\left[-\frac{1}{2}Q(x)\right] = \prod_{k=1}^n \exp\left[-\frac{1}{2}(x^k)^2\right]$, the conclusion follows from (ii) and Fubini's theorem. □

4.3.3 **Definition.** A *Gaussian distribution* is a measure μ on $\mathbf{R}^n$ with Fourier transform $\widehat{\mu}$ of the form

$$(i) \qquad \widehat{\mu}(t) = \exp\left(-\frac{1}{2}h(t)\right),$$

where $h(t)$ is a positive quadratic form.

4.3.4 Proposition. *Let μ be a Gaussian distribution given by 4.3.3(i). Suppose that h is positive definite. Then*

$$d\mu = c\exp\left[-\frac{1}{2}h_1(x)\right] dx,$$

where c is a normalizing constant such that $\int d\mu = 1$ and $h_1(x)$ is the adjoint of h, defined by

$$(ii) \qquad h_1(x) = \sup\{t.x : h(t) \leq 1\}.$$

PROOF. Let a basis be chosen such that $h(t) = \sum t_k^2$; then 4.3.2(iii) implies (i) with $h_1(x) = \sum x_k^2$. Using formula (ii), $h_1(x)$ can be defined without changing bases. □

4.3.5 Proposition. *Let μ be a Gaussian distribution of the form 4.3.3(i). Let*

$$V = \{t : h(t) = 0\} \quad \text{and} \quad V^\perp = \{x : t.x = 0 \ \ \forall t \in V\}.$$

Then μ is a measure with support $V^\perp$. Let $y \in V^\perp$ and let dy be the volume measure on $V^\perp$. Then

$$d\mu = c \ \exp\left[-\frac{1}{2}h_1(y)\right] dy,$$

where h_1 is the quadratic form defined for $y \in V^\perp$ by

$$h_1(y) = \sup\{t.y : h(t) \leq 1\}.$$

REMARK. The quadratic form h_1 is positive definite.

PROOF. Let x, $t \in \mathbf{R}^n$ be decomposed as

$$x = y + z, \quad \text{where} \quad y \in V^\perp, \ z \in V;$$
$$t = \eta + \zeta, \quad \text{where} \quad \eta \in V^\perp, \ \zeta \in V.$$

Then

$$\int e^{it(y+z)} d\mu = c \int_{V^\perp} e^{i\eta \cdot y} \exp\left[-\frac{1}{2}h_1(y)\right] dy. \square$$

5 Theorems on Convergence of Martingales

5.1 Martingales

5.1.1 Definition of a filtration

Let $(\Omega, \mathcal{A}, P)$ be a probability space. A *filtration* of the space is an increasing sequence $\{\mathcal{A}_n\}$ of sub-σ-algebras of $\mathcal{A}$ such that

$$\mathcal{A}_0 \subset \mathcal{A}_1 \subset \ldots \subset \mathcal{A}_n \subset \ldots.$$

Let $\mathcal{A}_\infty$ be the σ-algebra generated by all the $\mathcal{A}_n$; we write $\mathcal{A}_\infty = \lim \mathcal{A}_n$. The filtering sequence is said to converge to $\mathcal{A}$ if $\mathcal{A}_\infty = \mathcal{A}$.

The phenomenological meaning of an increasing sequence of σ-algebras is clear. Let $0, 1, 2, \ldots, n$ be the various instants of an "experiment".

Let $\mathcal{A}'_n$ be the Boolean algebra generated by all the observations made up to time n (in the sense of 1.1). Then $\mathcal{A}'_n$ encapsulates all the experimenter's knowledge of the system at time n. The σ-algebra generated by $\mathcal{A}'_n$ is written $\mathcal{A}_n$, and might be called the *σ-algebra of the past* at time n.

5.1.2 Sequence of r.v. adapted to a filtration

Let $(\Omega, \mathcal{A}, P)$ be a probability space equipped with a filtration $\mathcal{A}_n$. A sequence of r.v. $\{X_n\}$ in $L^0(\Omega, \mathcal{A})$ is said to be *adapted to the filtration* if $X_n \in L^0(\Omega, \mathcal{A}_n)$.

5.1.3 Given a sequence $\{Y_K\}$ in $L^0(\Omega, \mathcal{A})$, let $\mathcal{A}_k^Y$ be the σ-algebra generated by $Y_s^{-1}(\mathcal{B}_{\mathbf{R}})$, where $s \leq k$.

Then the A_k^Y form a filtration of $(\Omega, \mathcal{A}, P)$. Moreover, *the sequence of r.v. Y_n is adapted to the filtration $\mathcal{A}_n$ if and only if $\mathcal{A}_n \supset \mathcal{A}_n^Y$ for any n.*

5.1.4 Remark. $\mathcal{A}_n^Y$ might be called the σ-algebra of the past corresponding to the "experiment" that consists of observing the values of $Y_1(\omega), \ldots, Y_n(\omega)$.

5.1.5 Definition of a martingale

Definition. Let $(\Omega, \mathcal{A}, P)$ be a probability space equipped with a filtration $\{\mathcal{A}_n\}$. A sequence $\{X_n\}$ of r.v. is called a martingale if

(i) *the X_n are integrable: $X_n \in L^1(\Omega, \mathcal{A})$;*
(ii) *the sequence $\{X_n\}$ is adapted to the filtration $\{\mathcal{A}_n\}$; and*
(iii) $\mathbf{E}^{\mathcal{A}_n}(X_{n+1}) = X_n$, $n \geq 1$.

5.1.6 Proposition. *If $\{X_n\}$ is a martingale, then*

$$\mathbf{E}^{\mathcal{A}_n}(X_{n+p}) = X_n, \quad \forall n \text{ and } \forall p > 0.$$

Proof. Since $\mathcal{A}_n \subset \mathcal{A}_{n+1} \subset \mathcal{A}_{n+2} \subset \ldots \subset \mathcal{A}_{n+p-1}$, it follows from 2.3(iii) that

$$\mathbf{E}^{\mathcal{A}_n} = \mathbf{E}^{\mathcal{A}_n}\mathbf{E}^{\mathcal{A}_{n+1}} \ldots \mathbf{E}^{\mathcal{A}_{n+p-1}}.$$

By 5.1.5(iii), $\mathbf{E}^{\mathcal{A}_{n+p-1}}(X_{n+p}) = X_{n+p-1}$, $\mathbf{E}^{\mathcal{A}_{n+p-2}}(X_{n+p-1}) = X_{n+p-2}, \ldots,$ and finally

$$\mathbf{E}^{\mathcal{A}_n}(X_{n+1}) = X_n. \square$$

5.2 Energy equality

5.2.1 Proposition. *Let $\{X_k\}$ be a martingale relative to the filtration $\{\mathcal{A}_k\}$, and assume that $X_k \in L^2(\Omega, \mathcal{A})$ $(1 \leq k \leq n)$. Then, for $n > p$,*

$$\mathbf{E}(X_n^2) - \mathbf{E}(X_p^2) = \sum_{j=p}^{n-1} \mathbf{E}((X_{j+1} - X_j)^2).$$

PROOF. Set $e_j = X_{j+1} - X_j$. Then, for $m \leq j$,

$$(i) \qquad \mathbf{E}^{\mathcal{A}_m}(e_j) = \mathbf{E}^{\mathcal{A}_m}\mathbf{E}^{\mathcal{A}_j}(X_{j+1} - X_j) = \mathbf{E}^{\mathcal{A}_m}(X_j - X_j) = 0.$$

Writing $X_n = X_p + \sum_{j=1}^{n-1} e_j$ and expanding X_n^2, we obtain

$$\mathbf{E}(X_n^2) = \mathbf{E}(X_p^2) + \sum_{j=p}^{n-1} \mathbf{E}(e_j^2) + \sum_{\substack{j,j' \\ j \neq j'}} \mathbf{E}(e_j e_{j'}) + 2\sum_{j=p}^{n-1} \mathbf{E}(X_p e_j).$$

We now show that all the terms appearing in the last two sums are zero. Assume that $j < j'$. By 2.1.3, $\mathbf{E}(e_j e_{j'}) = \mathbf{E}(\mathbf{E}^{\mathcal{A}_{j+1}}(e_j e_{j'}))$. We now use the fact that $\mathbf{E}^{\mathcal{A}_{j+1}}(e_j e_{j'}) = e_j \mathbf{E}^{\mathcal{A}_{j+1}}(e_{j'})$. Since $j < j'$, (i) implies that $\mathbf{E}^{\mathcal{A}_{j+1}}(e_{j'}) = 0$, whence $\mathbf{E}^{\mathcal{A}_{j+1}}(e_j e_{j'}) = 0$. Similarly, $\mathbf{E}(X_p e_j) = \mathbf{E}\mathbf{E}^{\mathcal{A}_j}(X_p e_j) = \mathbf{E}(X_p \mathbf{E}^{\mathcal{A}_j}(e_j)) = 0$. $\square$

5.2.2 Corollary. *Let $\{X_n\}$ be a martingale. Then $\mathbf{E}(X_n^2)$ is an increasing sequence.*

PROOF. Apply the energy equality with $p = n - 1$.

5.3 Theory of L^2 martingales

5.3.1 **Definition.** $\{X_k\}$ is called an L^2 martingale if

$$\sup \mathbf{E}(X_k^2) < +\infty.$$

It follows from 5.2.2 that

5.3.2 $\lim \mathbf{E}(X_k^2)$ exists and is finite.

5.3.3 Structure theorem. *Let $\{X_k\}$ be an L^2 martingale. Then there exists $X_\infty \in L^2(\Omega, \mathcal{A}_\infty)$ such that*

$$(i) \qquad \|X_k - X_\infty\|_{L^2} \to 0 \quad \textit{and}$$

$$(ii) \qquad X_k = \mathbf{E}^{\mathcal{A}_k}(X_\infty).$$

X_∞ is called the final value of the martingale.

Conversely, let $\ldots \mathcal{G}_n \subset \mathcal{G}_{n+1} \subset \ldots$ be an increasing sequence of σ-algebras on $(\Omega, \mathcal{A})$, let $\mathcal{G}_\infty$ be the σ-algebra generated by all the $\mathcal{G}_n$, let $f \in L^2(\Omega, \mathcal{G}_\infty, P)$, and let $Y_k = \mathbf{E}^{\mathcal{G}_k}(f)$. Then

$$(iii) \qquad Y_k \textit{ is an } L^2 \textit{ martingale}$$

and

$$(iv) \qquad \|Y_k - f\|_{L^2} \to 0.$$

PROOF. We first prove the following lemma.

5.3.4 Lemma. $\mathbf{E}((X_{n+p} - X_n)^2) = \mathbf{E}(X_{n+p}^2) - \mathbf{E}(X_n^2)$.

PROOF.

$$\mathbf{E}((X_{n+p} - X_n)^2) = \mathbf{E}(X_{n+p}^2) + \mathbf{E}(X_n^2) - 2\mathbf{E}(X_{n+p}X_n) \quad \text{and}$$

$$\mathbf{E}(X_{n+p}X_n) = \mathbf{E}\mathbf{E}^{\mathcal{A}_n}(X_{n+p}X_n) = \mathbf{E}(X_n\mathbf{E}^{\mathcal{A}_n}(X_{n+p})) = \mathbf{E}(X_n^2).\square$$

PROOF OF THE THEOREM. Since the sequence $\alpha_n = \mathbf{E}(X_n^2)$ is convergent by hypothesis, $\forall \epsilon > 0\ \exists n_0\ \forall p > 0\ \alpha_{n+p} - \alpha_n < \epsilon$. By 5.3.4,

$$\|X_{n+p} - X_n\|_{L^2}^2 < \epsilon, \quad \forall n \geq n_0 \text{ and } \forall p > 0.$$

Thus X_k is a Cauchy sequence, which converges since L^2 is complete. Moreover, $X_k = \mathbf{E}^{\mathcal{A}_k}(X_{k+r})$ for all $r > 0$ by 5.1.6.

Let $r \to +\infty$. Then $X_{k+r} \to X_\infty$ in L^2; hence $\mathbf{E}^{\mathcal{A}_k}(X_{k+r}) \to \mathbf{E}^{\mathcal{A}_k}(X_\infty)$, and (ii) follows.

We now prove the converse. By 2.3.3(iii),

$$\mathbf{E}^{\mathcal{G}_n}\mathbf{E}^{\mathcal{G}_{n+k}} = \mathbf{E}^{\mathcal{G}_n}.$$

Applying this to f, we obtain

$$\mathbf{E}^{\mathcal{G}_n}(\mathbf{E}^{\mathcal{G}_{n+k}}(f)) = \mathbf{E}^{\mathcal{G}_n}(f),$$

or

$$\mathbf{E}^{\mathcal{G}_n}(Y_{n+k}) = Y_n.$$

Furthermore, since the projection operator $\mathbf{E}^{\mathcal{G}_n}$ has norm ≤ 1,

$$\|Y_n\|_{L^2} = \|\mathbf{E}^{\mathcal{G}_n}(f)\|_{L^2} \leq \|f\|_{L^2}.$$

The sequence $\{Y_n\}$ is an L^2 martingale. By the first part of the theorem, $\exists f_\infty \in L^2$ such that

$$\|Y_n - f_\infty\|_{L^2} \to 0.$$

By the lemma, $f = f_\infty$, and this completes the proof of the theorem. □

5.3.5 Lemma. *Let $(\Omega, \mathcal{A}, P)$ be a probability space filtered by $\{\mathcal{A}_n\}$. Set $\mathcal{A}_\infty = \lim \mathcal{A}_n$ (in the sense of 5.1.1). Then*

$$\|\mathbf{E}^{\mathcal{A}_n} f - \mathbf{E}^{\mathcal{A}_\infty} f\|_{L^2} \to 0 \quad \textit{for every} \quad f \in L^2.$$

PROOF. We write V_n for $L^2(\Omega, \mathcal{A}_n)$, a closed subspace of $L^2(\Omega, \mathcal{A})$. Let $V_\infty = \overline{\cup_n V_n}$ and let Π_{V_n} and Π_{V_∞} denote the respective orthogonal projections. Then Hilbert space theory shows that

$$\|\Pi_{V_n} f - \Pi_{V_\infty} f\|_{L^2} \to 0.$$

Moreover, $L^2(\Omega, \mathcal{A}_\infty) \supset V_n$ for all n. Since V_∞ is the smallest closed subspace of L^2 that contains all the V_n, $L^2(\Omega, \mathcal{A}_\infty) \supset V_\infty$.

Now let $B \in \mathcal{A}_\infty$. If we show that $\mathbf{1}_B \in V_\infty$, the density of the simple functions in $L^2(\Omega, \mathcal{A}_\infty)$ will imply that $L^2(\Omega, \mathcal{A}_\infty) \subset V_\infty$ and hence that

$$(i) \qquad L^2(\Omega, \mathcal{A}_\infty) = V_\infty.$$

To prove this, let $\mathcal{B}$ denote the set of subsets B of Ω such that $B \in \mathcal{A}_s$ for some s. From the set-theoretic point of view in $\mathcal{P}(\Omega)$, $\mathcal{B} = \cup_s \mathcal{A}_s$.

Then $\mathcal{B}$ is a Boolean algebra and $\mathcal{B} \subset \mathcal{A}_\infty$. By I-1.4, $\mathcal{A}_\infty$ is the monotone class generated by $\mathcal{B}$.

If $B \in \mathcal{B}$, then $\mathbf{1}_B \in V_\infty$. Let $\mathcal{M}$ denote the class of subsets D of Ω with indicator function satisfying $\mathbf{1}_D \in V_\infty$.

We now show that $\mathcal{M}$ is a monotone class. Let D_n be an increasing sequence of elements of $\mathcal{M}$, with limit D_∞. Then $\mathbf{1}_{D_n} \to \mathbf{1}_{D_\infty}$ everywhere, and by the dominated convergence theorem $\|\mathbf{1}_{D_n} - \mathbf{1}_{D_\infty}\|_{L^2} \to 0$.

Since $P(\Omega) = 1$, the analogous result for decreasing sequences follows by taking complements. Thus $\mathcal{M}$ is a monotone class and $\mathcal{M} = \mathcal{A}_\infty$, so (i) is true and the lemma is proved. □

5.4 Stopping times and the maximal inequality

5.4.1 Definition of stopping time

Let $\{X_n\}$ be a martingale defined on the space $(\Omega, \mathcal{A}, P)$ filtered by $\{\mathcal{A}_n\}$.

A *stopping time* $T(\omega)$ is a function on Ω, with strictly positive integer values, such that

$$(i) \qquad A_{T,p} = \{\omega : T(\omega) > p\} \in \mathcal{A}_p \quad \forall p \in \mathbf{N}.$$

(ii) Proposition. *If T^1 and T^2 are two stopping times, then*

$$T^3(\omega) = \inf(T^1(\omega), T^2(\omega)) =_{\text{Def}} (T^1 \wedge T^2)(\omega)$$

is a stopping time.

PROOF. $A_{T^3,p} = A_{T^1,p} \cap A_{T^2,p} \in \mathcal{A}_p$.

(iii) Any given time q can be thought of as a stopping time.

5.4.2 Truncated martingales

Definition. Given a martingale $\{X_n\}$ and a stopping time T, the *truncated martingale* is defined by

$$X_n^T(\omega) = X_{T(\omega)\wedge n}(\omega).$$

We proceed to justify this terminology by showing that $\{X_n^T\}$ is a martingale. Since

$$(i) \qquad X_m^T = \sum_{j=1}^{m-1} (X_{j+1} - X_j)\mathbf{1}_{A_{T,j}} + X_1$$

and all the functions on the right-hand side are $\mathcal{A}_m$-measurable, X_m^T is $\mathcal{A}_m$-measurable. Moreover, $\mathbf{E}^{\mathcal{A}_n}(X_{n+1}^T)$ can be computed by observing that, on the right-hand side of (i), all the functions except X_{n+1} are $\mathcal{A}_n$-measurable. Thus

$$\mathbf{E}^{\mathcal{A}_n}(X_{n+1}^T) = (\mathbf{E}^{\mathcal{A}_n}(X_{n+1}) - X_n)\mathbf{1}_{A_{T,n}} + X_n^T.$$

But

$$\mathbf{E}^{\mathcal{A}_n}(X_{n+1}) = X_n,$$

whence

$$\mathbf{E}^{\mathcal{A}_n}(X_{n+1}^T) = X_n^T. \square$$

5.4.3 Definition of the maximal function

Let $\{Y_n\}$ be a martingale, let

$$Y_n^* = \sup_{1\le p<n} |Y_p|,$$

and let

$$Y^* = \lim_{n\to+\infty} Y_n^*.$$

Y^* is called the *maximal function* of the martingale $\{Y_n\}$.

5.4.4 Doob's maximal inequality

Proposition. *Let* $\{Y_n\}$ *be a martingale on the space* $(\Omega, \mathcal{A}, P\}$ *filtered by* $\{\mathcal{A}_n\}$. *Then, for every constant* $\gamma > 0$,

$$(i) \qquad P(Y_n^* \geq \gamma) \leq \frac{2}{\gamma}[\mathbf{E}(|Y_n|) + \mathbf{E}(|Y_1|)] \quad (n = 0, 1, \ldots)$$

and

$$(ii) \qquad P(Y^* \geq \gamma) \leq \frac{4}{\gamma} \sup_n \mathbf{E}(|Y_n|).$$

PROOF. Let $A_\gamma^n = \{\omega : \sup_{p<n} Y_p(\omega) \geq \gamma\}$ and let

$$\begin{array}{rcll} T(\omega) & = & \inf\{p : Y_p(\omega) \geq \gamma\} & \text{if} \quad \omega \in A_\gamma^n \\ T(\omega) & = & n & \text{if} \quad \omega \notin A_\gamma^n. \end{array}$$

Then $T(\omega) \leq n$. Moreover,

$$\{\omega : T(\omega) > q\} = \cup_{p \leq q}\{\omega : Y_p(\omega) < \gamma\} \in \mathcal{A}_q \quad \text{if} \quad q < n$$

and

$$\{\omega : T(\omega) > n\} = \emptyset.$$

Thus T is a stopping time; let $\{Y_k^T\}$ be the martingale truncated by T. Then $\mathbf{E}(Y_1^T) = \mathbf{E}(Y_1)$ (since $Y_1^T = Y_1$) and

$$\mathbf{E}(Y_n^T) = \mathbf{E}(Y_n^T \ \mathbf{1}_{T<n}) + \mathbf{E}(Y_n^T \ \mathbf{1}_{T=n}).$$

$Y_n^T \geq \gamma$ on the event $\{\omega : T(\omega) < n\}$ and $Y_n^T = Y_n$ on $\{\omega : T(\omega) = n\}$; hence

$$\mathbf{E}(Y_n^T \ \mathbf{1}_{T<n}) \geq \gamma \mathbf{E}(\mathbf{1}_{T<n}) = \gamma P(A_\gamma^n).$$

Thus $\mathbf{E}(Y_1) - \mathbf{E}(Y_n \mathbf{1}_{T=n}) \geq \gamma P(A_\gamma^n)$, and

$$\gamma P(A_\gamma^n) \leq \mathbf{E}(|Y_n|) + \mathbf{E}(|Y_1|).$$

(i) follows by observing that

$$\{\omega : Y_n^*(\omega) \geq \gamma\} = A_\gamma^n \cup \left\{\omega : \sup_{p<n}[-Y_p] \geq \gamma\right\}.$$

To prove (ii), it suffices to note that $\{Y_n^*\}$ is an increasing sequence with limit Y^*. Hence $Y^* \geq \gamma \Rightarrow \forall \epsilon > 0 \ \exists n$ such that $P(Y_n \geq \gamma - \epsilon)$. Thus, by (i),

$$P(Y^* \geq \gamma) \leq \frac{4}{\gamma - \epsilon} \sup \mathbf{E}(|Y_n|).\square$$

5.5 Convergence of regular martingales

5.5.1 Definition of regular martingales

Let $\{Y_n\}$ be a martingale relative to the filtration $\mathcal{A}_n$ on $(\Omega, \mathcal{A}, P)$. $\{Y_n\}$ is called *regular* if there exists $Z \in L^1(\Omega, \mathcal{A}, P)$ such that

$$Y_n = \mathbf{E}^{\mathcal{A}_n}(Z) \quad \forall n.$$

EXAMPLE. Every L^2 martingale is regular by 5.3.3.

5.5.2 Final value of a regular martingale

Let $\mathcal{A}_\infty$ denote the σ-algebra generated by the union of the σ-algebras $\mathcal{A}_n$, and set

$$Y_\infty = \mathbf{E}^{\mathcal{A}_\infty}(Z).$$

Y_∞ is called the *final value* of the martingale $\{Y_n\}$.

5.5.3 Theorem on L^1 convergence. *Let $\{Y_n\}$ be a regular martingale and let Y_∞ be its final value. Then*

$$(i) \qquad Y_n = \mathbf{E}^{\mathcal{A}_n}(Y_\infty) \quad \textit{and}$$

$$(ii) \qquad \mathbf{E}(|Y_n - Y_\infty|) \to 0 \quad \textit{as} \quad n \to +\infty.$$

PROOF. Let $\varphi_M(t)$ be the function introduced in I-6.7 to define the truncation operator and set $Z^M = \varphi_M(Z)$. Then

$$(iii) \qquad \|Z^M - Z\|_{L^1} \to 0 \quad \text{as} \quad M \to 0.$$

Set $Y_{n,M} = \mathbf{E}^{\mathcal{A}_n}(Z^M)$. Then

$$\|Y_{n,M} - Y_n\|_{L^1} \leq \|Z^M - Z\|_{L^1}.$$

Thus $\|Y_{n,M}\|_{L^\infty} \leq M$, and hence $\{Y_{n,M}\}$ is a martingale. Using 5.3.3 and 5.3.5 and setting $Y_{\infty,M} = \mathbf{E}^{\mathcal{A}_\infty}(Z^M)$, we obtain

$$(iv) \qquad Y_{n,M} = \mathbf{E}^{\mathcal{A}_n}(Y_{\infty,M}) \quad \text{and} \quad \|Y_{n,M} - Y_{\infty,M}\|_{L^2} \to 0.$$

(i) is proved by using (iii) and the first formula of (iv), then letting M tend to infinity. Similarly, since the L^2 convergence in (iv) implies L^1 convergence by the Cauchy-Schwarz inequality, (ii) follows for $Y_{n,M}$. Letting $M \to \infty$ shows that (ii) holds for Y_n. □

5.5.4 Proposition (Almost sure convergence). *Let $\{Y_n\}$ be a regular martingale and let Y_∞ be its final value. Then*

$$Y_n(\omega) \to Y_\infty(\omega) \quad \textit{almost surely.}$$

PROOF. Let $\beta_q(\omega) = \sup_{n>q} |Y_n(\omega) - Y_q(\omega)|$ and let $\beta(\omega) = \lim_{q\to\infty} \beta_q(\omega)$.
For fixed q, let

$$Z_m = Y_{q+m} - Y_q \quad (m \geq 0).$$

Then $\{Z_m\}$ is a regular martingale relative to the filtration $\{\mathcal{A}_{q+m}\}$ and

$$\sup_m \|Z_m\|_{L^1} \leq \|Y_\infty - Y_q\|_{L^1} + \sup_{q'>q} \|Y_\infty - Y_{q'}\|_{L^1}.$$

By 5.5.3(ii), the right-hand side is less than ϵ if $q \geq q_0$. Hence, using the maximal inequality 5.4.4(ii),

$$P(\{\omega : \beta_q(\omega) > \gamma\}) < \frac{2\epsilon}{\gamma} \quad \text{if} \quad q \geq q_0.$$

Fixing γ, let $q \to \infty$. Since $\{\beta_q\}$ is a decreasing sequence of functions,

$$P(\{\omega : \beta(\omega) > \gamma\}) = 0, \quad \text{whence} \quad \beta(\omega) = 0 \text{ a.s.}$$

$\{Y_n(\omega)\}$ converges a.s. Let Z_∞ be its limit. Since $\{Y_n\}$ converges in L^1 to Y_∞, it has a subsequence $\{Y_{n_k}\}$ that converges a.s. to Y_∞; hence $Z_\infty = Y_\infty$. $\square$

5.6 L^1 martingales

5.6.1 **Definition.** A martingale $\{Y_n\}$ is called an L^1 martingale if

$$\sup_n \|Y_n\|_{L^1} < +\infty.$$

EXAMPLE. Every regular martingale is an L^1 martingale.

5.6.2 Proposition. *Let $\{Y_n\}$ be an L^1 martingale. Let $T_1 \leq T_2 \leq \ldots \leq T_j \leq \ldots$ be an increasing sequence of stopping times such that, for every j, $T_j(\omega) < +\infty$ a.s. Let $Y_{T_j}(\omega) = Y_{T_j(\omega)}(\omega)$.*
Then

$$\sum_{j=1}^{\infty} (Y_{T_{j+1}}(\omega) - Y_{T_j}(\omega))^2 < +\infty \quad a.s.$$

PROOF. Set $a = \sup \|Y_n\|_{L^1}$ and let Y^* be the maximal function. Then, by 5.4.4, $P(Y^* \geq p) \leq 4ap^{-1}$, whence

$$(i) \qquad Y^*(\omega) < +\infty \quad a.s.$$

Fix p and let f be the continuously differentiable convex function defined by $f(t) = t^2$ if $|t| \leq p$ and $f(t) = 2p|t| - p^2$ if $|t| \geq p$. Let g be the nonnegative function defined by

$$g(v_1, v_2) = f(v_2) - f(v_1) - (v_2 - v_1)f'(v_1).$$

Then $g(v_1, v_2) = (v_2 - v_1)^2$ if $|v_i| \leq p$ $(i = 1, 2)$.

$$(ii) \qquad \mathbf{E}[f(Y_n)] \leq 2p\mathbf{E}(|Y_n|) \leq 2pa$$

$$(iii) \qquad \mathbf{E}[(Y_{j+1}-Y_j)f'(Y_j)] = \mathbf{E}\mathbf{E}^{\mathcal{A}_j} = \mathbf{E}[f'(Y_j)\mathbf{E}^{\mathcal{A}_j}(Y_{j+1}-Y_j)] = 0;$$

$$\mathbf{E}(f(Y_n)-f(Y_1)) = \mathbf{E}\left[\sum_{j=1}^{n-1}(f(Y_{j+1})-f(Y_j))\right] = \sum_{j=1}^{n-1}\mathbf{E}(g(Y_{j+1},Y_j));$$

$$\mathbf{E}\left(\sum_{j=1}^{+\infty}(Y_{j+1}-Y_j)^2\mathbf{1}_{\{Y^*\leq p\}}\right) \leq 2pa.$$

Hence, letting $p\to+\infty$ and using (i),

$$(iv) \qquad \sum_{j=1}^{+\infty}(Y_{j+1}(\omega)-Y_j(\omega))^2 < +\infty \text{ a.s.}$$

We now generalize this "local version" of 5.2.1 to an increasing sequence of stopping times $T_1 \leq T_2 \leq \ldots \leq T_j \leq \ldots$. Set $Y_{T_j}(\omega) = Y_{T_j(\omega)}(\omega)$. We would like to show that

$$(v) \qquad \sum_{j=1}^{+\infty}(Y_{T_{j+1}}(\omega)-Y_{T_j}(\omega))^2 < +\infty \text{ a.s.}$$

Once (ii) and (iii) have been generalized, the same calculation will give (v). Letting $A_{T_{j,q}} = \{\omega : T_j > q\}$, we have, as in 5.4.2,

$$f(Y_{T_j})-f(Y_1) = \sum_{q=1}^{\infty}(f(Y_{q+1})-f(Y_q))\mathbf{1}_{A_{T_{j,q}}}.$$

$$\mathbf{E}(f(Y_{T_j})-f(Y_1)) = \sum_{q=1}^{+\infty}\mathbf{E}(g(Y_{q+1},Y_q)\mathbf{1}_{A_{T_{j,q}}}) \leq \sum_q g(Y_{q+1},Y_q) \leq 2pa,$$

whence

$$(ii)' \qquad \mathbf{E}(f(Y_{T_j})) \leq 4pa.$$

Let $\mathcal{A}_{T_j}$ denote the σ-algebra generated by the $\mathcal{A}_q \cap T_j^{-1}(q)$, where $q\in\mathbf{N}$. Then

$$(iii)' \qquad \mathbf{E}^{\mathcal{A}_{T_j}}((Y_{T_{j-1}}-Y_{T_j})f'(Y_{T_j})) = 0.$$

This proves (v). □

5.6.3 Fatou's theorem. *Let Y_n be an L^1 martingale. Then* $\lim_{n\to\infty} Y_n(\omega)$ *exists a.s.*

PROOF. For a proof by contradiction, assume that Fatou's theorem fails; then there exists $b>0$ such that

$$(vi) \qquad G = \left\{\omega : \limsup_{n,n'\to\infty}|Y_n(\omega)-Y_{n'}(\omega)| > 2b\right\} \quad \text{satisfies} \quad P(G)>0.$$

Let $T_1(\omega) = 1$ and let the later stopping times be defined recursively by

$$T_{j+1}(\omega) = \inf\{q : q > T_j(\omega) \text{ and } |Y_{T_j}(\omega) - Y_q(\omega)| > b\}.$$

Since the sequence $T_j(\omega)$ is increasing, $|Y_{T_{j+1}}(\omega) - Y_{T_j}(\omega)| > b$. This contradicts 5.6.2. □

IMPORTANT REMARK. Nonzero L^1 martingales can be constructed with $\lim Y_n(\omega) = 0$ a.s. It is thus impossible to reconstruct the martingale from this limit, as was done for regular martingales. Hence the importance of the regularity criterion that will be given in Section 5.8. In Section 5.7, we will develop a concept that is both interesting for its own sake and crucial for stating the regularity criterion.

5.7 Uniformly integrable sets

5.7.1 **Definition.** A subset H of L^1 is called *uniformly integrable* if for every $\epsilon > 0$ there exists $\eta > 0$ such that $\mathbf{E}(|h|\ \mathbf{1}_A) < \epsilon$ for all $h \in H$ and for every $A \in \mathcal{A}$ with $P(A) \leq \eta$.

5.7.2 Proposition. *Let H be a subset of L^1. Then the following two statements are equivalent:*

(i) H is uniformly integrable.

(ii)
$$\lim_{q \to \infty} \left[\sup_{h \in H} \int_{|h|>q} |h|\ dP \right] = 0.$$

PROOF. To prove that (i) $\Rightarrow$ (ii), we first show that (i) implies

(iii)
$$\exists M < +\infty \quad \text{such that} \quad \|h\|_{L^1} < M \ \forall h \in H.$$

Let $\eta > 0$ be the number associated with $\epsilon = 1$ by Definition 5.7.1. Then (iii) follows from setting $M = \frac{1}{\eta} + 1$.

By Chebyshev, $P(|h| > q) \leq q^{-1}M$. Since this expression tends to zero as $q \to \infty$, (i) implies (ii) formally.

We now prove that (ii) $\Rightarrow$ (i). For a proof by contradiction, suppose that there exist $\epsilon_0 > 0$ and sequences $\{h_n\}$ in H and $\{A_n\}$ in $\mathcal{A}$ such that

$$\mathbf{E}(h_n \mathbf{1}_{A_n}) > \epsilon_0 \quad \text{and} \quad P(A_n) \to 0.$$

Let q_0 be chosen so that

$$\int_{|h|>q_0} h\ dP < \frac{\epsilon_0}{2} \quad \forall h \in H.$$

Set $B_n = \{\omega : |h_n(\omega)| > q_0\}$. Then

$$\epsilon_0 < \mathbf{E}(h_n \mathbf{1}_{A_n}(\mathbf{1}_{B_n} + \mathbf{1}_{B_n^0})) \leq \mathbf{E}(h_n \mathbf{1}_{B_n}) + q_0 P(A_n).$$

Since the first term on the right-hand side is less than $\frac{\epsilon_0}{2}$ and the second tends to zero as $n \to \infty$, this gives a contradiction. □

5.7.3 Proposition. *Let H be a uniformly integrable subset of L^1 and let H_1 be the closure of H in the topology of almost sure convergence. Then H_1 is uniformly integrable.*

PROOF. Set

$$\varphi(\epsilon) = \sup \mathbf{E}(|h| \; \mathbf{1}_A), \quad \text{where} \quad h \in H \quad \text{and} \quad P(A) < \epsilon.$$

Let $\{h_n\}$ be a sequence of elements of H which converges almost surely to h_0. Fatou's lemma implies that

$$\varphi(\epsilon) \geq \liminf \mathbf{E}(|h| \; \mathbf{1}_A) \geq \mathbf{E}(|h_0| \; \mathbf{1}_A),$$

whence

$$\sup \mathbf{E}(|h| \; \mathbf{1}_A) \leq \varphi(\epsilon), \quad \forall h_1 \in H_1, \forall A \text{ such that } P(A) < \epsilon. \square$$

5.7.4 Theorem (Generalization of Lebesgue's dominated convergence theorem). *Let $\{u_n\}$ be a sequence of integrable functions on a measure space $(X, \mathcal{A}, \mu)$, $\mu(X) < +\infty$, such that*

(i) the family $\{u_n\}$ is uniformly integrable and
(ii) u_n converges a.s. to u_0.

Then

$$\|u_n - u_0\|_{L^1} \to 0.$$

PROOF. By Egoroff's theorem, there exist $\epsilon > 0$ and $B \in \mathcal{A}$ such that $\mu(B^c) < \epsilon$ and u_n converges uniformly to u_0 on B.

Then

$$\|u_n - u_0\|_{L^1} \leq \mathbf{E}(|u_n - u_0| \; \mathbf{1}_B) + \mathbf{E}(|u_n| \; \mathbf{1}_{B^c}) + \mathbf{E}(\|u_0\| \; \mathbf{1}_{B^c}).$$

The first term on the right-hand side tends to zero by uniform convergence, the second by uniform integrability, and the third by the same reasoning as in 5.7.3. □

5.8 Regularity criterion

5.8.1 Theorem. *Let $\{X_n\}$ be an L^1 martingale. Then the following conditions are equivalent:*

(i) $\{X_n\}$ is regular.
(ii) $\{X_n : 1 \leq n \leq \infty\}$ is uniformly integrable.

PROOF THAT (ii) $\Rightarrow$ (i). We know by Fatou's theorem (5.6.2) that $X_q(\omega)$ converges almost surely to Z. By 5.7.4, this implies that $\|X_q - Z\|_{L^1} \to 0$. Hence, using the identity

$$X_n = \mathbf{E}^{\mathcal{A}_n}(X_q) \quad q > n,$$

fixing n, and letting q go to infinity,

$$X_n = \mathbf{E}^{\mathcal{A}_n}(Z). \square$$

PROOF THAT (i) $\Rightarrow$ (ii). For a c to be fixed later, set $B_n = \{\omega : |X_n(\omega)| > c\}$. Then, since $B_n \in \mathcal{A}_n$ and $|X_n| \leq \mathbf{E}^{\mathcal{A}_n}(|Z|)$,

$$\mathbf{E}(|X_n| \; \mathbf{1}_{B_n}) \leq \mathbf{E}((\mathbf{1}_{B_n}\mathbf{E}^{\mathcal{A}_n}(|Z|)) = \mathbf{E}(\mathbf{E}^{\mathcal{A}_n}(|Z|)\mathbf{1}_{B_n})) = \mathbf{E}(|Z| \; \mathbf{1}_{B_n}).$$

Hence, with b also to be fixed later,

$$\begin{aligned}
\int_{|X_n|>c} |X_n| \, dP &\leq \int_{|X_n|>c} |Z| \, dP \\
&= \int_{\{|X_n|>c\}\cap\{|Z|>b\}} |Z| \, dP + \int_{\{|X_n|>c\}\cap\{|Z|<b\}} |Z| \, dP \\
&\leq \int_{|Z|>b} |Z| \, dP + bP(|X_n| > c).
\end{aligned}$$

But, by Chebyshev's inequality, $P(|X_n| > c) \leq \frac{1}{c}\mathbf{E}(|X_n|) \leq \frac{1}{c}\mathbf{E}(|Z|)$, whence

$$\int_{|X_n|>c} |X_n| \, dP \leq \int_{|Z|>b} |Z| \, dP + \frac{b}{c}\mathbf{E}(|Z|).$$

Let $b = q^{1/2}$ and $c = q$; then the right-hand side tends to zero as $q \to \infty$, and the conclusion follows by 5.7.2. $\square$

6 Theory of Differentiation

If f is a *continuous* function defined on $[0,1] \subset \mathbf{R}$ and $F(x) = \int_0^x f$, then F is differentiable for every x and $F'(x) = f(x)$. The same result holds for $f \in L^1$, provided that "for every x" is replaced by "almost everywhere"; this is another theorem of Lebesgue.

The derivative is computed as the limit of quotients of the form

$$\text{(i)} \qquad \frac{1}{\epsilon}[F(x+\epsilon) - F(x)] = \frac{1}{\nu(A_c)}\rho(A_\epsilon),$$

where ν is Lebesgue measure, $\rho(A) = \int_A f$, and $A_\epsilon = [x, x+\epsilon]$.

In this section, we study the limits of quotients of the form (i) on an abstract measure space. A.s. convergence will be obtained for an appropriate choice of the A_ϵ: the A_ϵ will be the atoms of an increasing sequence $\mathcal{A}_n$ of finite sub-σ-algebras of $\mathcal{A}$, "converging to $\mathcal{A}$".

Quotients of the form (i), which thus form a martingale for the filtration $\mathcal{A}_n$, will be used to prove the *Radon-Nikodym theorem.*

Conditional probabilities can immediately be defined for conditionings by finite σ-algebras; the existence of conditional distributions in the general case will depend essentially on a convergence theorem for vector-valued martingales. The convergence of such martingales will be clear for Radon measures. A *structure theorem* will allow all separable measure spaces to be realized by means of Lebesgue measure on $\mathbf{R}$.

6.0 Separability

The measure space $(X, \mathcal{A}, \mu)$ is called *separable* if there exists a sequence that is dense in L^1_μ; in other words, if L^1_μ satisfies the first separability axiom I-2.4.1(i).

Consider the case of Radon measures on a compact space Y. If Y is metrizable, then $C(Y)$ satisfies the first separability axiom and, since $C(Y)$ is dense in L^1_μ, the same holds for L^1_μ. The same result is true if Y is locally compact, metrizable, and the countable union of compact sets.

6.1 Separability and approximation by finite σ-algebras

Proposition. *A measure space $(\Omega, \mathcal{A}, P)$ is separable if and only if there exists an increasing sequence of σ-algebras $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \ldots \subset \mathcal{A}_n \ldots$ such that each σ-algebra $\mathcal{A}_n$ is finite and*

$$\mathbf{E}^{\mathcal{A}_n}(f) \to f \quad \textit{for every} \quad f \in L^1(\Omega, \mathcal{A}, P).$$

The sequence of σ-algebras $\mathcal{A}_n$ is said to P-generate $\mathcal{A}$.

PROOF. Assume that $(\Omega, \mathcal{A}, P)$ is separable, and let $f_1, \ldots, f_n, \ldots$ be a dense sequence in L^1. Approximating each f_n by a sequence of simple functions gives a countable family Φ of simple functions which is dense in L^1_P.

Let $g_1, \ldots, g_n, \ldots$ be an enumeration of this sequence and let $\mathcal{A}_i$ be the σ-algebra generated by $g_k^{-1}(\mathcal{A})$, $1 \le k \le i$.

With each $f \in L^1$ we now associate a sequence $\{X_k\}$ defined by

$$\mathbf{E}^{\mathcal{A}_k} f = X_k.$$

Then $\{X_k\}$ is a regular martingale, which converges in L^1 by 5.5.3. Let $X_\infty = \lim X_k$. Since $\|X_\infty\|_{L^1} \le \|f\|_{L^1}$, a bounded operator $\pi : L^1 \to L^1$ can be defined by setting $\pi(f) = X_\infty$, and

$$\mathbf{E}^{\mathcal{A}_k} X_\infty = \lim_{q \to \infty} \mathbf{E}^{\mathcal{A}_k} \mathbf{E}^{\mathcal{A}_q} f = \mathbf{E}^{\mathcal{A}_k} f = X_k.$$

That is, $\pi^2 = \pi$. The image V of π is closed, for if $u_n = \pi(f_n)$ and $u_n \to u_0$, then since $\pi u_n = u_n$ and π is continuous it follows that $\pi u_0 = u_0$. V is thus a closed vector subspace of $L^1(X, \mathcal{A}, \mu)$.

Let f_0 be a simple function in Φ. Then $f_0 \in L^0(\mathcal{A}_k)$ for k sufficiently large, and hence $\mathbf{E}^{\mathcal{A}_k}(f_0) = f_0$. Thus $f_0 \in V$. Since the family Φ of simple functions is dense in L^1, it follows that $V = L^1$.

The proof in the other direction is clear. For each k, $\mathbf{E}^{\mathcal{A}_k}(L^1)$ is a finite-dimensional subspace of L^1 and hence separable. The union of these spaces is separable and dense in L^1. □

6.2 The Radon-Nikodym theorem

6.2.1 Theorem. *Let $(\Omega, \mathcal{A}, \mu)$ be a separable measure space and let μ and ν be finite measures defined on $\mathcal{A}$. Then the following statements are equivalent:*

(i) For every $A \in \mathcal{A}$, $\mu(A) = 0 \Rightarrow \nu(A) = 0$.
(ii) There exists $k \in L^1$, $k \geq 0$, such that $\nu(A) = \int_A k d\mu$.

REMARK. The function k is called the *density* of ν with respect to μ and is sometimes written $k = \frac{d\nu}{d\mu}$.

PROOF. It is trivial that (ii) $\Rightarrow$ (i). Indeeed, if $k\mathbf{1}_A$ is a function that is zero a.e., then its integral is zero. To prove that (i) $\Rightarrow$ (ii), assume that

$$\mu(X) < +\infty. \tag{iii}$$

This hypothesis can easily be dropped later, by taking an exhaustion sequence $\{A_n\}$ for X.

Multiplying by a constant reduces the proof to the case where

(iv) $\mu(X) = 1$ and $(X, \mathcal{A}, \mu)$ will be considered as a probability space.

We now prove that hypothesis (i) implies the following *quantitative version.*

6.2.2 Lemma. *Assume that 6.2.1(i) holds. Then, for every $\epsilon > 0$, there exists $\delta > 0$ such that*

$$\mu(A) < \delta \quad \textit{implies} \quad \nu(A) < \epsilon.$$

PROOF. Otherwise there would exist ϵ_0 and A_k such that

$$\mu(A_k) < 2^{-k} \quad \text{and} \quad \nu(A_k) \geq \epsilon_0.$$

Set $G_n = \cup_{k>n} A_k$; then $\mu(G_n) < 2^{-n+1}$. Since G_n is a decreasing sequence,

$$\mu\left(\cap_n G_n\right) = 0. \tag{i}$$

Furthermore, since $\nu(G_n) \geq \nu(A_n) \geq \epsilon_0$,

$$(ii) \qquad \nu\left(\cap_n G_n\right) = \lim \nu(G_n) \geq \epsilon_0.$$

But (i) and (ii) contradict 6.2.1(i). □

6.2.3 Associated martingales

Let $\mathcal{A}_1, \ldots, \mathcal{A}_n, \ldots$ be the increasing sequence of finite σ-algebras constructed in 6.1.

Let $\mathcal{E}_p = \{e_1, \ldots, e_s\}$ be the atoms of $\mathcal{A}_p$, and let a function $Y_p \in L^0(\mathcal{A}_p)$ be defined by setting

$$\begin{aligned} Y_p(e_r) &= \frac{\nu(e_r)}{\mu(e_r)} && \text{if } e_r \in \mathcal{E}_p \text{ and } \mu(e_r) \neq 0, \\ Y_p(e_r) &= 0 && \text{otherwise.} \end{aligned}$$

Then $\mathbf{E}(Y_p) = \sum' Y_p(e_r)\mu(e_k)$, where the sum $\sum'$ is restricted to those atoms such that $\mu(e_r) \neq 0$. Since $\mu(e_r) = 0 \Rightarrow \nu(e_r) = 0$, it follows that $\mathbf{E}(Y_p) = \sum \nu(e_r) = \nu(X)$.

More generally, let $\mathcal{A}_{p+1}$ be the σ-algebra following $\mathcal{A}_p$. An atom e_r of $\mathcal{A}_p$ can be decomposed into atoms of $\mathcal{A}_{p+1}$: $e_r = g_{r,1} \cup g_{r,2} \cup \ldots \cup g_{r,s}$.

Since the function Y_{p+1} is constant on each atom g,

$$\mathbf{E}^{\mathcal{A}_p}(Y_{p+1})(e_r) = \frac{1}{\mu(e_r)} \sum_{j=1}^{s} Y_{p+1}(g_{r,j})\mu(g_{r,j}).$$

But $Y_{p+1}(g_{r,j})\mu(g_{r,j}) = \nu(g_{r,,j})$ by the definition of Y_{p+1}. Since $\sum \nu(g_{r,j}) = \nu(e_r)$,

$$\mathbf{E}^{\mathcal{A}_p}(Y_{p+1}) = Y_p,$$

and we have proved the following result:

The Y_p form a martingale.

6.2.4 Lemma. *The martingale $\{Y_p\}$ constructed in 6.2.3 satisfies the uniform integrability condition.*

PROOF. Let p be fixed. Given $\epsilon > 0$, we must show that there exists η such that

$$(i) \qquad \int_A Y_p d\mu = \mathbf{E}(Y_p \, \mathbf{1}_A) < \epsilon \quad \text{for any } A \in \mathcal{A} \text{ such that } \mu(A) < \eta.$$

By 2.6.1(iii),

$$\mathbf{E}(Y_p \, \mathbf{1}_A) = \mathbf{E}(\mathbf{E}^{\mathcal{A}_p}(Y_p \, \mathbf{1}_A)) = \mathbf{E}(Y_p \mathbf{E}^{\mathcal{A}_p}(\mathbf{1}_A)).$$

Set $\varphi = \mathbf{E}^{\mathcal{A}_p}(\mathbf{1}_A)$; then $0 \leq \varphi \leq 1$ and $\mathbf{E}(\varphi) = \mu(A)$.

Introducing the atoms e_r of $\mathcal{A}_p$, we have 2.4.1. Since φ is $\mathcal{A}_p$-measurable, φ is constant on each atom e_r of $\mathcal{A}_p$; thus

$$\mathbf{E}(Y_p\varphi) = \sum \varphi(e_r)\nu(e_r) \quad \text{and} \quad \mathbf{E}(\varphi) = \sum \varphi(e_r)\mu(e_r).$$

Define a partition Φ_s of $\mathcal{E}_p$ by $e_r \in \Phi_s$ if $\varphi(e_r) \in [2^{-s-1}, 2^{-s}]$. Then

$$\mathbf{E}(Y_p\varphi) \leq \sum 2^{-s}\nu(H_s), \quad \text{where} \quad H_s = \bigcup_{e_r\in\Phi_s} e_r,$$

and

$$\mu(A) = \mathbf{E}(\varphi) \geq \sum 2^{-s-1}\mu(H_s).$$

Let s_0 be chosen so that $2^{-s_0+1}\nu(X) < \frac{\epsilon}{2}$. Then

$$\mathbf{E}(Y_p\varphi) < \sum_{0\leq s<s_0} 2^{-s}\nu(H_s) + \frac{\epsilon}{2}.$$

Let η' be the number associated by Lemma 6.2.2 with $\epsilon' = \frac{\epsilon}{4}$, and let $\eta = 2^{-s_0-1}\eta'$. Then, if $\mu(A) < \eta$, we have $\mu(H_s) < \eta'$ for $0 \leq s < s_0$. It follows from 6.2.2 that $\nu(H_s) < \epsilon'$, and thus

$$\mathbf{E}(Y_p\varphi) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This proves (i). □

6.2.5 Proof of the Radon-Nikodym theorem

Since $\{Y_p\}$ is an L^1 martingale and is uniformly integrable, there exists a function $k \in L^1_\mu$ such that $Y_p = \mathbf{E}^{\mathcal{A}_p}(k)$.

We begin by showing that

$$\nu(A) = \mathbf{E}(k\ \mathbf{1}_A).$$

By the construction of $\{Y_p\}$,

$$\int \psi d\nu = \mathbf{E}(Y_p\psi) \quad \text{if} \quad \psi \in \mathcal{L}^0(\mathcal{A}_p).$$

In particular,

$$\mathbf{E}(\mathbf{E}^{\mathcal{A}_p} k \mathbf{E}^{\mathcal{A}_p}\ \mathbf{1}_A) = \int (\mathbf{E}^{\mathcal{A}_p}\ \mathbf{1}_A) d\nu.$$

Set $\varphi_p = \mathbf{E}^{\mathcal{A}_p}(\mathbf{1}_A)$; then $0 \leq \varphi_p \leq 1$. The martingale $\mathbf{E}^{\mathcal{A}_p}(\mathbf{1}_A)$ converges μ-a.e. to $\mathbf{1}_A$ by 5.5.4, and convergence μ-a.e. implies convergence ν-a.e. by 6.2(i). Hence, by Lebesgue's dominated convergence theorem,

$$\int \mathbf{E}^{\mathcal{A}_p}(\mathbf{1}_A) d\nu \to \nu(A).$$

Since $\{Y_p\}$ is uniformly integrable, so is $\{Y_p\varphi_p\}$. Moreover, $Y_p\varphi_p$ converges μ-a.e. to $k\mathbf{1}_A$, and Theorem 5.7.4 implies that

$$\int Y_p\varphi_p d\mu \to \int_A k\ d\mu.$$

Thus

$$\int_A k\ d\mu = \nu(A).\square$$

6.3 Duality of the L^p spaces

Theorem. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let $1 \leq p < +\infty$, and let q be the conjugate exponent to p. Then the space of continuous linear functionals on L^p can be identified with L^q. As in I-9.4.3, the dual pairing is written*

$$\langle f, g\rangle = \int fg\ d\mu, \quad \textit{where} \quad f \in L^p,\ g \in L^q.$$

PROOF. Using an exhaustion sequence $\{A_n\}$ of X, we can reduce the proof to the case where $(\Omega, \mathcal{A}, \mu)$ is a probability space.

A *positive linear functional* l on L^p is a linear functional such that $l(f) \geq 0$ for every $f \geq 0$, $f \in L^p$. As in II-5, it can be shown that every linear functional on L^p can be written as the difference of two positive linear functionals. It thus suffices to prove the theorem when l is positive.

Since $\mu(X) < +\infty$, $L^\infty \subset L^p$ and we can define

$$\nu(A) = l(\mathbf{1}_A) \geq 0.$$

Let $C_n = \cup_{1\leq i\leq n} A_i$ and let $C_\infty = \cup_{i=1}^\infty A_i$. Since the A_i can be assumed to be disjoint, $\sum_{1\leq i\leq n} \mathbf{1}_{A_i} = \mathbf{1}_{C_n}$. Thus

$$\nu(C_n) = \sum l(A_i) = \sum \nu(A_i).$$

$\mathbf{1}_{C_n} \to \mathbf{1}_{C_\infty}$ everywhere and $\mathbf{1}_{C_n} \leq 1$; hence, by Lebesgue's dominated convergence theorem, $\|\mathbf{1}_{C_n} - \mathbf{1}_{C_\infty}\|_{L^p} \to 0$. It follows that ν is a *measure defined on* $\mathcal{A}$. Furthermore, $\mu(A) = 0$ implies $\mathbf{1}_A = 0$ in L^1_μ, whence $\nu(A) = l(\mathbf{1}_A) = 0$. Thus hypothesis 6.2.1(i) is satisfied, and the Radon-Nikodym theorem implies the existence of a nonnegative $k \in L^1_\mu$ such that

$$\int_A k\ d\mu = \nu(A).$$

Using linear combinations of characteristic functions, we see that

$$(i) \qquad l(\varphi) = \int k\varphi\ d\mu$$

for all simple functions. If we show that

$$(ii) \qquad k \in L^q,$$

each side of (i) will define a continuous linear functional on L^p; since they coincide on the dense set of simple functions, they will be equal everywhere.

Let $B_n = \{x : k(x) < n\}$ and let $k_n = k\mathbf{1}_{B_n}$; then $k_n \in L^\infty$. If $q < +\infty$, by Fatou-Beppo Levi the negation of (ii) is equivalent to the assertion that $\|k_n\|_{L^q} \to \infty$ as $n \to \infty$.

Let $\alpha = p^{-1}q$ and let

$$u_n = \frac{1}{\|k_n\|_{L^q}^\alpha} k_n^\alpha.$$

Then $\|u_n\|_{L^p} = 1$ and

$$l(u_n) = \int u_n k = \int u_n k_n = \|k_n\|_{L^q} \to \infty \quad \text{as } n \to \infty,$$

contradicting the inequality

$$|l(u_n)| \leq \|l\| \; \|u_n\|_{L^p}.$$

The case $q = \infty$ is treated in the same way, using the inequality

$$l(1 - \mathbf{1}_{B_n}) \geq n\|1 - \mathbf{1}_{B_n}\|_{L^1}. \square$$

6.4 Isomorphisms of separable probability spaces

6.4.1 Atoms of a measure space

Let $(\Omega, \mathcal{A}, P)$ be a measure space. $A \in \mathcal{A}$ is called a *P-atom* if $P(A) > 0$ and if, for any $B \in \mathcal{A}$ such that $\mathbf{1}_B \leq \mathbf{1}_A$ a.e., either $\mathbf{1}_B = 0$ a.e. or $\mathbf{1}_B = \mathbf{1}_A$ a.e. This notion of atom corresponds to the one introduced in I-6.2, except that we now consider the classes defined by equality a.e.

6.4.2 Structure theorem (nonatomic case). *Let $(\Omega, \mathcal{A}, P)$ be a separable probability space which is complete and has no P-atoms. Then there exists $f \in L^\infty(\Omega, \mathcal{A})$ such that $0 \leq f \leq 1$ and f is a probability space isomorphism from $(\Omega, \mathcal{A}, P)$ onto $[0, 1]$ equipped with Lebesgue measure.*

PROOF. Let $\{\mathcal{A}_n\}$ be the increasing sequence of finite sub-σ-algebras of $\mathcal{A}$ constructed in 6.1. Note that we could regroup the atoms of $\mathcal{A}_n$ that have measure zero with an atom of strictly positive measure, to produce a new sub-σ-algebra $\mathcal{A}'_n$ such that $P(A) > 0$ if $A \in \mathcal{A}'_n$ and $A \neq \emptyset$. Assume that this has been done.

We next enumerate the atoms of $\mathcal{A}_1$, say $e_{1,1}, \ldots, e_{1,s}$, where $s = s(1)$, and then the atoms of $\mathcal{A}_2$, consistently with the enumeration for $\mathcal{A}_1$. That is, all the atoms into which $e_{1,1}$ is decomposed appear first, then the atoms into which $e_{1,2}$ is decomposed, and so on.

With the atoms of $\mathcal{A}_q$ listed as $e_{q,1}, \ldots, e_{q,s}$, $s = s(q)$, there exists a strictly increasing integer-valued function φ such that

$$e_{q,s} \text{ is decomposed into the atoms } e_{q+1,j}, \text{ with } \varphi(s) \leq j < \varphi(s+1).$$

Having defined this *coherent enumeration* of the atoms of $\mathcal{A}_n$, we set $\beta_{q,r} = P(e_{q,r})$ and define

$$f_q(x) = \frac{1}{2}\beta_{q,l} + \sum_{r<l} \beta_{q,r}, \quad \text{where} \quad x \in e_{q,l}.$$

Then $f_q \in L^\infty(\mathcal{A}_q)$ and

$$\begin{aligned}\mathbf{E}^{\mathcal{A}_q}(f_{q+1})(e_{q,s}) &= \sum_{r<\varphi(s)} \beta_{q+1,r} \\ &+ \frac{1}{P(e_{q,s})} \sum_{\varphi(s)\leq j\leq\varphi(s+1)} P(e_{q+1,j}) \left[\frac{1}{2}\beta_{q+1,j} + \sum_{m<j} \beta_{q+1,m}\right].\end{aligned}$$

In the second sum, observe that

$$\frac{1}{2}\left[\sum_{\varphi(s)\leq j\leq\varphi(s+1)} \beta_{q+1,j}\right]^2 = \frac{1}{2}[P(e_{q,s})]^2.$$

Similarly, the first sum can be written $\sum_{t<s} \beta_{q,t}$, whence

$$\mathbf{E}^{\mathcal{A}_q}(f_{q+1}) = f_q.$$

The f_q form a martingale; since $0 \leq f_q \leq 1$, they form an L^2 *martingale.* This martingale converges a.s. to its final value $f \in L^\infty(\mathcal{A})$, and $f_q = \mathbf{E}^{\mathcal{A}_q}(f)$.

Furthermore, let

$$\eta_q = \sup_r \beta_{q,r}.$$

Then $\{\eta_q\}$ is a decreasing sequence. Assume for contradiction that

$$(i) \qquad \lim \eta_q = \epsilon > 0.$$

Then there exists a decreasing sequence of atoms $a_n \in \mathcal{A}_n$ such that

$$P(\lim \downarrow a_n) = \lim P(a_n) = \epsilon > 0.$$

Let $C = \lim \downarrow a_n$. Since the probability space $(\Omega, \mathcal{A}, P)$ has no P-atoms, we can find $D \in \mathcal{A}$ such that $D \subset C$ and $P(D) > 0$, $P(C - D) > 0$.

Since the σ-algebras $\mathcal{A}_n$ P-generate $\mathcal{A}$,

$$(ii) \qquad \mathbf{E}^{\mathcal{A}_n}(\mathbf{1}_D) \to \mathbf{1}_D \quad \text{a.s.}$$

But, since $\mathbf{1}_D \leq \mathbf{1}_C$, $\mathbf{E}^{\mathcal{A}_n}(\mathbf{1}_D)$ is constant on the atom a_n; that is,

$$(iii) \qquad \lim \mathbf{E}^{\mathcal{A}_n}(\mathbf{1}_D) = \gamma_n \mathbf{1}_C, \quad \text{where } \gamma_n \text{ is a constant.}$$

This contradicts (ii). Hence (i) cannot hold, and

$$(iv) \qquad \lim \eta_q = 0.$$

(v) The distribution of f is Lebesgue measure.

Since the functions f_n converge a.s. to f, their distributions converge to the distribution of f. Let $u \in C([0,1])$ and consider

$$\mathbf{E}(u(f_n(\omega))) = \sum_r \beta_{q,r} u \left(\frac{1}{2}\beta_{q,r} + \sum_{l<r} \beta_{q,l} \right).$$

The right-hand side is a Riemann sum for u, and since the mesh of the partitions tends to zero by (iv), the Riemann sums converge to $\int u \, dx$, whence (v).

(vi) Let $\mathcal{A}' = f^{-1}(\mathcal{B}_{\mathbf{R}})$, where $\mathcal{B}_{\mathbf{R}}$ is the Borel algebra of $\mathbf{R}$. Then $L_P^\infty(\mathcal{A}) = L_P^\infty(\mathcal{A}')$.

Let $\beta'_{q,t} = \sum_{j<t} \beta_{q,j}$. Then, by the construction of the f_j,

$$f_j^{-1}((\beta'_{q,r}, \beta'_{q,r+1})) = e_{q,r} \quad \text{if } j > q.$$

By the a.s. convergence of the f_j,

$$f^{-1}([\beta'_{q,r}, \beta'_{q,r+1}]) \supset e_{q,r} \supset f^{-1}((\beta'_{q,r}, \beta'_{q,r+1})).$$

Since $P(f^{-1}(\beta'_{q,r}))$ = Lebesgue measure of $\{\beta_{q,r}\} = 0$, the two inverse images above differ by sets of probability zero. Hence $L_P^\infty(\mathcal{A}') \supset L_P^\infty(\mathcal{A})$, and (vi) follows.
$\square$

6.4.3 Structure theorem (general case). *Let $(\Omega, \mathcal{A}, P)$ be a separable complete probability space. Then there exists a discrete measure $\gamma = \sum c_k \delta_{\xi_k}$ on $[0,1]$ satisfying the following two conditions:*

$$(i) \qquad \|\gamma\| = \sum c_k \leq 1.$$

(ii) Setting

$$d\mu = d\gamma + (1 - \|\gamma\|) d\xi,$$

there exists a function f in $L^\infty(\Omega, \mathcal{A})$ which is an isomorphism from $(\Omega, \mathcal{A}, P)$ onto $[0,1]$ equipped with the completion with respect to μ of its Borel algebra.

Proof. Let $A_1, \ldots, A_\alpha, \ldots$ denote the P-atoms of $\mathcal{A}$. Since $\sum_\alpha P(A_\alpha) \leq 1$, the set of P-atoms is countable. Let $c_k = P(A_k)$ and let $\xi_k = \frac{1}{k}$; then the measure γ is well defined. If $\|\gamma\| = 1$, the desired isomorphism is clear. If $\|\gamma\| < 1$, set

$$\begin{aligned} \widetilde{\Omega} &= \Omega - \cup_k A_k, \\ \widetilde{\mathcal{A}} &= \mathcal{A} \wedge \widetilde{\Omega}, \end{aligned}$$

and

$$\widetilde{P}(\widetilde{A}) = \frac{1}{P(\widetilde{\Omega})} P(\widetilde{A}) \quad \text{for} \quad \widetilde{A} \in \widetilde{\mathcal{A}}.$$

Applying Theorem 6.4.2 to $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ shows that $\widetilde{f} \in L^\infty(\widetilde{\Omega}, \widetilde{\mathcal{A}})$. Now let

$$f = \widetilde{f} + \sum \frac{1}{k} \mathbf{1}_{A_k}. \square$$

6.5 Conditional probabilities

We would like to express conditional expectation by an integral. This has already been done in 2.4.2 in the case of conditioning relative to a finite sub-σ-algebra.

6.5.1 Theorem. *Let $(\Omega, \mathcal{A}, P)$ be a separable complete probability space, let $(X, \mathcal{B})$ be a measure space, let $f \in M_P((\Omega, \mathcal{A}); (X, \mathcal{B}))$, and let $\mu = f_* P$ be the distribution of f. For a σ-algebra $\mathcal{A}'$ on Ω, let $\pi(\Omega, \mathcal{A}')$ be the set of probability measures defined on $\mathcal{A}'$.*

Then there exist

(i) a σ-algebra $\mathcal{A}' \subset \mathcal{A}$ such that $L_P^\infty(\mathcal{A}') = L_P^\infty(\mathcal{A})$ and
(ii) a mapping $x \mapsto \nu_x$ from X to $\pi(\Omega, \mathcal{A}')$ that is defined μ-a.e. and satisfies

$$\mathbf{E}(u(f(\omega))h(\omega)) = \int_X u(x) d\mu(x) \left[\int_\Omega h(\omega) d\nu_x(\omega) \right]$$

for any $u \in L_\mu^\infty(\mathcal{B})$, $h \in \mathcal{L}^\infty(\mathcal{A}')$.

(The expression in brackets on the right-hand side is a function in $L_\mu^1(\mathcal{B})$.)

PROOF. Using Theorem 6.4.2 on isomorphisms of probability spaces and noting that the "atomic set" appearing in 6.4.3 can be handled easily, we reduce the proof to the case where $\Omega = [0, 1]$, $\mathcal{A}$ is the σ-algebra of Lebesgue-measurable sets, and P is Lebesgue measure. Taking $\mathcal{A}' = \mathcal{B}_{\mathbf{R}}$, this reduction to $[0, 1]$ allows us to use the theory of Radon measures.

Let $g_n = x^n$, $x \in [0, 1]$. Let W denote the finite linear combinations of g_n with rational coefficients, that is the polynomials with rational coefficients. For every $w \in W$, the conditional expectation of w given f is defined in the complement of a μ-negligible set. Taking a countable union of such negligible sets, we can find $B_0 \in \mathcal{B}$ such that $\mu(B_0) = 0$ and

$$l_x(w) = \mathbf{E}(w(\omega) | f(\omega) = x) \text{ is defined } \forall x \in B_0^c.$$

Then l_x is a linear functional on the $\mathbf{Q}$-vector space W. Since $|l_x(w)| \leq \|w\|_C$, the Hahn-Banach theorem implies that l_x extends to a linear functional l'_x defined on $C([0, 1])$.

Hence, by II-5.2, there exists a Radon measure ν_x on $[0,1]$ such that

$$l'_x(h) = \int h(\omega) d\nu_x(\omega), \quad \forall h \in C([0,1]).$$

In particular,

$$\mathbf{E}(w(\omega)|f(\omega) = x) = \int w(\omega)\, d\nu_x(\omega).$$

This formula extends by continuity from W to $L^\infty_P(\Omega)$.

Note finally that $l'_x(w) \geq 0$ if $w \geq 0$; whence ν_x is positive. Taking $f = \mathbf{1}_B$ shows that $\nu_x(\Omega) = 1$. □

6.6 *Product of a countably infinite set of probability spaces*

Theorem. *Let $(\Omega_n, \mathcal{A}_n, P_n)$ be a countably infinite set of probability spaces. Then there exists a unique probability space $(\Omega, \mathcal{A}, P)$ with the following two properties:*

(i) For every q, there exists a morphism from the product of the first q probability spaces $(\Omega_n, \mathcal{A}_n, P_n)$ to $(\Omega, \mathcal{A}, P)$.

(ii) Furthermore, $(\Omega, \mathcal{A}, P)$ is the smallest probability space satisfying (i). More precisely, if $(\Omega', \mathcal{A}', P')$ is a probability space satisfying (i), then there exists a morphism of probability spaces $\Phi : (\Omega', \mathcal{A}', P') \to (\Omega, \mathcal{A}, P)$.

PROOF. By the structure theorem (6.4.3), we can reduce the proof to the case where $\Omega_n = [0,1]$, $\mathcal{A}_n$ is the Borel algebra, and P_n is a Radon measure μ_n which is the sum of a discrete measure and a multiple of Lebesgue measure. Let

$$\Omega = [0,1]^{\mathbf{N}}.$$

Then Ω is a *compact space*, which will be equipped with its Borel algebra. Define an injection

$$f_q : [0,1]^q \to [0,1]^{\mathbf{N}}$$

by setting

$$f_q(\xi, \ldots, \xi_p) = (\xi, \ldots, \xi_q, 0, 0, \ldots).$$

Let

$$(f_q)_*(\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_q) = P_q.$$

Then P_q converges vaguely to a Radon measure P_∞, and $(\Omega, \mathcal{B}_\Omega, P_\infty)$ is the desired probability space. □

V

Gaussian Sobolev Spaces and Stochastic Calculus of Variations

Introduction

In Chapter IV, we began by basing probability theory on the theory of abstract measure spaces of Chapter I. We then studied convergence in distribution by means of the Fourier transform on $\mathbf{R}^d$. Thus both abstract integration theory and classical analysis were necessary to obtain the limit theorems of probability theory. These two sources of Chapter IV derive from the dual nature of distributions. Although a distribution is attached to a very abstract object, a random variable on a probability space, it can also be thought of as given by a Radon measure on $\mathbf{R}$. Borrowing an image from Plato, we might say that distributions have a daemonic nature: they come simultaneously from *celestial objects* (the abstract theory of measure spaces) and *terrestrial objects* (analysis on $\mathbf{R}$).

In this chapter, we study the "regularity of distributions". The concept of regularity is based on the existence of a standard Radon measure on $\mathbf{R}$, Lebesgue measure. A distribution is called *regular* if it has a density k with respect to Lebesgue measure, *very regular* if k is a C^∞ function, and so on. Lebesgue measure is defined in terrestrial terms as the *translation-invariant* Radon measure on $\mathbf{R}$.

To study the regularity of distributions, we will have to go up to the celestial level of quasi-invariant measures. A *Gaussian probability space* is a probability space equipped with a sequence of independent Gaussian random variables that generates the underlying σ-algebra. On such a space, the probability measure is *quasi-invariant* under the action of distinguished translations, those of Cameron-Martin. The action of translations on $L^2(\mathbf{R}^d)$ led in Chapter III to the definition of the Sobolev spaces $H^s(\mathbf{R}^d)$.

Proceeding similarly here, we will define further celestial objects, spaces of infinitely differentiable random variables. We can then use differential calculus on both $\mathbf{R}$ and the probability space. The interaction, through a random variable, of these two kinds of differential calculus will make it possible to study the regularity of distributions.

The use on an abstract probability space of a natural underlying differential structure, as developed here, is commonly called "stochastic calculus of variations".

1 Gaussian Probability Spaces

1.1 **Definition.** Let $(\Omega, \mathcal{A}, P)$ be a probability space and let X be an $\mathbf{R}^n$-valued random variable defined on Ω. X is called a *Gaussian random variable* if the distribution of X is a Gaussian measure on $\mathbf{R}^n$. (See IV-4.3.3.) Gaussian measures and Gaussian random variables are sometimes called *normal.*

REMARK. If X is Gaussian, X is in L^p $\forall p < +\infty$.

1.2 **Definition.** Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\{X_n\}$ be a sequence of independent normal random variables. $(\Omega, \mathcal{A}, P)$ is said to be a *Gaussian space* if the σ-algebra generated by all the X_n is equal to $\mathcal{A}$.

We intend to construct a basis for $L^2(\Omega, \mathcal{A}, P)$.

1.3 Hermite polynomials

On $\mathbf{R}$, we define the Gaussian measure $\nu_1(A) = \int_A \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}}$.

1.3.1 On $L^2(\mathbf{R}, \nu_1)$, we consider the scalar product

$$(\varphi|\psi) = \int \varphi(x)\psi(x)d\nu_1,$$

the differentiation operator $d = \frac{d}{dx}$, and the operator δ defined by

$$\delta\varphi(x) = -\frac{d\varphi}{dx} + x\varphi = -\mathrm{e}^{x^2/2}\frac{d}{dx}\left(\mathrm{e}^{-x^2/2}\varphi\right).$$

1.3.2 Lemma. *When φ and ψ are C^1 functions with compact support,*

(*i*) $$(d\varphi|\psi) = (\varphi|\delta\psi).$$

(ii) Moreover, (i) remains true if φ and ψ are absolutely continuous and in $L^2(\mathbf{R}, \nu_1)$, with $d\varphi$ and $\delta\psi$ in L^2.

(iii) $$d\delta - \delta d = \textit{Identity}.$$

PROOF. (i) follows from an integration by parts, and (ii) from approximating φ and ψ by compactly supported C^1 functions. A different proof will be given later, in 2.2.3. □

1.3.3 **Definition.** The Hermite polynomials are defined by setting $H_0 = 1$ and $H_n = \delta^n 1$ for $n \geq 1$. Here $\delta^n = \delta \circ \ldots \circ \delta$, n times. It is immediate that

$$\begin{aligned} H_1 &= \delta 1 = x, \\ H_2 &= \delta\delta 1 = x^2 - 1, \quad \text{and} \\ H_3 &= \delta^3 1 = x^3 - 3x. \end{aligned}$$

1.3.4 Proposition. *H_n is a polynomial of degree n whose highest-degree term is x^n. The following relations hold:*

$$\begin{aligned} &(i) & \delta H_n &= H_{n+1}; \\ &(ii) & dH_n &= nH_{n-1}; \\ &(iii) & (\delta + d)H_n &= xH_n; \\ &(iv) & \delta d H_n &= nH_n. \end{aligned}$$

PROOF.

(i) follows immediately from the definition.
(ii) is proved by induction, using 1.3.2.(iii):

$$dH_n = d\ \delta H_{n-1} = \delta\ dH_{n-1} + H_{n-1} = (n-1)\delta H_{n-2} + H_{n-1}.$$

(iii) follows from the definition of the operator δ (1.3.1).
(iv) follows from (ii) and the definition of H_n. □

1.3.5 Corollary. *Let $\mathcal{F}(g(x))(\xi) = \int_{-\infty}^{+\infty} e^{i\xi x} g(x)dx$ be the Fourier transform of g at the point ξ. Then $\mathcal{F}(H_n(x)e^{-x^2/2}) = i^n \xi^n e^{-\xi^2/2}$.*

PROOF.

$$\begin{aligned} \mathcal{F}(\delta^n 1.\mathrm{e}^{-x^2/2})(\xi) &= (\mathrm{e}^{i\xi x} | \delta^n 1) = (d^n \mathrm{e}^{i\xi x} | 1) = i^n \xi^n (\mathrm{e}^{i\xi x} | 1) \\ &= i^n \xi^n \int_{-\infty}^{+\infty} \mathrm{e}^{i\xi x - x^2/2} \frac{dx}{\sqrt{2\pi}} = i^n \xi^n \mathrm{e}^{-\xi^2/2}. \square \end{aligned}$$

1.3.6 Theorem. $\left\{ \frac{1}{(n!)^{\frac{1}{2}}} H_n \right\}$ *is an orthonormal basis of $L^2(\mathbf{R}, \nu_1)$.*

PROOF. (i) We show that the polynomials H_n are dense in $L^2(\mathbf{R}, \nu_1)$. Otherwise there would exist $\varphi \in L^2$ such that $(\varphi | x^k) = 0 \ \forall k = 0, 1, \dots$

Let

$$F(t) = \int \varphi(v) \mathrm{e}^{itv - v^2/2} dv.$$

Setting $t = \sigma + i\tau$, we have

$$\int |\varphi(v)| \mathrm{e}^{-v^2/2} \mathrm{e}^{-\tau v} dv \leq \left[\int |\varphi(v)|^2 \mathrm{e}^{-v^2/2} dv \right]^{1/2} \left[\int \mathrm{e}^{-2\tau v} \mathrm{e}^{-v^2/2} dv \right]^{1/2}.$$

Thus I-7.8.4 can be used to differentiate under the integral sign, showing that F is an entire function of t. Since

$$F^{(k)}(0) = i^k \int v^k \varphi(v) \mathrm{e}^{-v^2/2} dv = 0$$

for every k, $F \equiv 0$. Applying the inverse Fourier transform, we see that $\varphi \equiv 0$.

(ii) The polynomials H_k are linearly independent since the coefficient of x^k in H_k is 1.

(iii) We show that the functions $\frac{1}{(k!)^{1/2}} H_k$ form an orthonormal system. If $s > k$, then

$$(H_k | H_s) = (d^s \delta^k 1 | 1) = 0,$$

since $d^s \delta^k 1 = 0$. If $s = k$, then $d^s \delta^s 1$ is the product of $s!$ and the coefficient of the highest-degree term of H_s; that is, $d^s \delta^s 1 = s!$. □

1.4 Hermite series expansion

1.4.1 Theorem. *Let g be a C^∞ function on $\mathbf{R}$ such that g and all its derivatives are in $L^2(\mathbf{R}, \nu_1)$. The expansion of g with respect to the basis $\frac{1}{(n!)^{1/2}} H_n$ is*

$$g(x) = \sum_{n=0}^{\infty} \frac{1}{n!} \mathbf{E}[g^{(n)}] H_n(x),$$

where $\mathbf{E}(g^{(n)}) = (g^{(n)} | 1)$ and $g^{(n)}$ is the n*th-order derivative of g ($g^{(0)} = g$).*

PROOF. Let $g(x) = \sum_{n=0}^{\infty} C_n H_n(x)$ be the Hermite series expansion of g. Integrating term by term and using the orthogonality of the polynomials H_k, we have

$$\int_{-\infty}^{+\infty} H_k(x) g(x) \frac{\mathrm{e}^{-x^2/2} dx}{\sqrt{2\pi}} = C_k k!$$

Moreover, $(H_k | g) = (\delta^k 1 | g) = (1 | d^k g)$. Hence $C_k = \dfrac{\mathbf{E}[g^{(k)}]}{k!}$. □

1.4.2 EXAMPLE. $\exp(-\frac{t^2}{2} + tx) = \sum_{n=0}^{\infty} \frac{t^n}{n!} H_n(x)$.

PROOF.
$$\frac{d^n}{dx^n} \exp\left(-\frac{t^2}{2} + tx\right) = t^n \exp\left(-\frac{t^2}{2} + tx\right)$$
and
$$\int_{-\infty}^{+\infty} t^n \exp\left(-\frac{t^2}{2} + tx\right) \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}} = t^n. \square$$

1.4.3 Corollary. $\frac{H_n(x)}{n!} = \frac{1}{2\pi i} \int_\gamma z^{-(n+1)} \exp(-\frac{z^2}{2} + zx)dz$, *where* γ *is a simple closed curve around the origin in* $\mathbf{C}$.

PROOF. This follows from 1.4.2 and the Cauchy formula. $\square$

1.5 The Ornstein-Uhlenbeck operator on $\mathbf{R}$

1.5.1 **Definition.** $\mathcal{L} = \delta d = -\frac{d^2}{dx^2} + x\frac{d}{dx}$ is called the Ornstein-Uhlenbeck operator on $\mathbf{R}$.

1.5.2 Lemma. $\mathcal{L} H_n = nH_n$.

PROOF. By 1.3.4(iv). $\square$

1.5.3 **Definition.** Let P_θ be the operator defined by
$$P_\theta f(y) = \int_{-\infty}^{+\infty} f(x\cos\theta + y\sin\theta)\mathrm{e}^{-x^2/2}\frac{dx}{\sqrt{2\pi}}.$$

REMARK. The integral above takes the same value for θ and $\pi - \theta$; it depends only on $\sin\theta$.

1.5.4 Proposition.

(i) $(P_\theta\varphi|\psi) = (\varphi|P_\theta\psi)$;
(ii) $dP_\theta = \sin\theta P_\theta d$;
(iii) $P_\theta\delta = \sin\theta\ \delta P_\theta$;
(iv) $\mathcal{L}P_\theta = P_\theta\mathcal{L}$;
(v) $P_\theta H_n = (\sin\theta)^n H_n$.

PROOF. The measure $\exp(-\frac{x^2+y^2}{2})\frac{dx\ dy}{\sqrt{2\pi}}$ is rotation invariant. (i) follows from this; (ii) is immediate; (iii) follows from (i) and (ii) and the fact that $(\delta\varphi|\psi) = (\varphi|d\psi)$. (ii) and (iii) imply (iv).

(v) is proved by alternately using (iii) and the fact that $\delta H_n = H_{n+1}$:
$$P_\theta H_n = P_\theta\delta H_{n-1} = \sin\theta\ \delta P_\theta H_{n-1} = \sin\theta\ \delta P_\theta\delta H_{n-2} = \sin^2\theta\ \delta^2 P_\theta H_{n-2}.$$
Iterating this gives
$$(\sin\theta)^n\delta^n P_\theta H_0 = (\sin\theta)^n\delta^n 1 = (\sin\theta)^n H_n. \square$$

1.5.5 Proposition. *Let* $\theta(t) = \arcsin(e^{-t})$, *where* $t > 0$. *Then*

(i) $\frac{d}{dt}(P_{\theta(t)}f) = -\mathcal{L}(P_{\theta(t)}f)$, *and*
(ii) $P_{\theta(t)} \circ P_{\theta(t')} = P_{\theta(t+t')}$.

PROOF. Since P_θ depends only on $\sin\theta$, we can set $\theta(t) = \arcsin(\mathrm{e}^{-t})$. Then

$$\begin{aligned}\frac{d}{d\theta}(P_\theta f)(y) &= \int_{-\infty}^{+\infty}(-x\sin\theta + y\cos\theta)f'(x\cos\theta + y\sin\theta)\frac{\mathrm{e}^{-x^2/2}dx}{\sqrt{2\pi}}\\ &= y\cos\theta P_\theta(df)(y) - \sin\theta\cos\theta P_\theta(d^2f)(y),\end{aligned}$$

where the second term comes from an integration by parts. Using 1.5.4(ii),

$$\frac{d}{d\theta}(P_\theta f)(y) = (ydP_\theta f(y) - d^2P_\theta f(y))\frac{\cos\theta}{\sin\theta}.$$

Since $\frac{d\theta}{dt} = -\frac{\sin\theta}{\cos\theta}$,

$$\frac{d}{dt}(P_{\theta(t)}f) = \frac{d}{d\theta}(P_\theta f)\frac{d\theta}{dt} = -\mathcal{L}P_{\theta(t)}f.$$

This proves (i).

We now prove (ii). By 1.5.4(v), $P_\theta \circ P_{\theta'}\ H_n = \sin(n\theta)\sin(n\theta')H_n$. Since $\sin\theta(t)\sin\theta(t') = \sin\theta(t+t') = \mathrm{e}^{-(t+t')}$, this implies (ii) for finite linear combinations of Hermite polynomials and hence, passing to the limit, for L^2. □

1.5.6 Lemma. $(P_\theta f)(y) = \int_{-\infty}^{+\infty} f(x)K_\theta(x,y)e^{-x^2/2}\frac{dx}{\sqrt{2\pi}}$, *where*

$$K_\theta(x,y) = \frac{1}{|\cos\theta|}\exp\left[\frac{2xy\sin\theta - \sin^2\theta(x^2+y^2)}{2\cos^2\theta}\right].$$

PROOF. This follows from the change of variables $u = x\cos\theta + y\sin\theta$ in 1.5.3. □

1.5.7 REMARK. Since the operator P_θ is self-adjoint with respect to the scalar product (see 1.5.4(i)), the kernel K_θ is symmetric in x and y.

1.5.8 Examples of expansion in Hermite series

(i) $H_n(x\cos\theta + y\sin\theta) = \sum_{p=0}^{n}\binom{n}{p}(\cos\theta)^p(\sin\theta)^{n-p}H_p(x)H_{n-p}(y)$
(ii) $K_\theta(x,y) = \sum_{n=0}^{\infty}\frac{(\sin\theta)^n}{n!}H_n(x)H_n(y) = \exp(\sin\theta\ \delta_1\delta_2)\mathbf{1}(x)\mathbf{1}(y)$, *where* $\delta_1 = -\frac{d^2}{dx^2} + x\frac{d}{dx}$ *and* $\delta_2 = -\frac{d^2}{dy^2} + y\frac{d}{dy}$.

PROOF.

(i) $(\frac{d}{dx})^p H_n(x\cos\theta + y\sin\theta) = \frac{(\cos\theta)^p n!}{(n-p)!} H_{n-p}(x\cos\theta + y\sin\theta)$ if $p \le n$. To evaluate $P_\theta H_{n-p}$, use 1.5.4(v) followed by Theorem 1.4.1.

(ii) Expand $y \mapsto K_\theta(x,y)$ in a Hermite series, using Theorem 1.4.1:

$$\begin{aligned}\int_{-\infty}^{+\infty}\left[\left(\frac{d}{dy}\right)^n K_\theta(x,y)\right] e^{-y^2/2}\frac{dy}{\sqrt{2\pi}} &= \int_{-\infty}^{+\infty} K_\theta(x,y)(\delta^n \mathbf{1})e^{-y^2/2}\frac{dy}{\sqrt{2\pi}} \\ &= \int_{-\infty}^{+\infty} K_\theta(x,y)H_n(y)e^{-y^2/2}\frac{dy}{\sqrt{2\pi}}.\end{aligned}$$

By 1.5.4(v), this equals $(\sin\theta)^n H_n(x)$. □

1.6 Canonical basis for the L^2 space of a Gaussian probability space

1.6.1 **Notation.** Let $\mathbf{R}^{\mathbf{N}}$ be the set of real-valued sequences and let $\mathcal{B}_\infty$ be the Borel algebra on $\mathbf{R}^{\mathbf{N}}$. Projection onto the first n coordinates is denoted by $\pi_n : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}^n$. It follows from the structure theorem (IV-6.6) that there exists a measure ν on $\mathbf{R}^{\mathbf{N}}$ such that the direct image $(\pi_n)_*\nu$ of ν under π satisfies $(\pi_n)_*\nu = \nu_n$, where $\nu_n = \prod_{i=1}^n \left(\frac{e^{-x_i^2/2}}{\sqrt{2\pi}}\right) dx_i$. $\mathcal{B}_n$ denotes the inverse image under π_n of $\mathcal{B}_\infty$.

1.6.2 Proposition. *The increasing sequence $\{\mathcal{B}_n\}$ of σ-algebras is a filtration of the space $(\mathbf{R}^{\mathbf{N}}, \mathcal{B}_\infty, \nu)$.*

The space $(\mathbf{R}^{\mathbf{N}}, \mathcal{B}_\infty, \nu)$ is a Gaussian probability space and $\mathcal{B}_\infty$ is the σ-algebra generated by the Gaussian variables X_n of projection onto the nth coordinate.

PROOF. Follows from the definitions. □

1.6.3 Proposition. *Let $f \in L^2(\Omega, \mathcal{A}, P)$ There exists $\tilde{f} : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}$ such that*

$$f(\omega) = \tilde{f}(X_1(\omega), X_2(\omega), \ldots, X_n(\omega), \ldots).$$

PROOF. By Dynkin's theorem, IV-1.5.4.

1.6.4 Lemma. *If $(\Omega, \mathcal{A}, P)$ is a Gaussian space and $\{X_n\}$ is a sequence of Gaussian random variables that generates $\mathcal{A}$, then $X_n \in L^{2p}(\Omega, \mathcal{A}, P)$ for $1 \le p < \infty$.*

PROOF. The integral $\displaystyle\int x^{2p} e^{-x^2/2} dx$ converges. □

1.6.5 **Definition.** Let $\mathcal{E}$ be the set of sequences of integers $(n_1, n_2, \ldots, 0, \ldots)$ for which all but finitely many terms are zero. For $\mathbf{p} = (n_1, \ldots, n_k, 0, \ldots) \in$

$\mathcal{E}$ and $\omega \in \Omega$, let

$$\mathcal{H}_{\mathbf{p}}(\omega) = \prod_{i=1}^{k} H_{n_i}(X_i(\omega)),$$

where H_{n_i} are the Hermite polynomials on $\mathbf{R}$. We write $\mathbf{p}! = \prod_{i=1}^{k} n_i!$.

1.6.6 Theorem. $\{(\mathbf{p}!)^{-1/2}\mathcal{H}_p(\omega)\}$ *is an orthonormal basis of* $L^2(\Omega, \mathcal{A}, P)$.

PROOF. We prove the theorem for $\mathbf{R}^2$. By IV-3.5.1, linear combinations of the form $f = \sum f_i h_i$, where $f_i(\xi_1)$, $h_i(\xi_2) \in L^2(\mathbf{R}, \nu_1)$, are dense in $L^2(\mathbf{R}^2, \nu_2)$. Approximating the functions f_i, h_i by their expansions in Hermite polynomials shows that the set of functions of the form

$$H_{k_1}(\xi_1)H_{k_2}(\xi_2)$$

generates $L^2(\mathbf{R}^2, \exp(-\frac{\xi_1^2+\xi_2^2}{2})\frac{d\xi_1 d\xi_2}{2\pi})$.

Moreover,

$$(H_{k_1}(\xi_1)H_{k_2}(\xi_2)|H_{s_1}(\xi_1)H_{s_2}(\xi_2)) = \frac{1}{k_1!k_2!}\delta(s_1, k_1)\delta(s_2, k_2). \square$$

1.6.7 Theorem (Taylor-Stroock formula). *Set*

$$\mathbf{E}(\partial_{\mathbf{p}}\tilde{f}) = \int \partial_1^{n_1}\partial_2^{n_2}\dots\partial_k^{n_k}\tilde{f}(x_1,\dots,x_k)\nu_k(dx)$$

for $\mathbf{p} = (n_1, n_2, \dots, n_k, 0, \dots, 0, \dots)$. *If* $\mathbf{E}(\partial_{\mathbf{p}} f)$ *exists for every* $\mathbf{p}$, *then*

$$f(\omega) = \sum_{\mathbf{p}\in\mathcal{E}} \frac{1}{\mathbf{p}!}\mathbf{E}(\partial_{\mathbf{p}}\tilde{f})\mathcal{H}_{\mathbf{p}}(\omega).$$

PROOF. It suffices to prove 1.6.7 when $\Omega = \mathbf{R}^n$ and $f : \mathbf{R}^n \to \mathbf{R}$; that is,

$$\tilde{f}(x_1,\dots,x_n) = \sum_{\mathbf{p}\in\mathcal{E}} \frac{1}{\mathbf{p}!}\mathbf{E}(\partial_{\mathbf{p}}\tilde{f})H_{\mathbf{p}}(x_1,\dots,x_n).$$

The proof proceeds as in 1.4.1. $\square$

1.7 Isomorphism theorem. *There exists an isomorphism* φ *between* $L^2(\Omega, \mathcal{A}, P)$ *and* $L^2(\mathbf{R}^{\mathbf{N}}, \mathcal{B}_\infty, \nu)$.

1.8 The Cameron-Martin theorem on $(\mathbf{R}^N, \mathcal{B}_\infty, \nu)$*: quasi-invariance under the action of* ℓ^2

1.8.1 Proposition. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space and let* $\{M_n\}$ *be a sequence of integrable random variables such that* $\sup_n \mathbf{E}(|M_n|^p) = C_p < +\infty$ *for* $p > 1$. *Then* $\{M_n\}$ *is uniformly integrable.*

PROOF. $\int_A |M_n| dP \leq \|M_n\|_{L^p} \|\mathbf{1}_A\|_{L^q} \leq C_p P(A)^{1/q}$ for $A \in \mathcal{A}$ (cf. IV-5.7.1). □

In what follows, we consider the Gaussian probability space $(\mathbf{R}^{\mathbf{N}}, \mathcal{B}_\infty, \nu)$ (cf. 1.6.1 and 1.6.2).

1.8.2 **Notation.** Let ℓ^2 be the space of sequences $x = (x_1, \ldots, x_n, \ldots) \in \mathbf{R}^{\mathbf{N}}$ such that $\sum_{i=1}^{\infty} x_i^2 < +\infty$. The scalar product $(x|y) = \sum_{i=1}^{\infty} x_i y_i$ is associated with the norm $|x| = \sqrt{\sum x_i^2}$ on ℓ^2. Let $\tau_y : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}^{\mathbf{N}}$ denote the mapping defined by $\tau_y(x) = x + y$ and let $(\tau_y)_* \nu$ denote the image of the measure ν under τ_y. (See IV-1.4.3.)

1.8.3 Theorem (Cameron-Martin). *If $y \in \ell^2$, then the image measure $(\tau_y)_*\nu$ is absolutely continuous with respect to ν and the density is given by*

$$\frac{d((\tau_y)_*\nu)}{d\nu}(z) = \exp\left(\sum_{k=1}^{\infty} y_k z_k - \frac{1}{2} \sum_{k=1}^{\infty} y_k^2 \right).$$

PROOF. Let $S_n(z) = \sum_{k=1}^{n} y_k z_k$.

(i) The sequence $\{S_n\}$ on $(\mathbf{R}^{\mathbf{N}}, \mathcal{B}_\infty, \nu)$ is an L^2 martingale relative to the filtration $\{\mathcal{B}_n\}$. Hence $\{S_n\}$ converges a.s. Let $S_\infty = \lim_{n\to\infty} S_n$. Then $S_\infty < +\infty$ a.s.

(ii) The a.s. convergence of S_n implies its convergence in distribution. This follows from IV-1.8.5.

(iii) Set $M_n(z) = \exp(\sum_{k=1}^{n} y_k z_k - \frac{1}{2} \sum_{k=1}^{n} y_k^2)$.

The sequence $\{M_n\}$ is a martingale relative to the filtration $\{\mathcal{B}_n\}$ and, for all n, $\mathbf{E}(M_n) = 1$. By Fatou's theorem (IV-5.6.3), the limit $M_\infty = \lim_{n\to\infty} M_n$ exists a.s.

(iv)

$$\mathbf{E}(M_n^p) = \mathbf{E}\left[\exp\left(\sum_{k=1}^{n} p y_k z_k - \frac{1}{2} p^2 \sum_{k=1}^{n} y_k^2 \right) \right].$$

But

$$\mathbf{E}\left[\exp\left(\sum_{k=1}^{n} (p y_k) z_k - \frac{1}{2} p^2 \sum y_k^2 \right) \right] = 1.$$

Hence

$$\mathbf{E}[M_n^p] = \exp\left(\frac{1}{2}(p^2 - p) \sum_{k=1}^{n} y_k^2 \right).$$

M_n is therefore bounded in L^p and hence uniformly integrable. By IV-5.8, M_n converges in L^p to M_∞. Thus M_n is the conditional expectation of M_∞. Given a function f depending on the first r coordinates, we have

$$\mathbf{E}((\tau_y)_* f) = \mathbf{E}(M_r f) = \mathbf{E}(M_\infty f).$$

The equality $\mathbf{E}(M_r f) = \mathbf{E}(M_\infty f)$ extends by continuity to all f in L^p. □

REMARK. Although ν is not invariant under the Cameron-Martin translations τ_y, the measure $(\tau_y)_*\nu$ is absolutely continuous with respect to ν. This property of $(\tau_y)_*\nu$ is called *quasi-invariance.*

2 Gaussian Sobolev Spaces

2.1 Finite-dimensional spaces

2.1.1 Notation

Let $f \in L^2(\mathbf{R}^k, \nu_k)$. We write $\mathbf{E}(f) = \int f(x_1, \ldots, x_k) d\nu_k$. By 1.5.8(iii), f can be expanded in a Hermite series. If $\mathbf{p} = (p_1, \ldots, p_k) \in \mathbf{N}^k$, we set $\mathbf{p}! = p_1! p_2! \ldots p_k!$ and $|\mathbf{p}| = p_1 + p_2 + \ldots + p_k$.

Then

$$f(x_1 \ldots, x_k) = \sum_{p_1 \cdots p_k} c_{p_1 \ldots p_k}(f) \frac{H_{p_1}(x_1) \ldots H_{p_k}(x_k)}{\mathbf{p}!}.$$

Let

$$f_{p_1 \ldots p_k}(x_1, \ldots, x_k) = c_{p_1 \ldots p_k}(f) \frac{H_{p_1}(x_1) \ldots H_{p_k}(x_k)}{\mathbf{p}!}.$$

Then

$$\|f\|^2_{L^2(\nu_k)} = \sum_{p_1 \cdots p_k} \|f_{p_1 \ldots p_k}\|^2_{L^2(\nu_k)} = \sum_{p_1 \cdots p_k} \frac{|c_{p_1 \ldots p_k}|^2}{\mathbf{p}!}.$$

For a C^1 function $\varphi : \mathbf{R}^k \to \mathbf{R}$, we have the partial differential operators

$$\partial_j \varphi = \frac{\partial}{\partial x^j} \varphi \quad \text{and} \quad \delta_j \varphi = (-\partial_j + x_j)\varphi.$$

2.1.2 Operators on $L^2(\mathbf{R}^k)$

An operator T defined on the polynomials can be extended to a *formal operator* on $L^2(\mathbf{R}^k)$ as follows: for $f \in L^2(\nu_k)$, let

$$Tf = \sum_{p_1 \cdots p_k} c_{p_1 \ldots p_k}(f) \frac{1}{\mathbf{p}!} T[H_{p_1}(x_1) \ldots H_{p_k}(x_k)].$$

The *domain* of T consists of those $f \in L^2$ such that $Tf \in L^2$.

Restricting our attention to differential operators, we consider ∂_j, δ_j, and $\mathcal{L} = \sum_{j=1}^k \delta_j \partial_j$. Let

$$(i) \qquad \mathcal{L}f = \sum_j \delta_j \partial_j f = \sum_{p_1 \cdots p_k} c_{p_1 \ldots p_k}(f) |\mathbf{p}| \frac{H_{p_1}(x_1) \ldots H_{p_k}(x_k)}{\mathbf{p}!},$$

and let the gradient operator be defined by $\nabla f = (\partial_1 f, \ldots, \partial_k f)$. For $z = (z_1, \ldots, z_j, \ldots, z_k)$, with $z_j \in L^2(\nu_k)$, $j = 1, \ldots, k$, we set

$$(ii) \qquad \delta z = \sum_{j=1}^{k} \delta_j z_j.$$

When $\partial_j f \in L^2(\nu_k)$ for $j = 1, \ldots, k$, we set

$$(iii) \qquad \|\nabla f\| = \left(\sum_{j=1}^{k} (\partial_j f)^2_{L^2(\nu_k)} \right)^{1/2}.$$

Similarly, if $\partial_{j_1} \ldots \partial_{j_2} f \in L^2(\nu_k)$ $\forall j_1, \ldots, j_s$, we set

$$(iv) \qquad \begin{aligned} \|\nabla^2 f\| &= \left(\sum_{j_1, j_2} (\partial_{j_1} \partial_{j_2} f)^2 \right)^{1/2}, \\ \|\nabla^s f\| &= \left(\sum_{j_1, \ldots j_s} (\partial_{j_1} \ldots \partial_{j_s} f)^2 \right)^{1/2}. \end{aligned}$$

We intend to determine the domains of the operators ∇, ∇^2, and $\mathcal{L}$; that is, the set of functions $f \in L^2(\nu_k)$ whose images under these operators are in $L^2(\nu_k)$. Recall that the Sobolev space $W^p_{r,loc}$ was defined in III-3.4.3.

2.1.3 **Definition.**

$$\mathbf{D}^p_r(\mathbf{R}^k) = \left\{ f \in W^p_{r,loc}(\mathbf{R}^k) : \sum_{s=0}^{r} \|\nabla^s f\|_{L^p(\nu_k)} < +\infty \right\}.$$

2.1.4 Theorem. $\mathbf{D}^p_r(\mathbf{R}^k)$*, with the norm* $\|f\|_{D^p_r} = \sum_{k=0}^{r} \|\nabla^k f\|_{L^p(\nu_k)}$*, is a complete space.*

PROOF. By III-3.3.6. □

2.2 Using Hermite series to characterize $\mathbf{D}^2_s(\mathbf{R})$ *in the Gaussian* L^2 *space*

Let $f \in L^2 = L^2(\mathbf{R}, \nu_1)$ and let $\sum_{n=0}^{\infty} c_n(f) \frac{H_n}{n!} = \sum_{n=0}^{\infty} f_n$ be its Hermite series expansion.

2.2.1 **Definition.** On L^2, the *formal operators* d, δ, and $\mathcal{L}$ are defined by

$$df = \sum_{n\geq 1} c_n(f)\frac{dH_n}{n!} = \sum_{n\geq 0} c_{n+1}(f)\frac{H_n}{n!};$$
$$\delta f = \sum_{n\geq 0} c_n(f)\frac{H_{n+1}}{n!} = \sum_{n\geq 0} c_n(f)(n+1)\frac{H_{n+1}}{(n+1)!};$$
$$\mathcal{L}f = \sum_{n\geq 1} c_n(f)\frac{H_n}{(n-1)!}.$$

The reader can easily verify that $\mathcal{L} = \delta d$.

2.2.2 Lemma.

(i) $|c_n(f)|^2 = n!\|f_n\|^2_{L^2}$.

For an integer $s \geq 1$,

(ii) $c_n(d^s f) = c_{n+s}(f)$*;*
(iii) $c_n(\delta^s f) = 0$ *if* $n < s$ *and* $c_n(\delta^s f) = \frac{n!}{(n-s)!}c_{n-s}(f)$ *if* $n \geq s$*;*
(iv) $c_n(\mathcal{L}^s f) = n^s c_n(f)$.

PROOF. (i) follows from 1.3.6, since $f_n = c_n(f)\frac{H_n}{n!}$.
The other identities follow from the definitions and from the relations $dH_n = nH_{n-1}$ and $\delta H_n = H_{n+1}$. □

2.2.3 Proposition. *Let* $f \in L^2$ *and let* $\sum_{n\geq 0} f_n$ *be its Hermite series expansion. Then for* s *an integer,* $s \geq 1$*, properties (i) through (iv) below are equivalent.*

(i) $$d^s f \in L^2.$$

(ii) $$\sum_{n\geq 1} n^s\|f_n\|^2_{L^2} < +\infty.$$

(iii) $$f \in \mathbf{D}^2_s(\mathbf{R}).$$

(iv) $$\delta^s f \in L^2.$$

In particular, $\mathbf{D}^2_s(\mathbf{R})$ *is the domain of the operators* d^s *and* δ^s *on* L^2*. If both* f *and* g *are in* $\mathbf{D}^2_1(\mathbf{R})$*, then*

(v) $$(df|g)_{L^2} = (f|\delta g)_{L^2}.$$

PROOF. (i) ⇒ (ii). $d^s f = \sum_{n\geq 0} c_{n+s}(f)\frac{H_n}{n!}$ since $dH_n = nH_{n-1}$. Hence, by 2.2.2,

$$\|d^s f\|^2_{L^2} = \sum_{n\geq 0} |c_{n+s}(f)|^2\frac{1}{n!} = \sum_{n\geq 0}\frac{(n+s)!}{n!}\|f_{n+s}\|^2.$$

This proves (ii), since $\frac{(n+s)!}{n!} \sim (n+s)^s$ as $n \to \infty$.

(ii) $\Rightarrow$ (i). If (ii) holds, the series $\sum_{n\geq 0} c_{n+s}(f)\frac{H_n}{n!}$ converges in L^2 and $d^s f \in L^2$.

(iv) $\Rightarrow$ (ii). By 2.2.1, $\delta^s f = \sum_{n\geq 0} c_n(f)(n+1)(n+2)\dots(n+s)\frac{H_{n+s}}{(n+s)!}$. Hence, by 2.2.2,

$$\begin{aligned}\|\delta^s f\|_{L^2}^2 &= \sum_{n\geq 0} |c_n(f)|^2 \frac{(n+1)(n+2)\dots(n+s)}{n!} \\ &= \sum_{n\geq 0} (n+1)(n+2)\dots(n+s)\|f_n\|^2.\end{aligned}$$

(ii) $\Rightarrow$ (iii). We give the proof for the case $s = 1$. Let $\varphi \in \mathcal{D}$; then

$$\begin{aligned}\textstyle\int \varphi'(x)f(x)dx &= \sum_{n\geq 0} \int \varphi'(x)c_n(f)\frac{H_n(x)}{n!}dx \\ &= -\sum_{n\geq 0} \int \varphi(x)c_n(f)\frac{nH_{n-1}(x)}{n!}dx \\ &= -\int \varphi(x)df(x)dx.\end{aligned}$$

Hence $f \in W^2_{1,loc}$. Since $df \in L^2$, it follows that $f \in \mathbf{D}^2_1(\mathbf{R})$.

(iii) $\Rightarrow$ (ii). As above, we give the proof only when $s = 1$. Let df be the weak derivative of f. Then $df \in L^2$ and $df = \sum_{n\geq 0} c_n(df)\frac{H_n}{n!}$ for $\varphi \in \mathcal{D}$, and hence

$$\int \varphi'(x)f(x)ds = -\int \varphi(x)df(x)dx.$$

This implies that $c_{n+1}(f) = c_n(df)$, and (ii) follows.

(v) is proved by using the orthogonality of the Hermite polynomials. □

2.2.4 Proposition. *If $s = 2p$, then (i), (ii), (iii), and (iv) of 2.2.3 are equivalent to*

(vi) $$\mathcal{L}^p f \in L^2.$$

PROOF. We verify only that (ii) $\Rightarrow$ (vi). Since $\mathcal{L}f = \sum_{n\geq 0} c_n(f)n\frac{H_n}{n!}$,

$$\|\mathcal{L}f\|_2^2 = \sum_{n\geq 0} c_n(f)^2\frac{n^2}{n!} = \sum_{n\geq 0} n^2\|f_n\|^2.$$

This implies equivalence when $s = 1$. The proof for $s > 1$ is similar. □

2.2.5 Lemma. *The following identities hold:*

(i) $$\|\delta f\|_{L^2}^2 = \|df\|_{L^2}^2 + \|f\|_{L^2}^2$$

and if k is an integer, $k \geq 1$,

$$(ii) \qquad \|d^k \delta f\|_{L^2}^2 = \|d^{k+1} f\|_{L^2}^2 + (2k+1)\|d^k f\|_{L^2}^2 + k^2 \|d^{k-1} f\|_{L^2}^2.$$

PROOF OF (i).

$$\|\delta f\|_{L^2}^2 = \sum_{n\geq 0} (n+1)\|f_n\|_{L^2}^2 = \|f_0\|_{L^2}^2 + \sum_{n\geq 0}(n+2)\|f_{n+1}\|_{L^2}^2,$$

$$\|df\|_{L^2}^2 = \sum_{n\geq 0}(n+1)\|f_{n+1}\|_{L^2}^2.$$

Hence

$$\|\delta f\|_{L^2}^2 - \|df\|_{L^2}^2 = \|f_0\|_{L^2}^2 + \sum_{n\geq 0} \|f_{n+1}\|_{L^2}^2 = \|f\|_{L^2}^2.$$

PROOF OF (ii). Since $c_{n+1}(\delta f) = (n+1)c_n(f)$ and $|c_n(f)|^2 = n!\|f_n\|_{L^2}^2$,

$$\begin{aligned}\|d^k \delta f\|_{L^2}^2 &= \sum_{n\geq 0} |c_{n+k}(\delta f)|^2 \times \frac{1}{n!} = \sum_{n\geq 0} |c_{n+k-1}(f)|^2 (n+k)^2 \frac{1}{n!} \\ &= \sum_{n\geq 0} \frac{(n+k)!}{n!}(n+k)\|f_{n+k-1}\|^2.\end{aligned}$$

Since

$$\|d^{k+1} f\|_{L^2}^2 = \sum_{n\geq 0} \frac{(n+k+1)!}{n!}\|f_{n+k+1}\|^2,$$

we can compute $\|d^k \delta f\|_{L^2}^2 - \|d^{k+1} f\|_{L^2}^2$ by observing that

$$\frac{(n+k+2)!(n+k+2)}{(n+2)!} - \frac{(n+k+1)!}{n!} = \frac{(n+k+1)!}{(n+2)!}[(2k+1)n + k^2 + 4k + 2].$$

□

2.2.6 Lemma (Differentiation of composite functions). *Let $g \in \mathbf{D}_1^4(\mathbf{R})$ and let $\mu = g_* \nu_1$ be the image of ν_1 under g. If $\varphi \in L^4(\mu)$ and $d\varphi \in L^4(\mu)$, then $\varphi \circ g \in \mathbf{D}_1^2(\mathbf{R})$ and $d(\varphi \circ g)(x) = (d\varphi)(g(x))dg(x)$.*

PROOF. d is the extension of the differentiation operator $\frac{d}{dx}$. By Hölder's inequality, if $f_1 \in L^4$ and $f_2 \in L^4$, then $f_1 f_2 \in L^2$. □

2.2.7 Lemma. *If $f \in \mathbf{D}_1^4(\mathbf{R})$ and $g \in \mathbf{D}_1^4(\mathbf{R})$, then $fg \in \mathbf{D}_1^2(\mathbf{R})$ and*

$$(i) \qquad d(fg) = f dg + g df.$$

If $f \in \mathbf{D}_2^4(\mathbf{R})$ and $g \in \mathbf{D}_2^4(\mathbf{R})$, then $fg \in \mathbf{D}_2^2(\mathbf{R})$ and

$$(ii) \qquad \mathcal{L}(fg) = \mathcal{L}(f)g + \mathcal{L}(g)f + df\ dg.$$

PROOF. (i) and (ii) follow from identities obtained when f and g are differentiable, since d (respectively $\mathcal{L}$) is the extension of the operator $\frac{d}{dx}$ (respectively δd — see 1.5.1). Hölder's inequality implies that $d(fg) \in L^2$ and $\mathcal{L}(fg) \in L^2$. □

2.3 The spaces $\mathbf{D}_s^2(\mathbf{R}^k)$ $(k \geq 1)$

2.3.1 Proposition. *Let $f \in L^2(\nu_k)$ and let $f = \sum_{p_1\ldots p_k} f_{p_1\ldots p_k}$ be its Hermite series expansion. Then the following statements are equivalent:*

(i) $$\|\nabla f\| \in L^2(\mathbf{R}^k, \nu_k).$$

(ii) $$\sum_{p_1\cdots p_k} |\mathbf{p}| \|f_{p_1\ldots p_k}\|_{L^2(\nu_k)} < +\infty.$$

(iii) $$f \in \mathbf{D}_1^2(\mathbf{R}^k).$$

PROOF. Note that

$$c_{p_1\ldots p_{j-1}p_jp_{j+1}\ldots p_k}(f) = c_{p_1\ldots p_{j-1}p_j-1p_{j+1}\ldots p_k}(\partial_j f).$$

Hence

$$\|\partial_j f\|^2_{L^2(\nu_k)} = \sum_{p_1\cdots p_k} p_j \frac{|c_{p_1\ldots p_k}(f)|^2}{\mathbf{p}!} = \sum_{p_1\cdots p_k} p_j \|f_{p_1\ldots p_k}\|^2.$$

(i) $\Leftrightarrow$ (ii) then follows from the relation $\|\nabla f\|^2_{L^2} = \sum_{j=1}^k \|\partial_j f\|^2_{L^2(\nu_k)}$. For (ii) $\Leftrightarrow$ (iii), see Proposition 2.2.3. □

2.3.2 Proposition. *Let $f = \sum_{p_1\ldots p_k} f_{p_1\ldots p_k}$ be the Hermite series expansion of $f \in L^2(\nu_k)$. The following properties are equivalent:*

(i) $$\mathcal{L}f \in L^2(\nu_k).$$

(ii) $$\sum_{p_1\cdots p_k} |\mathbf{p}|^2 \|f_{p_1\cdots p_k}\|^2 < +\infty.$$

(iii) $$f \in \mathbf{D}_2^2(\mathbf{R}^k).\square$$

PROOF. See Proposition 2.2.3. □

2.3.3 **Definition.** Let $\nabla^r f = (\partial_{j_1}\partial_{j_2}\ldots\partial_{j_r} f)_{j_1\ldots j_r}$.

If $\partial_{j_1}\partial_{j_2}\ldots\partial_{j_r} f \in L^2(\nu_k)$ for every $j_1\ldots,j_r$, then

$$\|\nabla^r f\|^2_{L^2(\nu_k)} = \sum_{j_1\cdots j_r} \|\partial_{j_1}\ldots\partial_{j_r} f\|^2_{L^2(\nu_k)}.$$

2.3.4 Proposition. *If $f \in \mathbf{D}_2^2(\mathbf{R}^k)$, then*

$$\|\mathcal{L}f\|^2_{L^2} = \|\nabla^2 f\|^2_{L^2} + \|\nabla f\|^2_{L^2}.$$

PROOF. It suffices to check the formula for differentiable functions. In this case, $\mathcal{L} = \sum_j \delta_j\partial_j$, where $\delta_j = -\partial_j^2 + x_j\partial_j$.

Hence

$$\|\mathcal{L}f\|^2_{L^2} = \sum_{\substack{j_1=1,\ldots,k\\ j_2=1,\ldots,k}} (\delta_{j_1}\partial_{j_1} f|\delta_{j_2}\partial_{j_2} f) = \sum_{j_1,j_2} \|\partial_{j_1}\partial_{j_2} f\|^2_{L^2} + \sum_j \|\partial_j f\|^2_{L^2},$$

where we have used 1.3.2(i) and (ii). □

2.4 Approximation of $L^p(\mathbf{R}^N, \nu)$ *by* $L^p(\mathbf{R}^n, \nu)$

Let $X_i : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}$ denote projection onto the ith coordinate and let $\mathcal{B}_n$ be the σ-algebra generated by the random variables $\{X_i\}_{1 \leq i \leq n}$. Then $\{\mathcal{B}_n\}_n$ is a filtration (see 1.6.2).

2.4.1 Lemma. *Let* $f \in L^p(\mathbf{R}^{\mathbf{N}}, \nu)$. *Then*

(i) $f_n = \mathbf{E}^{\mathcal{B}_n}(f)$ *is a martingale in* $L^p(\mathbf{R}^{\mathbf{N}}, \nu)$ *and*

(ii) $\lim_{n\to\infty} \|f_n - f\|_{L^p(\mathbf{R}^{\mathbf{N}})} = 0$. *Thus* f *is the final value of a martingale in* $L^p(\mathbf{R}^{\mathbf{N}})$ *relative to the filtration* $\{\mathcal{B}_n\}$ *(see IV-5.5.2).*

PROOF.

(i) By IV-5.1.5.

(ii) By V-1.8.1 and IV-5.8.1. □

2.5 The spaces $\mathbf{D}^p_r(\mathbf{R}^N)$

Next, starting with the Gaussian Sobolev spaces $\mathbf{D}^p_r(\mathbf{R}^k)$ on the finite-dimensional space $\mathbf{R}^k$ and using the martingale approximation of Lemma 2.4.1, we will study the Gaussian spaces $\mathbf{D}^p_r(\mathbf{R}^{\mathbf{N}})$. If $f \in L^p(\mathbf{R}^{\mathbf{N}}, \nu)$, the function $f_n = E^{\mathcal{B}_n}(f)$ depends only on the first n variables:

$$f_n(x) = \varphi_n(X_1(x), \ldots, X_n(x)) = \varphi_n(x_1, \ldots, x_n).$$

2.5.1 **Definition.** We say that $f \in \mathbf{D}^p_r(\mathbf{R}^{\mathbf{N}})$ if $f_n = \mathbf{E}^{\mathcal{B}_n}(f) \in \mathbf{D}^p_r(\mathbf{R}^{\mathbf{N}})$ for all n and $\sup_n \|f_n\|_{\mathbf{D}^p_r(\mathbf{R}^{\mathbf{N}})} < +\infty$.

In this case, we set $\|f\|_{\mathbf{D}^p_r(\mathbf{R}^{\mathbf{N}})} = \sup_n \|f_n\|_{\mathbf{D}^p_r(\mathbf{R}^{\mathbf{N}})}$.

2.5.2 Operators on $L^2(\mathbf{R}^N, \nu)$

Let $f \in L^2(\mathbf{R}^{\mathbf{N}}, \nu)$ and let its Hermite series decomposition (see 1.6.7) be

$$f(x) = \sum_{\mathbf{p} \in \mathcal{E}} \frac{1}{\mathbf{p}!} c_{\mathbf{p}}(f) \prod_{i=1}^{k} H_{p_i}(x_i),$$

where $\mathbf{p} = (p_1, \ldots, p_k, 0, \ldots, 0)$.

As in 2.1.2, we set

$$\partial_j f(x) = \sum_{\mathbf{p} \in \mathcal{E}} \frac{1}{\mathbf{p}!} c_{\mathbf{p}}(f) \partial_j \left(\prod_{i=1}^{k} H_{p_i}(x_i) \right)$$

and

$$\mathcal{L}f(x) = \sum_{\mathbf{p}\in\mathcal{E}} \frac{1}{\mathbf{p}!} c_{\mathbf{p}}(f) \mathcal{L}\left(\prod_{i=1}^{k} H_{p_i}(x_i)\right).$$

2.5.3 Lemma. *Let* $f \in \mathbf{D}_1^2(\mathbf{R^N})$. *Then*

(*i*) $$\partial_j \mathbf{E}^{\mathcal{B}_n}(f) = \mathbf{E}^{\mathcal{B}_n}(\partial_j f) \quad \textit{if} \quad j \leq n;$$

(*ii*) $$\partial_j \mathbf{E}^{\mathcal{B}_n}(f) = 0 \quad \textit{if} \quad j > n.$$

Proof. Let $f \in L^2(\mathbf{R^N}, \nu)$ and let its Hermite series expansion be

$$f = \sum_{p_1,\ldots,p_k} c_{p_1\ldots p_k}(f) \frac{1}{\mathbf{p}!} H_{p_1}(x_1)\ldots H_{p_k}(x_k).$$

Then

$$\mathbf{E}^{\mathcal{B}_n}(f) = \sum_{\substack{p_1,\ldots,p_k \\ 1\leq k\leq n}} c_{p_1\ldots p_k}(f) \frac{1}{\mathbf{p}!} H_{p_1}(x_1)\ldots H_{p_k}(x_k),$$

since the variables X_i are independent and $\mathbf{E}^{\mathcal{B}_n}(H_p(x_k)) = 0$ if $k > n$. (i) and (ii) follow immediately. □

2.5.4 Lemma. *Let* $f \in \mathbf{D}_1^2(\mathbf{R^N})$ *and let* $f_n = \mathbf{E}^{\mathcal{B}_n}(f)$. *Then, for* $k \in \mathbf{N}$, *the sequence* $\{\partial_k f_n\}_{n\in\mathbf{N}}$ *converges in* $L^2(\mathbf{R^N})$ *and*

(*i*) $$\partial_k f = \lim_{n\to\infty} \partial_k \mathbf{E}^{\mathcal{B}_n}(f).$$

Proof. Let k be fixed. Then $\{\partial_k f_n\}$ is a martingale by 2.5.3(i); by 2.5.1 it is an L^2 martingale, which converges by IV-5.3.3

To prove (i), note that $\partial_k f - \lim_n \partial_k f_n$ is a continuous linear map from $D_1^2(\mathbf{R^N})$ to $L^2(\mathbf{R^N})$ which vanishes on the set of functions depending on a finite number of coordinates. □

2.5.5 Lemma. *Let* $f \in \mathbf{D}_2^2(\mathbf{R^N})$. *Then* $\mathcal{L}f = \lim_{n\to\infty} \mathbf{E}^{\mathcal{B}_n}(\mathcal{L}f)$.

Proof. Check that $\mathbf{E}^{\mathcal{B}_n}(\mathcal{L}f) = \mathcal{L}\mathbf{E}^{\mathcal{B}_n}(f)$ on the Hermite series decomposition of f. □

2.5.6 Theorem. *Let* $f \in L^2(\mathbf{R^N}, \nu)$. *Then the following statements are equivalent:*

(*i*) $f \in \mathbf{D}_1^2(\mathbf{R^N})$.
(*ii*) *For every* k, $k \geq 1$, $\partial_k f \in L^2(\mathbf{R^N}, \nu)$ *and* $\sum_k \|\partial_k f\|_{L^2}^2 < +\infty$.

Furthermore, the space $\mathbf{D}_1^2(\mathbf{R^N})$ *is complete in the metric given by the norm* $\|f\|_{D_1^2}^2 = \|f\|_{L^2}^2 + \sum_k \|\partial_k f\|_{L^2}^2$.

PROOF. (i) $\Rightarrow$ (ii): By 2.5.1,

$$\|f_n\|_{D_1^2}^2 = \|f_n\|_{L^2}^2 + \sum_{j=1}^{n} \|\partial_j f_n\|_{L^2}^2 \leq c,$$

where c is a constant independent of n.

Thus $\sum_{j=1}^{p} \|\partial_j f_n\|_{L^2}^2 \leq c$ for $p \leq n$. As $n \to \infty$, this inequality persists: $\sum_{j=1}^{p} \|\partial_j f\|_{L^2}^2 \leq c$. Letting $p \to \infty$ gives (ii).

(ii) $\Rightarrow$ (i): The same procedure as for (i) $\Rightarrow$ (ii).

To show that $\mathbf{D}_1^2(\mathbf{R}^{\mathbf{N}})$ is complete, let $\{f^{(q)}\}$ be a Cauchy sequence in $\mathbf{D}_1^2(\mathbf{R}^{\mathbf{N}})$ and set $f_n^{(q)} = \mathbf{E}^{\mathcal{B}_n}(f^{(q)})$. Then

$$\|f_n^{(q)} - f_n^{(k)}\|_{\mathbf{D}_1^2(\mathbf{R}^{\mathbf{N}})} \leq \|f^{(q)} - f^{(k)}\|_{\mathbf{D}_1^2(\mathbf{R}^{\mathbf{N}})}.$$

Since $\mathbf{D}_1^2(\mathbf{R}^n)$ is complete, the sequence $\{f_n^{(q)}\}_{q \in \mathbf{N}}$ converges in $\mathbf{D}_1^2(\mathbf{R}^n)$ to f_n. It is straightforward to show that $\{f_n\}_{n \in \mathbf{N}}$ is a martingale associated with the filtration $\{\mathcal{B}_n\}$; it converges to $f \in \mathbf{D}_1^2(\mathbf{R}^{\mathbf{N}})$. $\square$

3 Absolute Continuity of Distributions

3.1 *The Gaussian Space on* $\mathbf{R}$

Let $g : \mathbf{R} \to \mathbf{R}$, $g \in L^2(\mathbf{R})$. We seek sufficient conditions on g for the direct image measure $g_*\nu_1$ to be absolutely continuous with respect to Lebesgue measure on $\mathbf{R}$.

3.1.1 Lemma. *Let θ be a finite positive Borel measure on $\mathbf{R}$. Suppose that, for every φ which is C^1 and bounded on $\mathbf{R}$,*

$$(i) \qquad \left| \int \varphi'(\xi) d\theta(\xi) \right| \leq c \sup_{\xi \in \mathbf{R}} |\varphi(\xi)|.$$

Then θ is absolutely continuous with respect to Lebesgue measure $d\xi$ on $\mathbf{R}$, and its density k is in $L^2(d\xi)$ and satisfies

$$\int k^2 d\xi \leq c\theta(\mathbf{R}) \quad \text{and} \quad k(\xi) \leq c d\xi \quad \text{a.e.}$$

PROOF. Let φ be a bounded increasing C^1 function such that $\varphi(-\infty) = 0$. Then $\varphi(\xi) = \int_{-\infty}^{\xi} \varphi'(u) du$.

It follows from (i) that, for every nonnegative continuous function φ',

$$(ii) \qquad \int \varphi'(\xi) d\theta(\xi) \leq c \int_{-\infty}^{+\infty} \varphi'(u) du.$$

(ii) extends to nonnegative Borel functions.

Hence, for every nonnegative Borel function, $\int f du = 0$ implies $\theta(f) = 0$. By the Radon-Nikodym theorem (IV-6.2.1), θ is absolutely continuous with respect to Lebesgue measure. Let $k(\xi) = \frac{d\theta}{d\xi}$. Then $k(\xi) \geq 0$ and, since

$$\left|\int \varphi'(\xi)k(\xi)d\xi\right| \leq c \sup_{\xi \in \mathbf{R}} |\varphi(\xi)|,$$

$k(\xi) \leq c\ d\xi$ a.e.

It remains to check that the density $k = \frac{d\theta}{d\xi}$ is in $L^2(d\xi)$. This follows from the inequality

$$\int k^2(\xi)d\xi \leq c \int k(\xi)d\xi = c\theta(\mathbf{R}).\square$$

3.1.2 Corollary. *If $g \in L^2(\nu_1)$, $\|dg\|^{-1} \in L^2(\nu_1)$, and $\delta(\frac{1}{dg}) \in L^1(\nu_1)$, then $\mu = g_*\nu_1$ is absolutely continuous with respect to Lebesgue measure, and its density $k = d\mu/d\xi$ is in $L^2(d\xi)$.*

PROOF.

$$\int \varphi'(\xi)d(g_*\nu_1)(\xi) = \int \varphi'(g(x))d\nu_1 = (\varphi'(g(x))|1)$$

and

$$\begin{aligned}(\varphi'(g(x))|1) &= \int d(\varphi \circ g)(x) \times \frac{1}{dg(x)} d\nu_1 \\ &= \left(\varphi \circ g|\delta(\tfrac{1}{dg})\right) \leq \sup_{\xi \in \mathbf{R}} |\varphi(\xi)| \int \left|\delta(\frac{1}{dg})\right| d\nu_1.\end{aligned}$$

The result follows from Lemma 3.1.1. $\square$

3.1.3 Let $g \in \mathbf{D}_2^2(\mathbf{R})$, the Sobolev space of order 2. Let $A = \{x \in \mathbf{R} : dg(x) \neq 0\}$, let $\mathbf{1}_A$ denote the indicator function of A, and let $\mathbf{1}_A\nu_1$ be the density measure $\mathbf{1}_A$ with respect to ν_1.

3.1.4 Theorem. *The image measure $g_*\mathbf{1}_A\nu_1$ is absolutely continuous with respect to Lebesgue measure.*

PROOF. Let $f(x) = \dfrac{dg(x)}{1 + (dg(x))^2}$. Since $g \in \mathbf{D}_2^2(\mathbf{R})$, we have $dg \in \mathbf{D}_1^2(\mathbf{R})$ and $\delta f \in L^2(\mathbf{R})$. Let $A_\epsilon = \{x : dg(x) > \epsilon\}$. When $x \in A_\epsilon$ and $0 < \epsilon < 1$, $dg(x)f(x) > \frac{\epsilon^2}{2}$.

Let ψ be a nonnegative function defined on $\mathbf{R}$. Then

$$\begin{aligned}\int \psi(\xi)g_*(dg(x)f(x)\nu_1) &= \int (\psi \circ g(x))dg(x)f(x)d\nu_1(x) \\ &\geq \frac{\epsilon^2}{2} \int_{A_\epsilon} (\psi \circ g(x))d\nu_1(x) = \frac{\epsilon^2}{2} \int \psi(\xi)(g_*(\mathbf{1}_{A_\epsilon}\nu_1)).\end{aligned}$$

Suppose that $\rho_A = g_*(\mathbf{1}_A \nu_1)$ is not absolutely continuous with respect to Lebesgue measure; then there exists a compact subset K of $\mathbf{R}$ such that $\int_K d\xi = 0$ and $\rho_A(K) > 0$.

Since $A = \cup_n A_{1/n}$,

$$\rho_A(K) = \int \mathbf{1}_K(g(x))\mathbf{1}_A(x)d\nu_1(x) = \lim_{n\to\infty} \int \mathbf{1}_K(g(x))\mathbf{1}_{A_{\frac{1}{n}}}(x)d\nu_1(x) > 0.$$

Hence there exists ϵ such that $\rho_{A_\epsilon}(K) > 0$. Let $\{u_n\}$ be a sequence of continuous functions on $\mathbf{R}$ such that (i) $0 \leq u_n \leq 1$, (ii) $\lim_{n\to\infty} u_n(\xi) = \mathbf{1}_K(\xi)$, and (iii) for some R, $u_n(\xi) = 0$ if $|\xi| \geq R$.

Set $\varphi_n(\xi) = \int_{-R}^{\xi} u_n(\lambda)d\lambda$. Then, by the dominated convergence theorem,

$$\lim_{n\to\infty} \varphi_n(\xi) = \int_{-R}^{\xi} \mathbf{1}_K(\lambda)d\lambda = 0.$$

Moreover,

$$(\varphi_n \circ g|\delta f) = ((\varphi_n' \circ g)dg|f) \geq \frac{\epsilon^2}{2}\int (\varphi_n' \circ g)\cdot \mathbf{1}_{A_\epsilon}\nu_1.$$

Since $(\varphi_n \circ g|\delta f) \to 0$ and $\int \varphi_n'(\xi)\rho_{A_\epsilon}(d\xi) \to \rho_{A_\epsilon}(K)$, this gives a contradiction. □

3.2 The Gaussian space on $\mathbf{R}^N$

Let $g = (g_1, \ldots, g_d) \in L^2(\mathbf{R}^{\mathbf{N}}, \nu)$ be a function with values in $\mathbf{R}^d$. We now seek sufficient conditions for the direct image measure $g_*\nu$ to be absolutely continuous with respect to Lebesgue measure on $\mathbf{R}^d$.

3.2.1 **Notation.** If $g = (g_1, \ldots, g_d)$ is such that $g_k \in \mathbf{D}_1^2(\mathbf{R}^{\mathbf{N}})$ for $k = 1, \ldots, d$, we set

$$\nabla g_k = (\partial_1 g_k, \partial_2 g_k, \ldots, \partial_j g_k, \ldots).$$

By 2.5.6, $\sum_{j=1}^{\infty} \|\partial_j g_k\|_{L^2(\nu)}^2 < +\infty$.

3.2.2 Lemma. *If*

$$\sum_{j=1}^{\infty} \|\partial_j g_k\|_{L^2(\nu)}^2 < +\infty \quad \textit{and} \quad \sum_{j=1}^{\infty} \|\partial_j g_p\|_{L^2(\nu)}^2 < +\infty,$$

then

$$\sum_{j=1}^{\infty} (\partial_j g_k|\partial_j g_p)_{L^2(\nu)} < +\infty.$$

PROOF. This follows immediately from Hölder's inequality. □

3.2.3 **Notation.** We set

(i) $|\nabla g_k|^2(x) = \sum_{j=1}^{\infty} |\partial_j g_k|^2(x)$,

(ii) $(\nabla g_k|\nabla g_p)(x) = \sum_{j=1}^{\infty} \partial_j g_k(x)\partial_j g_p(x)$.

The series (i) and (ii) are convergent in $L^1(\nu)$.

3.2.4 **Definition.** The matrix $\sigma_{ik} = ((\nabla g_i|\nabla g_k)(x))_{\substack{i=1,\dots,d\\k=1,\dots,d}}$ is called the covariance matrix associated with g.

3.2.5 Lemma. *If $g = (g_1, \dots, g_d) : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}^d$ and, for $k = 1, \dots, d$, $g_k \in \mathbf{D}_1^2(\mathbf{R}^{\mathbf{N}})$, then*

$$\sigma_{ik} \in L^1(\mathbf{R}^{\mathbf{N}}, \nu).$$

PROOF. This follows from Hölder's inequality. □

3.2.6 **Notation.** Let $g = (g_1, \dots, g_d) : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}^d$. Suppose that $g_i \in \mathbf{D}_1^2(\mathbf{R}^{\mathbf{N}})$ for $i = 1, \dots, d$ and that the inverse matrix σ_{ik}^{-1} exists ν-a.e. We set

(i) $z_{jk} = \sum_{i=1}^{d} \sigma_{ik}^{-1} \partial_j g_i$,
(ii) $z_k = (z_{1k}, z_{2k}, \dots, z_{jk}, \dots)$.

Lemma. *If $g_i \in \mathbf{D}_1^4(\mathbf{R}^{\mathbf{N}})$ for $i = 1, \dots, d$ and $\sigma_{ik}^{-1} \in L^4(\mathbf{R}^{\mathbf{N}})$, then $\sum_j \|z_{jk}\|_{L^2(\nu)}^2 < +\infty$ and, for every C^1 function $\varphi : \mathbf{R}^d \to \mathbf{R}$,*

$$(i) \qquad \sum_j \partial_j(\varphi \circ g)(x) z_{jk}(x) = \frac{\partial \varphi}{\partial \xi_k}(g(x)).$$

PROOF. $\partial_j(\varphi \circ g)(x) = \sum_{p=1}^{d} \frac{\partial \varphi}{\partial \xi_p}(g(x))\partial_j g_p(x)$, and hence

$$\partial_j(\varphi \circ g)(x) z_{jk}(x) = \sum_{p=1}^{d} \frac{\partial \varphi}{\partial \xi_p}(g(x)) \sum_{i=1}^{d} \sigma_{ik}^{-1} \partial_j g_i \partial_j g_p.$$

This implies (i). □

3.2.7 **Definition.** When $g_i \in \mathbf{D}_2^4(\mathbf{R}^{\mathbf{N}})$ for $i = 1, \dots, d$ and $\sigma_{ik}^{-1} \in \mathbf{D}_1^4(\mathbf{R}^{\mathbf{N}})$, we set

$$\delta z_k = \sum_{i=1}^{d} (\mathcal{L} g_i) \sigma_{ik}^{-1} - \sum_{i=1}^{d} (\nabla g_i | \nabla(\sigma_{ik}^{-1})).$$

3.2.8 Theorem. *For every function $\psi \in \mathbf{D}_1^2(\mathbf{R}^{\mathbf{N}})$,*

$$\sum_j (\partial_j \psi | z_{jk})_{L^2(\nu)} = (\psi | \delta z_k)_{L^2(\nu)}.$$

PROOF. $\sum_j (\partial_j \psi | z_{jk})_{L^2} = \left(\psi | \sum_{i=1}^d \delta_j(\sigma_{ik}^{-1} \partial_j g_i)\right)$ and, recalling that $\partial_j = \frac{d}{dx_j}$ and $\delta_j = -\frac{d}{dx_j} + x_j$, one can easily show that

$$\delta_j(f \partial_j g_i) = f \delta_j \partial_j g_i - \partial_j f \partial_j g_i.$$

The relation follows by summing over j. □

3.2.9 Proposition. *Let $\varphi : \mathbf{R}^d \to \mathbf{R}$ and let g be such that $g_i \in \mathbf{D}_2^4(\mathbf{R}^\mathbf{N})$ for $i = 1, \ldots, d$ and $\sigma_{ik}^{-1} \in \mathbf{D}_1^4(\mathbf{R}^\mathbf{N})$. Then*

$$\int_{\mathbf{R}^d} \frac{\partial \varphi}{\partial \xi_k} d(g_* \nu) = \int (\varphi \circ g)(x) \delta z_k(x) d\nu(x).$$

PROOF.

$$\int_{\mathbf{R}^d} \frac{\partial \varphi}{\partial \xi_k}(\xi) d(g_* \nu) = \int \frac{\partial \varphi}{\partial \xi_k}(g(x)) d\nu(x) = \sum_j (\partial_j(\varphi \circ g) | z_{jk})_{L^2(\nu)} = (\varphi \circ g | \delta z_k).$$

The last equality follows from 3.2.8. □

3.2.10 Lemma. *Let θ be a finite measure on $\mathbf{R}^d$. Suppose that there exists a constant C such that*

$$(i) \qquad \left| \int \frac{\partial \varphi(\xi)}{\partial \xi_i} d\theta(\xi) \right| \leq C \sup_{\xi \in \mathbf{R}^d} |\varphi(\xi)|, \quad i = 1, \ldots, d,$$

for every bounded C^1 function φ on $\mathbf{R}^d$.

Then θ is absolutely continuous with respect to Lebesgue measure $d\xi$ on $\mathbf{R}^d$.

PROOF. For the case $d = 1$, see 3.1.1. We prove the lemma when $d = 2$.

Let φ be a compactly supported C^1 function on $\mathbf{R}^2$. We first show that

$$(ii) \quad \left(\int \int_{\mathbf{R}^2} |\varphi|^2 dx_1 dx_2 \right)^{1/2} \leq \frac{1}{2} \left(\int \left| \frac{\partial \varphi}{\partial x_1} \right| dx_1 dx_2 + \int \left| \frac{\partial \varphi}{\partial x_2} \right| dx_1 dx_2 \right).$$

To see this, let

$$v(x_1) = \sup_{x_2 \in \mathbf{R}} |\varphi(x_1, x_2)| \quad \text{and} \quad w(x_2) = \sup_{x_1 \in \mathbf{R}} |\varphi(x_1, x_2)|.$$

Then

$$(iii) \qquad \int \int_{\mathbf{R}^2} |\varphi|^2 dx_1 dx_2 \leq \int v(x_1) dx_1 \int w(x_2) dx_2.$$

Since

$$(iv) \qquad v(x_1) \leq \int_{-\infty}^{+\infty} \left| \frac{\partial \varphi}{\partial x_2}(x_1, x_2) \right| dx_2$$

and

$$(v) \qquad w(x_2) \leq \int_{-\infty}^{+\infty} \left| \frac{\partial \varphi}{\partial x_1}(x_1, x_2) \right| dx_1,$$

(ii) is proved.

Let u be a nonnegative continuous function with compact support such that

$$(vi) \qquad \int\int_{\mathbf{R}^2} u(x_1, x_2) dx_1 dx_2 = 1.$$

For $\epsilon > 0$, we set $u_\epsilon(x) = \frac{1}{\epsilon^2} u(\frac{x}{\epsilon})$ and

$$\varphi_\epsilon(x) = \int_{\mathbf{R}^2} u_\epsilon(x - \lambda)\theta(d\lambda).$$

For every continuous compactly supported function ψ,

$$\begin{aligned} \int_{\mathbf{R}^2} \varphi_\epsilon(x)\psi(x)dx &= \int\int_{\mathbf{R}^4} u\left(\frac{x-\lambda}{\epsilon}\right)\frac{1}{\epsilon^2}\psi(x)dx\ \theta(d\lambda) \\ &= \int\int_{\mathbf{R}^4} u(z)\psi(\lambda + \epsilon z)dz\ \theta(d\lambda). \end{aligned}$$

Since ψ is continuous, $\psi(\lambda + \epsilon z)$ tends to $\psi(\lambda)$ as $\epsilon \to 0$. It follows that

$$(vii) \qquad \lim_{\epsilon \to 0} \int \varphi_\epsilon(x)\psi(x)dx = \int \psi(\lambda)\theta(d\lambda).$$

The measures $\varphi_\epsilon(x)$ thus converge vaguely to $\theta(dx)$ as $\epsilon \to 0$.

If ψ is C^1,

$$\left|\int \frac{\partial \varphi_\epsilon}{\partial x_i}\psi dx\right| = \left|\int \varphi_\epsilon \frac{\partial \psi}{\partial x_i} dx\right|.$$

It follows from (i) and (vii) that

$$\left|\int \frac{\partial \varphi_\epsilon}{\partial x_i}\psi dx\right| \leq C \sup_{\lambda \in \mathbf{R}^2} |\psi(\lambda)|.$$

Hence

$$(viii) \qquad \int_{\mathbf{R}^2} \left|\frac{\partial \varphi_\epsilon}{\partial x_i}\right|^2 dx \leq C.$$

For every ϵ, by Hölder's inequality,

$$\left|\int \varphi_\epsilon \psi dx\right| \leq \left[\int |\varphi_\epsilon|^2\right]^{1/2} \left[\int |\psi|^2\right]^{1/2}.$$

It follows that

$$\lim_{\epsilon \to 0}\left|\int \varphi_\epsilon \psi dx\right| = \left|\int \psi\theta(d\lambda)\right| \leq C\left(\int |\psi|^2\right)^{1/2}.$$

The mapping $\psi \mapsto \int \psi(\lambda)\theta(d\lambda)$ is thus a continuous linear functional on $L^2(dx)$. This implies the existence of $k \in L^2(dx)$ such that

$$\int \psi(\lambda)\theta(d\lambda) = \int \psi(x)k(x)dx. \square$$

3.2.11 Principal theorem. *Let $g = (g_1, \ldots, g_d) : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}^d$ be such that $g_i \in \mathbf{D}_2^4(\mathbf{R}^{\mathbf{N}})$ for $i = 1, \ldots, d$.*

Let $\sigma_{ik} = (\nabla g_i | \nabla g_k)$ be the covariance matrix. Suppose that $\sigma^{-1} \in \mathbf{D}_2^4(\mathbf{R}^{\mathbf{N}})$. Then the image measure $g_* \nu$ is absolutely continuous with respect to Lebesgue measure on $\mathbf{R}^d$.

PROOF. By 3.2.9,

$$\left| \int_{\mathbf{R}^d} \frac{\partial \varphi}{\partial \xi}(\xi) d(g_* \nu) \right| \leq \sup_{\xi \in \mathbf{R}^d} |\varphi(\xi)| \cdot \int |\delta z_k(x)| d\nu(x).$$

Let $C = \sup_k \int |\delta z_k(x)| d\nu(x)$; then $C < +\infty$ and hypothesis 3.2.10(i) is satisfied. $\square$

Appendix I

Hilbert Spectral Analysis

The spectral theorem in finite dimensions makes it possible to write a Hilbert space as a direct sum of eigenspaces of a hermitian endomorphism u. If the dimension is infinite, direct sums are replaced by "continuous sums". We will apply Bochner's theorem to obtain the spectral theorem by Fourier analysis.

1 Functions of Positive Type

Let f be a function defined on an abelian group G. f is said to be *of positive type* if, for any given $g_1, \ldots, g_n \in G$, the matrix

$$(f(g_i - g_j)), \quad 1 \leq i, j \leq n,$$

is *positive hermitian*. That is,

$$\sum_{j,k} \lambda_j \overline{\lambda}_k f(g_j - g_k) \geq 0, \quad \forall \lambda_1, \ldots, \lambda_n \in \mathbf{C}.$$

In particular, taking a single element, we find that the matrix

$$\begin{pmatrix} f(0) & f(g) \\ f(-g) & f(0) \end{pmatrix}$$

is positive hermitian. That is,

1.1. $$f(0) > 0, \quad \overline{f}(g) = f(-g), \quad \text{and} \quad |f(g)|^2 \leq f(0).$$

Let $P(G)$ denote the set of functions of positive type on G. Observe that $P(G)$ is a cone:

$$\lambda f + \mu h \in P(G) \quad \forall f, h \in P(G) \text{ and } \lambda, \mu \in \mathbf{R}^+.$$

1.2 Proposition. *Let Γ be an abelian group and let $\widehat{\Gamma}$ be its dual. Then*

$$(M_+^1(\Gamma))^\wedge \subset P(\widehat{\Gamma}).$$

PROOF. Let $\mu \in M_+^1(\Gamma)$, $\lambda_j \in \mathbf{C}$. Then, writing $\langle \gamma, \widehat{\gamma} \rangle$ for $\widehat{\gamma}(\gamma)$,

$$q(\lambda) = \sum \widehat{\mu}(\widehat{\gamma}_j - \widehat{\gamma}_k)\lambda_j \overline{\lambda}_k = \sum \int_\Gamma \langle \gamma, \widehat{\gamma}_j - \widehat{\gamma}_k \rangle \lambda_j \overline{\lambda}_k d\mu(\gamma).$$

But

$$\sum \lambda_i \overline{\lambda}_k \langle \gamma_i, \widehat{\gamma}_j - \widehat{\gamma}_k \rangle = \sum \lambda_i \overline{\lambda}_j \langle \gamma, \gamma_j \rangle \overline{\langle \gamma, \gamma_k \rangle} = \left| \sum \lambda_i \langle \gamma, \gamma_i \rangle \right|^2,$$

whence

$$q(\lambda) = \int_\Gamma \left| \sum \lambda_j \langle \gamma, \gamma_j \rangle \right|^2 d\mu(\gamma) \geq 0. \square$$

Algebra structure of the cone of functions of positive type

Proposition. *Let f and h be functions of positive type on the abelian group G. Then their product fh is of positive type.*

PROOF. Set $k = fh$ and let $g_1, \ldots, g_n \in G$ be given. We consider the matrix

$$k(g_i - g_j) = f(g_i - g_j)h(g_i - g_j)$$

and apply the following lemma.

1.4 Lemma. Let (A_j^i) and (B_j^i), $1 \leq i, j \leq n$, be positive hermitian matrices. Let

$$C_j^i = A_j^i B_j^i, \quad 1 \leq i, j \leq n.$$

Then C_j^i is a positive hermitian matrix.

PROOF. Let X_α^i (respectively Y_β^i) be an orthonormal system of eigenvectors of A (respectively B), and let μ_α (respectively γ_β) be the corresponding eigenvalue. Then

$$A_j^i = \sum_\alpha \mu_\alpha X_\alpha^i \overline{X}_\alpha^j \quad \text{and} \quad B_j^i = \sum_\beta \gamma_\beta Y_\beta^i \overline{Y}_\beta^j.$$

Hence

$$C_j^i = \sum_{\alpha,\beta} \mu_\alpha \gamma_\beta X_\alpha^i Y_\beta^i \overline{X}_\alpha^j \overline{Y}_\beta^j.$$

Set $Z_{\alpha,\beta}^i = X_\alpha^i Y_\beta^i$ and $\rho_{\alpha,\beta} = \mu_\alpha \gamma_\beta$. Then

$$C_j^i = \sum \rho_{\alpha,\beta} Z_{\alpha,\beta}^i \overline{Z}_{\alpha,\beta}^j.$$

Since

$$\sum \lambda_i \overline{\lambda}_j Z_{\alpha,\beta}^i \overline{Z}_{\alpha,\beta}^j = |\sum \lambda_i Z_{\alpha,\beta}^i|^2,$$

the matrix $Z_{\alpha,\beta}^i \overline{Z}_{\alpha,\beta}^j$ is positive. C is thus a linear combination, with positive coefficients, of positive matrices, and therefore is positive. $\square$

2 Bochner's Theorem

Bochner's Theorem. *Let* $\mathbf{Z}$ *be the group of integers. A function* f *on* $\mathbf{Z}$ *is of positive type if and only if there exists* $\mu \in M_+(\mathbf{T})$ *such that* $\widehat{\mu}(n) = f(n)$.

PROOF. ($\Leftarrow$) This follows from 1.2.

($\Rightarrow$) Consider

$$g_r(n) = r^{|n|} \quad \text{where} \quad r \in [0,1).$$

Then $\widehat{P}_r(n) = g_r(n)$, where $P_r(\theta)$ denotes the Poisson kernel (see III-2.2.1), and thus $g_r \in P(\mathbf{Z})$. By 1.3, $k_r = fg_r \in P(\mathbf{Z})$. Moreover, by 1.1,

$$|k_r(n)| \leq |f(0)| r^{|n|}.$$

Set

$$(i) \qquad \tilde{k}_r(\theta) = \sum_n k_r(n) \mathrm{e}^{-in\theta}.$$

The right-hand side is an absolutely convergent series and $\tilde{k}_r(\theta) \in C(\mathbf{T})$.

Next, let $\lambda_p = \mathrm{e}^{-ip\theta}$ if $|p| < N$ and $\lambda_p = 0$ otherwise. Then, since $k_r \in P(\mathbf{Z})$,

$$(ii) \qquad 0 \leq G_N(\theta) = \frac{1}{2N-1} \sum \lambda_p \overline{\lambda}_q k_r(p-q), \quad \forall N \in \mathbf{Z}.$$

We now rewrite $G_N(\theta)$ in a slightly different form by noting that $\lambda_p \overline{\lambda}_q = \mathrm{e}^{i(q-p)\theta}$ and summing over $p - q = n$:

$$G_N(\theta) = \sum_{|n|<2N-1} \left(1 - \frac{|n|}{2N-1}\right) \mathrm{e}^{-in\theta} k_r(n).$$

Letting $N \to +\infty$, the absolute convergence of (i) and inequality (ii) show that

$$(iii) \qquad \tilde{k}_r(\theta) \geq 0.$$

A positive linear functional can thus be defined on $C(\mathbf{T})$ by setting

$$l_r(u) = \int_0^{2\pi} u(\theta) \tilde{k}_r(\theta) \frac{d\theta}{2\pi}.$$

Integrating the series in (i) term by term yields

$$(iv) \qquad \|l_r\| = l_r(1) = \int_0^{2\pi} \tilde{k}_r(\theta) \frac{d\theta}{2\pi} = f(0).$$

Moreover, $l_r(\mathrm{e}^{iq\theta}) = f(q) r^{|q|} \to f(q)$ as $r \to 1$.

Hence, if $Q(\theta)$ is a trigonometric polynomial,

$$(v) \qquad \lim_{r\to 1} l_r(Q) \text{ exists.}$$

Since the trigonometric polynomials are dense in $C(\mathbf{T})$ (III-2.2.8) and the l_r are equicontinuous by (iv), it follows that

$$\lim_{r\to 1} l_r(u) \text{ exists for every } u \in C(\mathbf{T})$$

and defines a positive linear functional, that is a Radon measure $\mu \in M_+(\mathbf{T})$. In particular,

$$\widehat{\mu}(n) = \lim_{r\to 1} l_r(e^{in\theta}) = f(n).\square$$

3 Spectral Measures for a Unitary Operator

Let H be a complex Hilbert space, with hermitian inner product $(h_1|h_2)$ and norm $(h|h) = \|h\|^2$. A linear operator U is called *unitary* if it is invertible and $U^* = U^{-1}$. Recall that the adjoint A^* of a linear operator A is defined by the identity

$$(Ah_1|h_2) = (h_1|A^*h_2).$$

Theorem on existence of spectral measures. *Let U be a unitary operator on the Hilbert space H. For a trigonometric polynomial $P(\theta) = \sum C_m e^{im\theta}$, let*

$$(i) \qquad P(U) = \sum C_m U^m.$$

Given $h \in H$, there exists a unique $\mu_h \in M_+(\mathbf{T})$ such that, for any trigonometric polynomial P,

$$(ii) \qquad (P(U)h|h) = \int_{\mathbf{T}} P(\theta)d\mu_h(\theta).$$

μ_h is called the spectral measure of U relative to h.

PROOF. To prove the uniqueness of μ_h, write (ii) for a trigonometric polynomial consisting of a single monomial. This gives

$$\widehat{\mu}(m) = \gamma(m), \quad \text{where} \quad \gamma(m) = (U^m h|h),$$

and uniqueness follows from III-2.2.8. To prove existence it suffices, using Bochner's theorem, to prove that $\gamma(m)$ is a function of positive type on $\mathbf{Z}$. We must therefore consider the sign of

$$I = \sum_{p,q} \lambda_p \overline{\lambda}_q \gamma(p-q).$$

Since U is unitary, $U^{-1} = U^*$, whence $\gamma(p-q) = (U^p h|U^q h)$. Thus $\lambda_p \overline{\lambda}_q \gamma(p-q) = (\lambda_p U^p h|\lambda_q U^q h)$ and

$$I = \sum_{p,q} (\lambda_p U^p h|\lambda_q U^q h).$$

But this can be written

$$I = \left(\sum_p \lambda_p U^p h|\sum_q \lambda_q U^q h\right) = \left|\sum_r \lambda_r U^r h\right|^2 \geq 0. \square$$

4 Spectral Decomposition Associated with a Unitary Operator

Theorem. *Let U be a unitary operator on the Hilbert space H. Let $\mathcal{L}^\infty(B_{\mathbf{T}})$ be the algebra of bounded complex-valued functions which are measurable with respect to the Borel algebra of $\mathbf{T}$. Then there exists an algebra homomorphism*

$$\Phi : \mathcal{L}^\infty(\mathbf{T}) \to \mathit{End}(H)$$

that associates the operator U with the function $e^{i\theta}$ and preserves conjugation. That is,

$$(\Phi(f))^* = \Phi(\overline{f}).$$

PROOF. Recall that the scalar product on H can be obtained from the norm by the following *polarization identity*:

$$4(h_1|h_2) = \|h_1 + h_2\|^2 - \|h_1 - h_2\|^2 + \|ih_1 + h_2\|^2 - \|ih_1 - h_2\|^2.$$

Polarized spectral measures are defined by setting

$$4\gamma_{h_1,h_2} = \mu_{h_1+h_2} - \mu_{h_1-h_2} + \mu_{ih_1+h_2} - \mu_{ih_1-h_2}.$$

Thus, for every trigonometric polynomial P, it follows from polarizing 3(ii) that

$$(i) \qquad (P(U)h_1|h_2) = \int_{\mathbf{T}} P(\theta)d\gamma_{h_1,h_2}(\theta).$$

Fixing $f \in \mathcal{L}^\infty(\mathcal{B}_{\mathbf{T}})$, we define a sesquilinear functional q_f by

$$q_f(h_1, h_2) = \int_{\mathbf{T}} f(\theta)d\gamma_{h_1,h_2}(\theta).$$

This integral is well defined since f is a bounded Borel function. We have the following upper bound:

$$|q_f(h_1, h_2)| \leq 4\|h_1\| \; \|h_2\| \; \|f\|_{\mathcal{L}^\infty}.$$

Hence fixing h_1 gives a conjugate linear functional in h_2, and this form is represented by a scalar product. There exists a bounded linear operator $\Phi(f)$ such that

$$(ii) \qquad (\Phi(f)h_1|h_2) = \int_{\mathbf{T}} f d\gamma_{h_1,h_2}, \quad \forall h_1, h_2 \in \mathrm{H}.$$

Moreover, when f_n converges to f while remaining bounded, Lebesgue's dominated convergence theorem shows that

$$(iii) \qquad (\Phi(f_n)h_1|h_2) \to (\Phi(f)h_1|h_2).$$

In order to show that Φ is an algebra homomorphism, it suffices, using (iii), to check the assertion for trigonometric polynomials. In this case, (ii) and (i) show that $\Phi(P) = P(U)$, and the formula

$$\Phi(P_1P_2) = \Phi(P_1)\Phi(P_2)$$

clearly holds. Finally, by the polarization identity, $\gamma_{h_1,h_2} = \overline{\gamma}_{h_1,h_2}$, which implies that

$$(\Phi(f)h_1|h_2) = (h_1|\Phi(\overline{f})h_2). \square$$

Corollary. *Let $A \in \mathcal{B}_{\mathbf{T}}$. Then $\Phi(\mathbf{1}_A)$ is an orthogonal projection and*

$$\Phi(\mathbf{1}_A)\Phi(\mathbf{1}_B) = \Phi(\mathbf{1}_{A\cap B}).$$

PROOF. $(\Phi(\mathbf{1}_A))^* = \overline{\Phi(\mathbf{1}_A)} = \Phi(\mathbf{1}_A)$ and $(\Phi(\mathbf{1}_A))^2 = \Phi(\mathbf{1}_A^2) = \Phi(\mathbf{1}_A)$. These properties characterize orthogonal projections. $\square$

Corollary (Spectral decomposition). *Let $\Gamma(H)$ denote the set of closed vector subspaces of H. Let $\Gamma(H)$ be given the structure of an abstract Boolean algebra, with products given by intersections and complements by orthogonal complements. Then Φ defines a homomorphism φ from the Boolean algebra $\mathcal{B}_{\mathbf{T}}$ to $\Gamma(H)$ by setting*

$$\varphi(A) = \textit{Image of } \Phi(\mathbf{1}_A).$$

Moreover,

$$U(\varphi(A)) \subset \varphi(A).$$

5 Spectral Decomposition for Several Unitary Operators

Let $U_1, \dots, U_n$ be n pairwise-commuting unitary operators on the same Hilbert space H:

$$U_k U_l = U_l U_k, \quad 1 \leq k, l \leq n.$$

With every trigonometric polynomial

$$P(\theta_1, \dots, \theta_n) = \sum c_{m_1,\dots,m_n} \mathrm{e}^{im_1\theta_1 + \dots + im_n\theta_n}$$

on $\mathbf{T}^n$, we associate the operator

$$P(U_1, \dots, U_n) = \sum c_{m_1,\dots,m_n} U_1^{m_1} \dots U_n^{m_n}.$$

5.1 Theorem on existence of spectral measures. *To every $h \in H$ there corresponds a positive measure μ_h on $\mathbf{T}^n$ such that*

$$(P(U_1, \dots, U_n)h|h) = \int_{\mathbf{T}^n} P(\theta) d\mu_h(\theta).$$

This is proved by generalizing Bochner's theorem from $\mathbf{Z}$ to $\mathbf{Z}^n$. Theorem 5.1 leads to the *simultaneous spectral decomposition* of the operators U_k, $1 \leq k \leq n$, i.e. a representation of $\mathcal{L}^\infty(B_{\mathbf{T}^n})$ in End(H).

Appendix II

Infinitesimal and Integrated Forms of the Change-of-Variables Formula

In this appendix, we give a new proof of Theorem II-4.4. The variational method used here, coupled with the ideas of Chapter V, yields a proof in the setting of Gaussian spaces.

1 Notation

Let μ be a Borel measure on Euclidean space $\mathbf{R}^n$. Let $\{T_t : t \in [0,1]\}$ be a family of $\mathbf{R}^n$-valued measurable mappings, defined on an open set D of $\mathbf{R}^n$ and with the following properties:

(i) $T_t : D \to D' \subset \mathbf{R}^n$ is a diffeomorphism. The inverse diffeomorphism is denoted by A_t.

(ii) $\forall x \in D$ the mapping $t \to T_t x$ is differentiable. The differential is denoted by $\left(\frac{d}{dt} T_t\right)(x)$.

(iii) $\forall t \in [0,1]$ the direct image $(A_t)_* \mu$ under A_t of the measure μ is absolutely continuous with respect to μ. The density is denoted by $G_t = \dfrac{d((A_t)_* \mu)}{d\mu}$.

Let $f : \mathbf{R}^n \to \mathbf{R}^p$ be differentiable. $J_f(x)$ denotes the Jacobian of f at the point x.

1.1 **Definition.** The vector fields

$$Z_t(y) = \left(\frac{d}{dt}T_t\right)(A_t y)$$

are called velocity fields associated with $(T_t)_{t\in[0,1]}$.

REMARK. $y \mapsto Z_t(y)$ defines not only a vector field on D' but also a differentiable mapping from D' to $\mathbf{R}^n$.

1.2 **Definition.** Let Z_t be a vector field defined on D. Z is said to admit a divergence with respect to μ if there exists a function $\delta_\mu Z : D \to \mathbf{R}$ such that

$$\int J_f(x)(Z(x))dx = -\int f(x)\delta_\mu Z(x)dx$$

for every differentiable function $f : \mathbf{R}^n \to \mathbf{R}$ with support contained in D.

2 Velocity Fields and Densities

2.1 Theorem. *Let Z_t be the velocity field associated with T_t. Then the density $G_t(x) = \dfrac{d((A_t)_*\mu)}{d\mu}$ is given by*

$$G_t(x) = G_0(x)\exp\left[\int_0^t \delta_\mu(Z_s)(T_s y)ds\right] \qquad a.e.\ d\mu.$$

PROOF.

$$(i) \qquad \int_D f(x)G_t(x)d\mu(x) = \int_{D'} f(A_t y)d\mu(y).$$

Differentiating with respect to t gives

$$\frac{d}{dt}f(A_t y) = J_f(A_t y)\frac{d}{dt}A_t(y).$$

Furthermore,

$$(ii) \qquad J_{(f\circ A_t)}(y) = J_f(A_t y)J_{A_t}(y),$$

and hence

$$(iii) \qquad \frac{d}{dt}f(A_t y) = J_{(f\circ A_t)}(y)J_{A_t}(y)^{-1}\frac{d}{dt}A_t(y).$$

Since

$$(iv) \qquad (T_t \circ A_t)y = y,$$

we have

$$J_{T_t}(A_t y) J_{A_t} y = Id,$$

whence $(J_{A_t} y)^{-1} = J_{T_t}(A_t y)$. Differentiating (iv) with respect to t gives

$$(v) \qquad \left(\frac{d}{dt} T_t\right)(A_t y) = -J_{T_t}(A_t y) \frac{d}{dt} A_t y.$$

Substituting into (iii), we find that

$$\frac{d}{dt} f(A_t y) = -J_{(f \circ A_t)}(y) \left(\frac{d}{dt} T_t\right)(A_t y)$$

and

$$(vi) \qquad \begin{aligned} \int_{D'} \frac{d}{dt} f(A_t y) d\mu(y) &= \int_{D'} (f \circ A_t)(y)(\delta_\mu Z_t)(y) d\mu(y) \\ &= \int_D f(x)(\delta_\mu Z_t)(T_t x) G_t(x) d\mu(x), \end{aligned}$$

where the first equality follows from Definition 1.2 and the second from (i).
Differentiating (i) with respect to t shows that

$$\frac{d}{dt} G_t(x) = (\delta_\mu Z_t)(T_t x) \cdot G_t(x) \quad \mu\text{-a.e.}$$

□

2.2 Corollary. *Let $\mu = dx$ be Lebesgue measure on $\mathbf{R}^n$ and suppose that $T_0 = Id$. Then $\forall t \in [0,1]$*

$$\int_D f(T_t x) |det\ J_{T_t} x| dx = \int_{D'} f(x') dx', \quad \textit{where} \quad D' = T_t(D).$$

PROOF. It suffices to verify the relation

$$\frac{d}{dt} \log \det J_{T_t} x = \delta(Z_t)(T_t x),$$

where

$$Z_t(y) = \left(\frac{d}{dt} T_t\right)(A_t y)$$

and δZ is the divergence of Z with respect to dx. To do this, we use the following two lemmas.

2.3 Lemma.

$$\frac{d}{dt}(J_{T_t})(y) = (J_{Z_t})(T_t y) \circ J_{T_t} y.$$

PROOF.

$$\frac{d}{dt}(J_{T_t})(y) = J_{(\frac{d}{dt} T_t)}(y) = J_{(Z_t \circ T_t)}(y) = (J_{Z_t})(T_t y) \circ J_{T_t} y. \square$$

2.4 Lemma. *Let* (B_t) *be* $n \times n$ *matrices such that* $\frac{d}{dt}B_t = M_t B_t$*, where the* (M_t) *are also* $n \times n$ *matrices. Then* $\frac{d}{dt}\log|\det B_t| = \operatorname{trace} M_t$.

PROOF. Let $\Phi_i(t)$ be the ith column of B_t. It follows from

$$\frac{d}{dt}\Phi_i(t)_j = \sum_k (M_t)_{ik}\Phi_k(t)_j$$

that

$$\frac{d}{dt}\Phi_i(t) = \sum_k (M_t)_{ik}\Phi_k(t).$$

Hence

$$\frac{d}{dt}\det[\Phi_1(t), \ldots, \Phi_n(t)] = \sum_i (M_t)_{ii}\det[\Phi_1(t), \ldots, \Phi_n(t)].\square$$

CONCLUSION OF THE PROOF OF COROLLARY 2.2.

$$\delta(Z_t)(T_t) = \operatorname{trace}(J_{Z_t})(T_t x).\square$$

REMARKS. (1) Compare 2.2 with II-4.4.1, the change-of-variables theorem.

(2) Let $T_t = I + tM$, where M is an $n \times n$ matrix. Suppose that $I + tM$ is invertible for every $t \in [0, 1]$. Then $\frac{d}{dt}(I + tM) = M(I + tM)^{-1}(I + tM)$. Letting Λ denote the exterior product, we can express the determinant of $A + I$ as $\Delta(t) = \det(I + tM) = \sum_{k=0}^{n}(\operatorname{trace}\Lambda^k M)t^k$. By 2.4,

$$\frac{\Delta'}{\Delta}(t) = \operatorname{trace} M(I + tM)^{-1}.$$

Thus

$$\det\,(I + tM) = \exp\int_0^t \operatorname{trace}\,(M(I + sM)^{-1})ds.$$

2.5 Corollary.

$$\frac{d}{dt}[\text{vol }(T_t(D))] = \int_{T_t(D)} \delta Z_t(y)dy.$$

PROOF. By 2.2,

$$\text{vol }(T_t(D)) = \int_D |\det\, J_{T_t}(x)|\; dx$$

and

$$\begin{aligned}\frac{d}{dt}[\text{vol }(T_t(D))] &= \int_D \left[\frac{d}{dt}\log\det\, J_{T_t}(x)\right] \times |\det\, J_{T_t}(x)|dx \\ &= \int_D (\delta Z_t)(T_t x)|\det\, J_{T_t}(x)|dx.\end{aligned}$$

Applying 2.2 once more proves the assertion. $\square$

3 The n-dimensional Gaussian Space

Let $\mathbf{R}^n$ be given the measure $\mu = \prod_{i=1}^n \frac{\exp\left(-\frac{x_i^2}{2}\right) dx_i}{\sqrt{2\pi}}$ and let $(x|y) = \sum_{i=1}^n x_i y_i$ denote the scalar product of two vectors x, $y \in \mathbf{R}^n$.

3.1 Lemma. *Let Z be a differentiable vector field on $\theta \subset \mathbf{R}^n$. Then $\forall x \in \theta$*

$$\delta_\mu Z(x) = trace\ J_Z(x) - (Z(x)|x).$$

3.2 Theorem. *Let $(T_t)_{t\in[0,1]}$ be the mappings defined in Section 1 and let $H_t(x) = \frac{d((A_t)_*\mu)}{d\mu}$. Then*

$$H_t(x) = H_0(x) det\ |J_{T_t}(x)| \exp\left[-\int_0^t (Z_s(T_s x)|T_s x)\right] ds.$$

PROOF. This follows from Lemma 3.1 and Theorem 2.

EXAMPLE. Translations of the Gaussian space.

For a differentiable mapping $h : \mathbf{R}^n \to \mathbf{R}^n$, set $Tx = x - h(x)$ and $T_t x = x - th(x)$. Let A_t be the inverse of T_t. Then $A_t x = x + th(A_t x)$.

The velocity fields associated with T_t are

$$Z_t(x) = \left(\frac{d}{dt} T_t\right)(A_t x) = -h(A_t x).$$

We have

$$(Z_s(T_s x)|T_s x) = -(h(x)|x) + s(h(x)|h(x))$$

and

$$\exp\left[-\int_0^t (Z_s(T_s x)|T_s x) ds\right] = \exp\left[t(h(x)|x) - \frac{t^2}{2}(h(x)|h(x))\right].$$

Compare this with the Cameron-Martin theorem (V-1.8.3). In particular, if $Tx = x - y$ and $T_t x = x - ty$, then $A_t x = x + ty$, $\det(J_{T_t}(x)) = 1$, and

$$\frac{d((A_t)_*\mu)}{d\mu} = \exp\left[t \sum_{i=1}^n x_i y_i - \frac{t^2}{2} \sum_{i=1}^n y_i^2\right].$$

REMARK. This method can be extended to the infinite-dimensional Gaussian space.

Exercises for Chapter I

Problem I-1. If $\mathcal{G}$ is a family of subsets of a set X, we denote by $a(\mathcal{G})$ the Boolean algebra generated by $\mathcal{G}$ and by $\sigma(\mathcal{G})$ the σ-algebra generated by $\mathcal{G}$. A *partition* of X is a family $P = \{P_j\}_{j\in J}$ of nonempty subsets of X such that $P_i \cap P_j = \emptyset$ if $i \neq j$ and $\cup_{j\in J} = X$.
(1) Let $P = \{P_j\}_{j\in J}$ be a partition of X. Characterize
  (a) $a(P)$ if J is finite,
  (b) $a(P)$ if J is infinite,
  (c) $\sigma(P)$ if J is finite or countable, and
  (d) $\sigma(P)$ if J is uncountably infinite.
(2) Show that the family $\mathcal{A}$ of subsets of X is a Boolean algebra generated by a finite number of elements if and only if there exists a partition $P = \{P_j\}_{j\in J}$, with J finite, such that $\mathcal{A} = a(P)$.
(3) Let $\mathcal{A}$ be a σ-algebra on a countable set X. Show that there exists a partition P of X such that $\mathcal{A} = \sigma(P)$.
(4) Show that a σ-algebra never has a countable number of elements.

Problem I-2. Let $\mathcal{G}$ be a family of subsets of a set X such that $X \in \mathcal{G}$ and $\mathcal{G}$ is closed under finite intersections. An *r-family* is a family $\mathcal{R}$ of subsets of X which is closed under finite intersections of pairwise disjoint sets and such that, if B_1 and $B_2 \in \mathcal{R}$ with $B_1 \subset B_2$, then $B_2 \setminus B_1 \in \mathcal{R}$. Let $r(\mathcal{G})$ be the smallest r-family containing $\mathcal{G}$. Show that $r(\mathcal{G})$ equals the Boolean algebra $a(\mathcal{G})$ generated by $\mathcal{G}$.

METHOD. Consider the families

$$\begin{array}{rcl}\mathcal{R}_1 & = & \{B : B \in r(\mathcal{G}) \text{ and } A \cap B \in r(\mathcal{G})\ \forall A \in \mathcal{G}\} \text{ and} \\ \mathcal{R}_2 & = & \{B : B \in r(\mathcal{G}) \text{ and } A \cap B \in r(\mathcal{G})\ \forall A \in r(\mathcal{G})\},\end{array}$$

and show that they are r-families.

Problem I-3. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two nonempty families of subsets of a set X which are closed under finite intersections. Let $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}$ denote the σ-algebras generated by $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_1 \cup \mathcal{G}_2$, respectively. Let P be a measure of total mass 1 on $(X, \mathcal{A})$. Show that if

$$P(A_1 \cap A_2) = P(A_1)P(A_2) \quad \text{for all } A_1 \in \mathcal{G}_1 \text{ and } A_2 \in \mathcal{G}_2,$$

then the same equality holds for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$.
METHOD. Consider the families

$$\begin{array}{rcl}\mathcal{M}_1 & = & \{A : A \in \mathcal{A} \text{ and } P(A \cap A_2) = P(A)P(A_2)\ \forall A_2 \in \mathcal{G}_2\} \text{ and} \\ \mathcal{M}_2 & = & \{A : A \in \mathcal{A} \text{ and } P(A_1 \cap A) = P(A_1)P(A)\ \forall A_1 \in \mathcal{A}_1\},\end{array}$$

and apply the theorem on monotone classes, using Problem I-2.

REMARKS. 1. This result is especially useful in probability theory. Thus, if $X = \mathbf{R}^2$, $A_1(x) = \{(x_1, x_2) : x_1 < x\}$, and $A_2(y) = \{(x_1, x_2) : x_2 < y\}$, then $\mathcal{G}_1 = \{A_1(x) : x \in \mathbf{R}\}$ and $\mathcal{G}_2 = \{A_2(y) : y \in \mathbf{R}\}$ are closed under finite intersections and $\mathcal{A}$ is the set of Borel subsets of $\mathbf{R}^2$. If P is a probability measure on $(\mathbf{R}^2, \mathcal{A})$, it is the distribution of a pair (X_1, X_2) of real random variables. By Problem II-3, (X_1, X_2) is a pair of independent random variables if and only if

$$P[X_1 < x;\ X_2 < y] = P[X_1 < x]\ P[X_2 < y]$$

for all $(x, y) \in \mathbf{R}^2$.
2. The result can be extended from two factors to n factors by constructing monotone classes $\mathcal{M}_k$ for $k = 1, 2, \ldots, n$ and using induction on k.

Problem I-4. Let $x = \{x_n\}_{n=0}^{\infty}$ and let

$$\ell^\infty = \left\{ x : x_n \in \mathbf{R}\ \ \forall n \in \mathbf{N} \text{ and } \|x\|_\infty = \sup_n |x_n| < \infty \right\}.$$

Define $T : \ell^\infty \to \ell^\infty$ by $(Tx)_0 = x_0$ and $(Tx)_n = x_n - x_{n-1}$ if $n > 0$.
(1) If $e = (1, 1, \ldots, 1, \ldots)$, show that the equation $Tx = e$ has no solution x in ℓ^∞.
(2) Let $F = T\ell^\infty$ be the image of T. Assume without proof that there exists a continuous linear functional f on ℓ^∞ such that $f(x) = 0$ for every x in F, $f(e) = 1$, and $\sup\{|f(x)| : \|x\|_\infty \leq 1\} < +\infty$ (Hahn-Banach theorem). Show that if $x = \{x_n\}_{n=0}^{\infty}$ is such that $x_n \geq 0$ for every n, then $f(x) \geq 0$.
(3) Let $S : \ell^\infty \to \ell^\infty$ be defined by $(Sx)_n = x_{n+1}$ if $n \geq 0$. Show that $f(x) = f(Sx)$ for every x in ℓ^∞.

(4) Show that $\liminf_{n\to+\infty} x_n \geq 0$ implies that $f(x) \geq 0$. Conclude that $\liminf_{n\to+\infty} x_n \leq f(x) \leq \limsup_{n\to+\infty} x_n$ for every $x \in \ell^\infty$.
(5) Let $A \subset \mathbf{N}$ and let $\mathbf{1}_A \in \ell^\infty$ be defined by $\mathbf{1}_A(n) = 0$ if $n \neq A$ and $\mathbf{1}_A(n) = 1$ if $n \in A$. If $P(A) = f(\mathbf{1}_A)$, show that $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$ and that P does not satisfy the countable additivity axiom.

REMARKS. The linear functional f above is called a *Banach limit*; it cannot be written down explicitly since it is constructed by means of the Hahn-Banach theorem and the axiom of choice. Similarly, it is impossible to give an explicit example of an additive but not σ-additive measure on a σ-algebra.

Problem I-5. Let X be an uncountable set and let $\mathcal{A}$ be the σ-algebra generated by the family of 1-element subsets of X. (See Problem 1, question (1d).) Let $P : \mathcal{A} \to [0,1]$ be defined by

$$\begin{array}{lll} P(A) = 0 & \text{if} & A \text{ is finite or countable} \\ P(A) = 1 & \text{if} & A \text{ is cocountable.} \end{array}$$

(A is cocountable if A^c is finite or countable.) Show that P is a probability measure on $(X, \mathcal{A})$.

Problem I-6. Let $(X, \mathcal{A}, \mu)$ be a measure space and let f be a nonnegative measurable function on X. For every $t \geq 0$, set

$$F(t) = \mu\{x : f(x) > t\} \quad \text{and} \quad G(t) = \mu\{x : f(x) \geq t\}.$$

(1) Assume that $f(X) \subseteq \mathbf{N}$ and that f is integrable. Prove that

$$\int_X f(x)d\mu(x) = \sum_{n=0}^{\infty} F(n) = \sum_{n=1}^{\infty} G(n).$$

METHOD. Set $\mu_n = \mu\{x : f(x) = n\}$ and show that $\int_X f(x)\mu(dx) = \sum_{n=0}^{\infty} n\mu_n$.
(2) Assume that f^α is integrable for $\alpha > 0$. Prove that

$$\int_X f^\alpha(x)d\mu(x) = \alpha \int_0^{+\infty} t^{\alpha-1}F(t)dt = \alpha \int_0^{+\infty} t^{\alpha-1}G(t)dt.$$

METHOD. Show that (2) holds for $\alpha = 1$ by considering the functions $f_n(x) = \frac{[2^n f(x)]}{2^n}$, where $[a]$ means "the greatest integer $\leq a$", and using the monotone convergence theorem. The general case can then be reduced to the case $\alpha = 1$.

Problem I-7. If $0 < r < 1$, we write the Poisson kernel as

$$P_r(\theta) = 1 + 2\sum_{n=1}^{\infty} r^n \cos n\theta = \frac{1 - r^2}{1 - 2r\cos\theta + r^2}.$$

(1) Show that $r^2 + \cos\theta(1-2r) \geq 0$ if $0 \leq \theta \leq \pi$ and $\frac{1}{2} \leq r \leq 1$. Deduce that $\theta^2 P_r(\theta) \leq \frac{(1-r^2)\theta^2}{1-\cos\theta}$ and evaluate $\lim_{r\to 1}\int_0^\pi \theta^2 P_r(\theta)d\theta$.
(2) Show that $\int_0^\pi \theta^2 P_r(\theta)d\theta = \frac{\pi^3}{3} + 4\pi\sum_{n=1}^\infty \frac{(-r)^n}{n^2}$ and use this to derive another expression for $\lim_{r\to 1}\int_0^\pi \theta^2 P_r(\theta)d\theta$.
(3) Use (1) and (2) to find the sums of the series $\sum_{n=1}^\infty \frac{(-1)^n}{n^2}$, $\sum_{n=1}^\infty \frac{1}{(2n-1)^2}$, and $\sum_{n=1}^\infty \frac{1}{n^2}$.
(4) Express $\int_0^1 (\log(1-x))^2 \frac{dx}{x^2}$ as the sum of a double series and show that $\int_0^1 (\log(1-x^2))^2 \frac{dx}{x^2} = 2\sum_{n=1}^\infty \frac{1}{n^2}$.

Problem I-8. Evaluate $\sum_{n=1}^\infty \frac{(-1)^{n+1}}{n}$ by using the integral $\int_0^1 \frac{dx}{1+x}$ and the monotone convergence theorem.

Problem I-9. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $x \mapsto f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$ be a measurable mapping from X to $\mathbf{R}^n$. Suppose that $\mathbf{R}^n$ is equipped with a norm $\|\ \|$ such that $x \mapsto \|f(x)\|$ is integrable.
(1) Show that f_j is integrable for every $j = 1, 2, \ldots, n$.
(2) Defining $\int_X f(x)d\mu(x)$ in $\mathbf{R}^n$ by

$$\left(\int_X f_1(x)\mu(dx), \ldots, \int_X f_1(x)\mu(dx)\right),$$

show that $\|\int_X f(x)\mu(dx)\| \leq \int_X \|f(x)\|\mu(dx)$.

METHOD. On the dual space $(\mathbf{R}^n)^*$ consisting of linear functionals $a : v \mapsto \langle a, v\rangle$ on $\mathbf{R}^n$, introduce the dual norm $\|a\|^* = \sup_{v\neq 0}\frac{|\langle a,v\rangle|}{\|v\|}$ and use the fact that $\|v\| = \sup_{a\neq 0}\frac{|\langle a,v\rangle|}{\|a\|^*}$.

REMARKS. 1. *The shortest path between two points is a straight line.* Consider $\mathbf{R}^n$ with the Euclidean norm $\|v\| = \left[v_1^2 + v_2^2 + \cdots v_n^2\right]^{1/2}$. Let $X = [0,1]$ with Lebesgue measure. (See Chapter II.) Let F be a function from $[0,1]$ to $\mathbf{R}^n$ such that the derivative $f = F'$ exists everywhere and is continuous. Then $\int_0^1 \|f(x)\|dx$ can be interpreted as the Euclidean length of the curve described by F, and $\|\int_0^1 f(x)dx\| = \|F(1) - F(0)\|$ is the length of the line segment with endpoints $F(0)$ and $F(1)$.
2. *Case of equality.* It can be shown that, when the unit ball B is strictly convex (that is, when $\|v_1\| = \|v_2\| = \|\lambda v_1 + (1-\lambda)v_2\| = 1$ for $\lambda \in [0,1]$ holds only for $\lambda = 0$ or 1), the inequality is strict unless there exist $v \in \mathbf{R}^n$ and a function $g(x) \geq 0$ such that $f(x) = g(x)v$ μ-almost everywhere. The application to the Euclidean length of a curve is immediate.

Problem I-10. Let $X, X_1, \ldots, X_n, \ldots$ be measurable functions from a space $(E, \mathcal{E}, \mu)$ to an open set Ω of Euclidean space $\mathbf{R}^d$ such that

$$\forall\epsilon > 0 \quad \mu(\{\|X_n - X\| \geq \epsilon\}) \to 0 \quad \text{as } n \to \infty.$$

(1) Show that $\forall \epsilon > 0$ there exists a compact set $K \subset \Omega$ such that $\mu(\{X \notin K\}) \leq \epsilon$ and, for every n, $\mu(\{X_n \notin K\}) \leq \epsilon$.
(2) If $f : \Omega \to \mathbf{R}^m$ is continuous, then $\forall \epsilon > 0$

$$\mu(\{\|f(X_n) - f(X)\| \geq \epsilon\}) \to 0 \quad \text{as} \quad n \to \infty.$$

Problem I-11. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces such that $\mu(X)$ and $\nu(Y) > 0$. Let $a : X \to \mathbf{C}$ and $b : Y \to \mathbf{C}$ be functions, respectively $\mathcal{A}$ and $\mathcal{B}$ measurable, such that

$$a(x) = b(y) \quad \mu \otimes \nu\text{-almost everywhere on } X \times Y.$$

Show that there exists a constant λ such that $a(x) = \lambda$ μ-a.e. and $b(y) = \lambda$ ν-a.e.

Problem I-12. On a measure space $(X, \mathcal{A}, \mu)$, let f and g be complex functions such that $|f|^2$ and $|g|^2$ are μ-integrable and consider the function

$$h(x, y) = |f(x)g(y) - f(y)g(x)|^2.$$

(1) Show that $0 \leq \int_{X \times X} h(x, y) d\mu(x)\ d\mu(y)$, and use this to prove the Cauchy-Schwarz inequality:

$$\left| \int_X f(x)\overline{g(x)} d\mu(x) \right|^2 \leq \int_X |f(x)|^2 d\mu(x) \int_X |g(x)|^2 d\mu(x).$$

METHOD. Consider first the case where $f \geq 0$ and $g \geq 0$.
(2) Show that equality holds in Schwarz's inequality if and only if either $g(x) = 0$ μ-a.e. on X or there exists a constant $\lambda \in \mathbf{C}$ such that $f(x) - \lambda g(x) = 0$ μ-a.e. on X.

METHOD. Problem I-11 can be used.

Problem I-13. If X and Y are measurable real-valued functions defined on the measure space $(\Omega, \mathcal{A}, \mu)$ such that $\mu(\{Y \leq x < X\}) = 0$ for all real x, show that $\mu(\{Y < X\}) = 0$.

Problem I-14. Let $(X, \mathcal{A}, \mu)$ be a measure space, where $\mu(X)$ is not necessarily finite, let $(Y, \mathcal{B})$ be a measurable space, and let f be a measurable mapping from X to Y. Suppose that there exists a sequence $\{B_n\}$ in $\mathcal{B}$ such that $\cup_{n=1}^{\infty} B_n = y$ and $\mu(f^{-1}(B_n)) < \infty$ for every n.
(1) Show that $\nu(B) = \mu(f^{-1}(B))$ defines a measure ν on $(Y, \mathcal{B})$ (called the image of μ under f).
(2) Show that if $g \in L^1(\nu)$, then

$$\int_X g(f(x))\mu(dx) = \int_Y g(y)\nu(dy).$$

REMARKS. 1. The image measure always exists when μ is bounded; this is used extensively in probability theory, in Chapter IV. It does not always

exist if $\mu(X) = +\infty$. For example, if $X = \mathbf{R}^2$ is equipped with Lebesgue measure $\mu = dx\, dy$ and $f : \mathbf{R}^2 \to \mathbf{R} = Y$ is the projection $f(x, y) = x$, the image of μ does not exist.
2. If X and Y are metrizable locally compact spaces which are countable at infinity and μ is a Radon measure on X, a sufficient condition for existence of the image measure is that, for every compact set K in Y, $f^{-1}(K)$ should be relatively compact. See problems II-11, 12, and 13 and III-3.

Problem I-15. (1) Let f be square integrable on $[0, 1]$ and let $F(x) = \int_0^x f(t)dt$. Applying the Cauchy-Schwarz inequality to the product $f \times 1$ on $[0, x]$, show that $\lim_{x \downarrow 0} x^{-1/2} F(x) = 0$.
(2) Let g be square integrable on $[0, +\infty)$ and let $G(x) = \int_0^x g(t)dt$. Applying the Cauchy-Schwarz inequality to the product $g \times 1$ on $[a, x]$, with a sufficiently large, show that $\lim_{x \to +\infty} x^{-1/2} G(x) = 0$.

REMARK. It is easy to replace L^2 by L^p, with $p > 1$. If $\frac{1}{p} + \frac{1}{q} = 1$, we find that $x^{-1/q} F(x) \to 0$ as $x \to 0$ and $x^{-1/q} G(x) \to 0$ as $x \to +\infty$.

Exercises for Chapter II

Problem II-1. Let I be an *open* interval in $\mathbf{R}$, equipped with the Borel algebra $\mathcal{B}$. A function $F : I \longrightarrow \mathbf{R}$ is called increasing if $x < y$ implies that $F(x) \leq F(y)$. We set $F(x-0) = \lim_{y \uparrow x} F(y)$, $F(x+0) = \lim_{y \downarrow x} F(y)$, and $D_F = \{x : F(x-0) \neq F(x+0)\}$.
(1) If $F : I \longrightarrow \mathbf{R}$ is increasing, prove that D_F is finite or countable.

METHOD. If $[a,b] \subset I$, show that $D(n;[a,b]) = \{x \in [a,b] : F(x+0) - F(x-0) \geq \frac{1}{n}\}$ has a finite number of elements.

(2) If $F : I \longrightarrow \mathbf{R}$ is increasing, prove that there exists exactly one measure $\mu \geq 0$ on $(I, \mathcal{B})$ such that

$$F(y) - F(x) = \mu([x,y])$$

for all x, y such that $[x,y] \subset I$ and x, $y \notin D_F$.

Prove that $\mu(\{a\}) = F(a+0) - F(a-0)$ for every a in I.

METHOD. *Uniqueness:* Use the fact (II-3.2) that a Borel measure that is locally finite on an interval is regular, and hence determined by its values on open sets.
Existence: Imitate the construction of the Riemann integral. For every continuous function f with support contained in I, define the integral $\int f d\mu$ as the limit of integrals of step functions

$$\sum_i g(x_i)(F(x_i) - F(x_{i-1})).$$

(3) Let μ be a locally finite nonnegative measure on $(I, \mathcal{B})$ and let $x_0 \in I$. Set $F(x) = \mu([x_0, x))$ if $x > x_0$ and $F(x) = -\mu([x, x_0))$ if $x \leq x_0$. Show that F is increasing and that $F(y) - F(x) = \mu([x, y])$ if $y \notin D_F$.
(4) Let a relation on the set of increasing functions on I be defined as follows: $F_1 \sim F_2$ if there exists a finite or countable subset $D_{1,2}$ of I such that $F_1(y) - F_1(x) = F_2(y) - F_2(x)$ for all x and $y \in I \setminus D_{1,2}$. Show that this defines an equivalence relation on the set of increasing functions on I. Characterize the equivalence classes in terms of measure.

REMARKS. 1. Since perhaps as many as 90 per cent of the measures used in practice are measures on $\mathbf{R}$, a description of all the Radon measures ≥ 0 on an open interval is important. Historically, the first measures ≥ 0 were considered by Stieltjes, precisely by means of increasing functions.
2. With every increasing function F on an open interval I, we can thus associate a measure $\mu(dx)$, which is often written $dF(x)$ or $F(dx)$. Conversely, given a measure $\mu \geq 0$ on I, an increasing function F satisfying the hypotheses of part (2) is called a *distribution function* for μ. As we have seen, a distribution function for μ is not unique; we can modify (slightly) its value at points of discontinuity (the atoms of μ) and add an arbitrary constant. When μ is a probability measure on $\mathbf{R}$, there are three traditional choices for distribution functions:

$$F_1(x) = \mu((-\infty, x)), \quad F_2(x) = \mu((-\infty, x]), \quad \text{and} \quad F_3(x) = \frac{1}{2}[F_1(x) + F_2(x)].$$

The third appears in the inversion formula for a characteristic function.
3. If we consider a measure $\mu \geq 0$ on a closed interval of the form $(-\infty, b]$, $[a, +\infty)$, or $[a, b]$, we can define its distribution function as above. However, two measures can then have the same distribution function but different masses at the endpoints of the interval.
4. Many identities and inequalities use increasing functions on an interval. It is essential to express the latter in terms of measures in order to understand the former; this also gives a systematic method of proof, although not necessarily the shortest.

Problem II-2. Specify for which measure on the open interval I each of the following increasing functions is the distribution function (see Problem II-1).

(1) $I = \mathbf{R}$
  (a) $F(x) = x$ (b) $F(x) = [x]$ (c) $F(x) = \frac{1}{\pi} \arctan x$
(2) $I = (-1, +1)$
  (a) $F(x) = \tan \frac{\pi x}{2}$ (b) $F(x) = (\text{sign} x)|x|^{1/2}$ (c) $F(x) = \frac{1}{\pi} \arcsin x$
(3) $I = (0, +\infty)$
  (a) $F(x) = \log x$ (b) $F(x) = -[\frac{1}{x}]$ (c) $F(x) = (x-1)^+$

(Notation: $[a] = \sup\{n : n \in \mathbf{Z} \text{ and } n \leq a\}$, $a^+ = \sup\{0, a\}$, and sign $a = +1$ if $a > 0$, sign $0 = 0$, and sign $a = -1$ if $a < 0$.)

Problem II-3. Let I be an open interval in $\mathbf{R}$. A function G is called convex if its right derivative $\lim_{\epsilon\downarrow 0}[G(x+\epsilon)-G(x)] = G'_+(x)$ exists for every x in I and the function $x \mapsto G'_+(x)$ is increasing. (See I-9.2.1.)

Prove that G is convex if and only if there exists an increasing function F on I such that, for every x_0 in I,

$$G(x) - G(x_0) = \int_{x_0}^{x} F(t)dt.$$

METHOD. For one direction, show that $G'_+(x) = \lim_{\epsilon\downarrow 0} F(x+\epsilon)$. For the other, consider $H(x) = \int_{x_0}^{x} G'_+(x)dt$ and use without proof the fact that, if a function has a right derivative that is zero in an open interval I, it is constant in I.

REMARK. It can be shown that the definition of convex functions given here is equivalent to the following property:

$$G[(\lambda x + (1-\lambda)y] \leq \lambda G(x) + (1-\lambda)G(y) \quad \text{if } x,\, y \in I \text{ and } \lambda \in [0,1].$$

For a proof of this equivalence and further details of convex functions, the reader may consult Artin[1] or Zygmund[2].

Problem II-4. Let I be an *open* interval in $\mathbf{R}$. Recall (see Problem II-3) that a function $G : I \to \mathbf{R}$ is called *convex* if there exists an increasing function F on I such that, for every x_0 in I,

$$G(x) - G(x_0) = \int_{x_0}^{x} F(t)dt.$$

If μ is the measure on I given by the distribution function F (see Problem II-2), prove the following assertions.
(1) If $x_0 \leq x$, with x and $x_0 \in I$, then

$$\begin{aligned} G(x) - G(x_0) &= (x-x_0)F(x_0+0) + \int_I \mathbf{1}_{(x_0,x]}(u)(x-u)\mu(du) \\ &= (x-x_0)F(x_0-0) + \int_I \mathbf{1}_{[x_0,x]}(u)(x-u)\mu(du). \end{aligned}$$

(2) If $x_0 \geq x$, with x_0 and $x \in I$, then

$$\begin{aligned} G(x) - G(x_0) &= (x-x_0)F(x_0+0) - \int_I \mathbf{1}_{[x,x_0]}(u)(x-u)\mu(du) \\ &= (x-x_0)F(x_0-0) - \int_I \mathbf{1}_{[x,x_0)}(u)(x-u)\mu(du). \end{aligned}$$

---

[1]E. Artin, *The Gamma Function* (New York: Holt, Rinehart and Winston 1964), 1–6.
[2]A. Zygmund, *Trigonometric Series* (Cambridge: Cambridge University Press 1959), 21–26.

REMARKS. If μ has no atoms and $x_0 \leq x$, we can replace the notation $\int_I \mathbf{1}_{(x_0,x]}(u)g(u)du = \int_I \mathbf{1}_{[x_0,x]}(u)g(u)du$ by $\int_{x_0}^{x} g(u)\mu(du)$, since the latter is unambiguous in this case. If $x \leq x_0$, we write $\int_{x_0}^{x} g(u)\mu(du) = -\int_I \mathbf{1}_{[x,x_0]}(u)\mu(du)$, which permits us to state the relation of Chasles: $\int_a^c = \int_a^b + \int_b^c$ for arbitrary a, b, and c in I. However, this relation does not hold if μ has atoms.

Problem II-5. Let M_1 be the set of measures $\mu \geq 0$ on $(0, +\infty)$ equipped with its Borel algebra, such that $\int_0^{\infty} \mathbf{1}_{[x,+\infty)}(u)u\mu(du) < \infty$ for every $x > 0$.
(1) Let G be a convex function on $(0, +\infty)$ (see Problem II-4) such that $\lim_{x\to+\infty} G(x) = 0$. Prove that there exists a unique μ in M_1 such that

$$(i) \qquad G(x) = \int_0^{+\infty} (u-x)^+ \mu(du) \quad \text{for every } x > 0,$$

where $a^+ = \max(0, a)$, and that $\int_0^{+\infty} u\mu(du) = \lim_{x\to 0} G(x) \leq +\infty$.
(2) Conversely, let $\mu \in M_1$. Show that (i) defines a convex function G on $(0, +\infty)$ such that $\lim_{x\to+\infty} G(x) = 0$.

METHOD. Let $F(x)$ be as in Problem II-4 and show that $F(x) \leq 0$ and that $\lim_{x\to+\infty} xF(x) = 0$. Then use Problem II-4.

REMARK. The measure $x\mu(dx)$ is not necessarily bounded: $G(x) = \frac{1}{x}$ gives $\mu(dx) = \frac{dx}{x^3}$.

Problem II-6. Let M be the set of measures $\nu \geq 0$ on $(0, +\infty)$ equipped with its Borel algebra, such that $\nu([x, +\infty)) < +\infty$ for every $x > 0$. If k is a positive integer, we denote by C_k the set of functions g defined on $(0, +\infty)$ such that $G(x) = (-1)^{k-1}g^{(k-1)}(x)$ exists and is convex and also that $\lim_{x\to+\infty} g(x) = \lim_{x\to+\infty} G(x) = 0$.
(1) If $g \in C_k$, show that there exists a unique ν in M such that

$$(i) \qquad g(x) = \int_0^{\infty} \left[\left(1 - \frac{x}{u}\right)^+\right]^k \nu(du) \quad \text{for every } x > 0.$$

(2) Conversely, let $\nu \in M$. Show that (i) defines an element of C_k.

METHOD. (1) First use Taylor's formula to show that $\lim_{x\to+\infty} g^{(j)}(x) = 0$ for $j = 0, 1, \ldots, k-1$, then use Problem II-5.

REMARK. It is clear that the functions $f_u(x) = \left[(1 - \frac{x}{u})^+\right]^k$ play the role of extremals in C_k; formula (i) shows that the functions in C_k are "barycenters" of the f_u. Formula (i) plays a role in the probability distributions of Polya and Askey. (See Problem III-5.)

Problem II-7. Let u be a decreasing function defined on $(0, +\infty)$ such that $u \to 0$ as $x \to +\infty$ and $\int_0^{\infty} x^2 u(x)dx < \infty$. Show that, for every $y > 0$,

$$y^2 \int_y^{+\infty} u(x)dx \leq \frac{4}{9} \int_0^{+\infty} x^2 u(x)dx \quad \text{(K.F. Gauss)}.$$

Describe in detail the case of equality.

METHOD. Consider a measure μ on $(0, +\infty)$ for which $-u$ is a distribution function.

Problem II-8. Let u be a decreasing function defined on $(-a, +\infty)$, with $a > 0$, such that $u \to 0$ as $x \to +\infty$ and $\int_{-a}^{+\infty} u(x)dx < +\infty$. Show that

$$\int_0^y u(x)dx \leq \frac{y}{y+a} \int_{-a}^{+\infty} u(x)dx \quad \forall y > 0,$$

and describe in detail the case of equality.

METHOD. Consider a measure μ on $(-a, +\infty)$ for which $-u$ is a distribution function.

Problem II-9. Let F be an increasing function on $[a, b]$ and let f be an integrable function on $[a, b]$. Show that there exists a number ξ in $[a, b]$ such that

$$\int_a^b f(x)F(x)dx = F(a) \int_a^\xi f(x)dx + F(b) \int_\xi^b f(x)dx.$$

(Second mean value theorem for integrals)

METHOD. Show that this can be reduced to the case where $F(a) = 0$ and $F(b) = 1$, and consider a probability measure μ on $[a, b]$ such that $F(x) = \mu([a, x])$ for $x \notin D_F = \{x : a < x < b \text{ and } F(x - 0) < F(x + 0)\}$.

Problem II-10. Let μ be a probability measure on $[0, 1]$. Set $m = \int_0^1 x\mu(dx)$ and $\sigma^2 = \int_0^1 x^2\mu(dx) - m^2$. Show that $\sigma^2 \leq \frac{1}{4}$. Describe in detail the case of equality.

Problem II-11. Let f be a positive decreasing function on $(0, 1]$ such that $\int_0^1 f(x)dx = 1$, and let $\lambda \in [0, 1]$. Let $P(dx) = \lambda\delta_0(dx) + (1 - \lambda)f(x)dx$, where δ_0 is the Dirac measure at the origin, let $m(\lambda, f) = \int_0^1 xP(dx)$, and let $\sigma^2(\lambda, f) = \int_0^1 x^2P(dx) - m^2(\lambda, f)$.
(1) Show that $\sigma^2(\lambda, f) \leq 1/9$. Describe in detail the case of equality.
(2) Show that $\sigma^2(0, f) < 1/9$. Is this inequality the best possible?

METHOD. If D_f is the set of points of discontinuity of f in $(0, 1]$, consider the measure ν on $(0, 1]$ such that $f(x) = \nu([x, 1])$ if $x \notin D_f$ and show that $\mu(dt) = t\nu(dt)$ is a probability measure on $(0, 1]$.

REMARK. If G is a convex function from $(0, 1)$ to $[0, 1)$, it can be shown that the measure P on $[0, 1)$ which is the image under G of Lebesgue measure on $(0, 1)$ is of the type considered in the problem. Hence

$$\int_0^1 G^2(x)dx = \left[\int_0^1 G(x)dx\right]^2 \leq \frac{1}{9}.$$

Problem II-12. Let n be a positive integer and let $\alpha, a_1, \dots, a_n, c_1, \dots, c_n$ be real numbers such that $a_1 < a_2 < \dots < a_n$ and $c_j > 0$ for $j = 1, \dots, n$. Let $\overline{\mathbf{C}}$ and $\overline{\mathbf{R}}$ denote the complex and the real numbers completed by a point at infinity ∞. Consider the function $f : \overline{\mathbf{C}} \to \overline{\mathbf{C}}$ defined by $f(x) = \infty$ if $x \in \{\infty, a_1, \dots, a_n\}$ and

$$f(x) = x + \alpha - \sum_{j=1}^{n} \frac{c_j}{x - a_j} \quad \text{if} \quad x \notin \{\infty, a_1, \dots, a_n\}.$$

The function $T : \overline{\mathbf{R}} \to \overline{\mathbf{R}}$ is the restriction of f to $\overline{\mathbf{R}}$. Lebesgue measure on $\overline{\mathbf{R}}$ is the measure m such that $m(\{\infty\}) = 0$ and the restriction of m to $\mathbf{R}$ is the usual measure.

(1) Let $y \in \mathbf{R}$. Show that the equation in x given by $f(x) = y$ has exactly $n + 1$ real roots $\{x_j(y)\}_{j=0}^{n}$ such that $a_j < x_j(y) < a_{j+1}$ (with the convention that $a_0 = -\infty$ and $a_{n+1} = +\infty$). Show that $\sum_{j=0}^{n} x'_j(y) = 1$ and conclude that T preserves m. That is, for every F in $L^1(m)$,

$$\int_{\overline{\mathbf{R}}} F(T(x))m(dx) = \int_{\overline{\mathbf{R}}} F(x)m(dx).$$

(2) Prove by induction on the integer $k \geq 0$ that, for every $z \in \mathbf{C}$,

$$\sum_{j=0}^{n} [x_j(y) - z]^{-k-1} x'_j(y) = \frac{1}{k!} \left(\frac{\partial}{\partial z} \right)^k [y - f(z)]^{-1}.$$

(3) Let g be a nonnegative rational function such that $\int_{\overline{\mathbf{R}}} g(x)m(dx) < \infty$. Prove that there exists a rational function g_1 with the same properties and such that the image $g(x)m(dx)$ under T is $g_1(x)m(dx)$. Conclude from (2) that, if z_1 is a pole of g_1 with multiplicity $m_1 > 0$, there exists a pole z of g with multiplicity m such that $f(z) = z_1$ and $m_1 \leq m$. Calculate g_1 when

$$f(x) = x - \frac{1}{x} \quad \text{and} \quad g(x) = \frac{2x^2}{\pi(x^2 + 1)^2}.$$

(4) Let $z = a + ib \in \mathbf{C}$, with $b > 0$. The Cauchy measure γ_z on $\overline{\mathbf{R}}$ is defined by $\gamma_z(dx) = \frac{bm(dx)}{\pi[(x-a)^2+b^2]}$. Prove, using (3), that the image of γ_z under T is $\gamma_{f(z)}$.

REMARKS. 1. A Cayley function is a function of the form

$$f(x) = c_0 x + \alpha - \sum_{j=1}^{n} \frac{c_j}{x - a_j},$$

where $c_j \geq 0$, $j = 0, 1, \dots, n$ and α, $a_1, \dots, a_n$ are real. If $c_0 = 0$ and $n = 1$, it is a positive linear fractional transformation; that is, $f(x) = \frac{ax+b}{cx+d}$ with

a, b, c, and d real and $ad-bc>0$. It is easy to see that all Cayley functions can be obtained by composing positive linear fractional transformations with the Cayley functions corresponding to $c_0=1$.
2. It is easy to see that if f is a positive linear fractional transformation and T is its restriction to $\mathbf{R}$, then the image of γ_z under T is $\gamma_{f(z)}$. This observation, the remark above, and result (4) of the problem show that the property holds for all Cayley functions.
3. Conversely, let $T:\overline{\mathbf{R}}\to\overline{\mathbf{R}}$ be a rational function such that, for every z with positive imaginary part, the image of γ_z under T is a Cauchy distribution γ_{z_1} (where z_1 depends on z). It can be proved that T is the restriction to the real axis of a Cayley function.
4. On the other hand, a Cayley function with $c_0>0$ maps Lebesgue measure m to c_0m. If $c_0=0$, the image measure is no longer a Radon measure on $\mathbf{R}$. For example, $f(x)=-\frac{1}{x}$ maps $m(dx)$ to $\frac{m(dx)}{x^2}$.

Problem II-13. The half-plane $\mathbf{R}^2_+=\{(x,y) : x\in\mathbf{R}\ \text{ and }\ y>0\}$ is equipped with the measure $\mu(dx,dy)=\frac{dx\,dy}{y^2}$. What is the image ν on $[1,+\infty)$ of this measure under the mapping $(x,y)\mapsto v(x,y)=\frac{1}{2y}(1+x^2+y^2)$ (in the sense of Problem I-14)?

Problem II-14. Let $\{\mu_n\}_{n\geq 0}$ be a sequence of positive measures on $\mathbf{R}$, each with total mass ≤ 1. Suppose that μ_n converges weakly to μ_0 as $n\to\infty$ and that

$$M=\sup_n\int_{-\infty}^{+\infty}x^2\mu_n(dx)<\infty.$$

(1) Show that μ_n converges narrowly to μ_0 as $n\to\infty$.

(2) Show that $\displaystyle\int_{-\infty}^{+\infty}|x|\mu_n(dx)\to\int_{-\infty}^{+\infty}|x|\mu_0(dx)$ as $n\to\infty$.

(3) Show by a counterexample that $\displaystyle\int_{-\infty}^{+\infty}x^2\mu_n(dx)$ does not necessarily tend to $\displaystyle\int_{-\infty}^{+\infty}x^2\mu_0(dx)$.

METHOD. Use Theorem II-6.8.

Problem II-15. If g is a measurable function on $(0,+\infty)$ which is locally integrable, and if $A=\lim_{T\to+\infty}\int_1^T g(x)dx$ and $B=\lim_{\epsilon\to 0}\int_\epsilon^1 g(x)dx$ exist, we say that $\int_0^{+\infty}g(x)dx$ exists and equals $A+B$.

Let f be measurable and locally integrable on $(0,+\infty)$ and suppose that $\lim_{T\to+\infty}\int_1^T f(x)\frac{dx}{x}$ exists. Let a and b be positive.
(1) Suppose that $K=\int_0^\infty f(x)\frac{dx}{x}$ exists and let F be defined by $F(x)=\int_1^x f(t)dt$. Show that $\int_0^\infty[F(ax)-F(bx)]\frac{dx}{x^2}$ exists and express the integral in terms of a, b, and K.
(2) Suppose that $L=\lim_{\epsilon\to 0}f(x)$ exists. Show that $\int_0^\infty(f(ax)-f(bx))\frac{dx}{x}$ exists and express the integral in terms of a, b, and L.

Problem II-16. Writing $x^{-1} = \int_0^\infty e^{-yx}dy$ for $x > 0$ and applying Fubini's theorem, show that the integral $\int_0^\infty \sin x \frac{dx}{x}$ exists (in the sense of Problem II-15) and compute it. Use this to evaluate the integrals $\int_0^\infty (\cos ax - \cos bx)\frac{dx}{x^2}$ and $\int_0^\infty (\cos ax - \cos bx)\frac{dx}{x}$ if a, $b > 0$. (See Problem II-15.)

Problem II-17. For an interval I in $\mathbf{R}$, $L^p(I)$ denotes the set of real-valued functions (rather, equivalence classes of functions) whose pth power is integrable with respect to Lebesgue measure on I.
(1) Show that $L^{p'}([0,1]) \subset L^p([0,1])$ if $0 < p < p' \leq \infty$. Give an example of a function in $L^1([0,1]) \setminus L^2([0,1])$.
(2) Give examples of functions in $L^1(\mathbf{R}) \setminus L^2(\mathbf{R})$ and in $L^2(\mathbf{R}) \setminus L^1(\mathbf{R})$.
(3) ℓ^p is the set of real-valued sequences $a = \{a_n\}_{n\geq 0}$ such that $\sum |a_n|^p < \infty$. Show that $\ell^{p'}(\mathbf{N}) \supset \ell^p(\mathbf{N})$ if $0 < p < p' \leq \infty$. Give an example of a sequence in $\ell^2 \setminus \ell^1(\mathbf{N})$.

Problem II-18. Let $\mathbf{R}_+^{n+1}$ denote the set of pairs (a,p) with $p > 0$ and $a \in \mathbf{R}^n$. Euclidean space $\mathbf{R}^n$ is equipped with the scalar product $\langle a, t\rangle$ and the norm $\|a\|$. Let

$$K(a,p) = K_n p\left[\|a\|^2 + p^2\right]^{-(n+1)/2},$$

where K_n is the constant such that $\int_{\mathbf{R}^n} K(x,1)dx = 1$. The goal of this problem is to calculate

$$I_t(a,p) = \int_{\mathbf{R}^n} \exp i\langle x,t\rangle K(x-a,p)dx,$$

where $t \in \mathbf{R}^n$.

If $f : \mathbf{R}_+^{n+1} \to \mathbf{C}$, we write $D_0 f = \frac{\partial}{\partial p} f$ and $D_j f = \frac{\partial}{\partial a_j} f$ for $j = 1, \ldots, n$. f is said to be *harmonic* in $\mathbf{R}_+^{n+1}$ if

$$(D_0^2 + \cdots + D_n^2) f(a,p) = 0 \quad \text{for every } (a,p) \in \mathbf{R}_+^{n+1}.$$

(1) Show that K is harmonic in $\mathbf{R}_+^{n+1}$. Show that, if $p_0 > 0$ and $V = (\frac{p_0}{2}, \frac{3p_0}{2})$, there exists a constant C such that $|D_i K(a,p)|$ and $|D_i D_j K(a,p)|$ are less than $C(1+\|a\|^2)^{-\frac{n+1}{2}}$ for all $(a,p) \in \mathbf{R}^n \times V$ and i, $j = 0, 1, \ldots, n$.
(2) Let μ be a Radon measure on $\mathbf{R}^n$ such that

$$\int_{\mathbf{R}^n} (1 + \|x\|^2)^{-(n+1)/2} |\mu|(dx) < \infty$$

and let $F_\mu(a,p) = \int_{\mathbf{R}^n} K(x-a,p)\mu(dx)$. Show that F_μ is harmonic and that $\lim_{p\to+\infty} F_\mu(a,p) = 0$.
(3) Show that there exists a function $g : \mathbf{R}^n \to \mathbf{C}$ such that $I_t(a,p) = g(pt)\exp(i\langle a,t\rangle)$.
Use (2) to calculate g.

REMARKS. 1. In n dimensions, $K(x-a,p)$ is sometimes called the Poisson kernel; in $\mathbf{R}^n$, it is sometimes called the Cauchy distribution.
2. The calculation giving $K_n = \Gamma(\frac{n+1}{2})\pi^{-(n+1)/2}$ is carried out in Problem III-4.

Problem II-19. (1) Let μ and ν be positive measures on $\mathbf{R}$ such that there exists an interval $[a,b] \subset \mathbf{R}$ with $\mu([a,b]) = \mu(\mathbf{R})$ and $\nu([a,b]) = \nu(\mathbf{R})$. Show that $\mu = \nu$ if and only if

$$\int_{\mathbf{R}} x^n \mu(dx) = \int_{\mathbf{R}} x^n \nu(dx), \quad \forall n = 0, 1, 2, \ldots.$$

(2) Let μ be a positive measure on $[0,+\infty)$ (not necessarily bounded). Its *Laplace transform* is the function from $\mathbf{R}$ to $[0,+\infty]$ defined by

$$s \mapsto (L\mu)(s) = \int_0^\infty \mathrm{e}^{-sx}\mu(dx).$$

(a) If $E_\mu = \{s : (L\mu)(s) < \infty\}$, show that E_μ is an interval which, if nonempty, is unbounded on the right. Give examples where $E_\mu = \mathbf{R}$, $\emptyset$, $(0,+\infty)$, and $[0,+\infty)$.
(b) Use (1) to show that if there exists a number a such that $L\mu = L\nu < +\infty$ on $[a,+\infty)$, then $\mu = \nu$.

Problem II-20. Give examples of sequences $\{\mu_n\}_{n=1}^\infty$ of positive Radon measures on $\mathbf{R}$ such that there exists a positive Radon measure μ with $\lim_{n\to\infty} \mu_n = \mu$
(1) vaguely but not weakly;
(2) weakly but not narrowly; and
(3) narrowly but not in norm.

REMARK. If the sequence of positive measures $\{\mu_n\}_{n=1}^\infty$ converges vaguely to μ and $\mu(X) < \infty$, then $\mu_n \to \mu$ weakly, since $C_K(X)$ is dense in $C_0(X)$. It should also be noted that narrow and weak convergence coincide when X is complete.

Problem II-21. Let X be a locally compact space which is countable at infinity and let $M^1(X)$ be the set of signed Radon measures ν on X such that $|\nu|$ has finite total mass $\|\nu\|$. If $\{\nu_n\}_{n=1}^\infty$ is a sequence in $M^1(X)$ such that $r = \sup_n \|\nu_n\| < \infty$, show that there exist ν in $M^1(X)$ and an increasing sequence of integers $\{n_k\}_{k=1}^\infty$ such that $\nu_{n_k} \to \nu$ as $k \to \infty$. Show also that $\nu \geq 0$ if $\nu_n \geq 0$ for every n.

METHOD. Use Theorem II-6.6.

REMARK. When $X = \mathbf{R}$, $\nu_n \geq 0$, and $r = 1$, this property is often called Helly's theorem.

Problem II-22. On a locally compact space X which is countable at infinity, let μ and $\{\mu_n\}_{n=1}^\infty$ be positive Radon measures such that μ_n converges vaguely to μ as $n \to \infty$.

(1) If O is an arbitrary open set, show that $\mu(O) \le \liminf_{n\to\infty} \mu_n(O)$.
(2) Suppose that O is an open set with compact closure K and such that its boundary $\partial O = K \setminus O$ has μ-measure 0. Let $\{O_k\}_{k=1}^{\infty}$ be a decreasing sequence of open subsets of X such that $\cap_{k=1}^{\infty} O_k = K$. Let f_k be a function equal to 1 on K and to 0 on O_k^c and satisfying $0 \le f(x) \le 1$ for x in O_k. (Such a function exists by Urysohn's lemma, II-1.1.) Show that

$$\limsup_{n\to\infty} \mu_n(O) \le \int_X f_k(x)\mu(dx),$$

and conclude that $\mu_n(O) \to \mu(O)$ as $n \to \infty$.
(3) If μ and $\{\mu_n\}_{n=1}^{\infty}$ are Radon measures on $\mathbf{R}$, positive and with total mass less than or equal to 1, show that μ_n converges weakly to μ as $n \to \infty$ if and only if

$$\mu_n((a,b)) \to \mu((a,b)) \quad \text{as } n \to \infty$$

for all points of continuity of the distribution function $x \mapsto \mu((-\infty, x))$.

If, moreover, $\mu_n(\mathbf{R}) = \mu(\mathbf{R}) = 1$, show that $\mu_n \to \mu$ narrowly if and only if

$$\mu_n((-\infty, x)) \to \mu((-\infty, x)) \quad \text{as } n \to \infty$$

for every point of continuity of the right-hand side.

METHOD. Use Problems II-1 and II-21 together with Theorem II-6.8.

REMARK. In practice, (3) gives a necessary and sufficient condition for the convergence of probability distributions on $\mathbf{R}$; it is often taken as a definition in elementary texts.

Problem II-23. Let X be a locally compact space which is countable at infinity, and let μ and $\{\mu_n\}_{n=1}^{\infty}$ be Radon measures on X such that μ_n converges vaguely to μ.
(1) If O is an open set in X and μ^* is the restriction of μ to O, show that μ_n^* converges vaguely to μ^* as $n \to \infty$.
(2) Show by an example that the statement is false if O is replaced by a closed set.
(3) Suppose that $X = \mathbf{R}$ and that $\mu_n \ge 0$, $n = 1, 2, \ldots$. Let a and b be real numbers with $a < b$. Show that there exist numbers p and q and an increasing sequence of integers $\{n_k\}_{k=1}^{\infty}$ such that, for every continuous function f on $[a, b]$,

$$\int_{[a,b]} f\mu_{n_k} \to pf(a) + qf(b) + \int_{[a,b]} f\mu \quad \text{as } n \to \infty.$$

METHOD. Use Problem II-21.

Problem II-24. (1) Let O and O' be two open sets in $\mathbf{R}^n$, let f be a diffeomorphism from O onto O', and let φ be a measurable function on O'

such that $\int_{O'} \varphi(x')dx' < \infty$. Show that

$$\int_O \varphi(f(x))|\det J_f(x)|dx = \int_{O'} \varphi(x')dx',$$

where $|\det J_f(x)|$ is the Jacobian.
(2) Let $a \in \mathbf{R} \cup \{-\infty\}$. Let f and g be functions satisfying the following conditions: (i) f is continuously differentiable for $x > a$; (ii) g is defined and integrable on $[0, +\infty)$; (iii) $|f'(x + \frac{u^2}{2})| \leq g(u)$ for all $x > a$; and (iv) both $u \mapsto ug(u)$ and $u \mapsto f(x + \frac{u^2}{2})$ are integrable on $[0, +\infty)$. If $F(x) = \int_{-\infty}^{+\infty} f(x + \frac{u^2}{2})du$, show by a change of variables in polar coordinates that

$$f(x) = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} F'(x + \frac{v^2}{2})dv.$$

REMARK. The case $f(x) = e^{-x}$ is well known and is used in IV-4.3.2(i).

Problem II-25. Consider a subset X of $\mathbf{R}^n$ with positive measure, a measurable function $f : X \to \mathbf{R}^n$, and a nonnegative locally integrable function h on X. Let μ denote the image in $\mathbf{R}^n$ of the measure $h(x)dx$ on X under f (in the sense of Problem I-14) if this image measure exists.
(1) If X and U are open sets and f is a diffeomorphism from X to U, show that

$$\mu(du) = h(f^{-1}(u))|\det J_{f^{-1}}(u)|du.$$

(2) If there exist an open subset U of $\mathbf{R}^n$ and disjoint open sets X_1, $X_2, \ldots, X_d$ contained in X such that the restriction f_j of f to X_j is a diffeomorphism on U, and if $X \setminus \sum_{j=1}^d X_j$ has Lebesgue measure zero, show that

$$\mu(du) = \sum_{j=1}^{d} h(f_j^{-1}(u))|\det J_{f_j^{-1}}(u)|\mathbf{1}_U(u)du.$$

(3) If $X = (0, +\infty)^2$, $c(x) = x^{-3/2} \exp[-(ax + b/x)]$, $h(x, y) = c(x)c(y)$, and $f(x, y) = (u, v)$, with $u = x + y$ and $v = 1/x + 1/y$, calculate μ. Conclude from the result that the image of $hdxdy$ under the map $(x, y) \mapsto (x + y, 1/x + 1/y - 4/(x + y))$ is also a product measure.

REMARKS. 1. The use of the change-of-variables theorem (II-4.4.1) to calculate the image of a measure is important in practice, especially in probability theory.
2. Problem II-12 treats a special case of (2) for $n = 1$.
3. (3) shows that if X and Y are independent random variables of density $Kc(x)dx$ (a distribution called "inverse Gaussian"), then $X + Y$ and $1/X + 1/Y - 4/(X + Y)$ are independent. It seems difficult to justify this result by Fourier analysis.

Exercises for Chapter III

Problem III-1. Let G be the group O_d of $d \times d$ orthogonal matrices, acting on the Euclidean space $\mathbf{R}^d$. The scalar product and the norm are denoted by $\langle x, t\rangle$ and $\|t\|$, respectively. Let μ be a bounded complex measure on $\mathbf{R}^d$, with Fourier transform

$$\widehat{\mu}(t) = \int_{\mathbf{R}^d} \exp(i\langle x, t\rangle)\mu(dx) \quad (t \in \mathbf{R}^d).$$

Prove the equivalence of the following three properties:
(1) μ is invariant under every element of G.
(2) There exists $\varphi : [0, \infty) \to \mathbf{C}$ such that $\widehat{\mu}(t) = \varphi(\|t\|)$ for every t.
(3) The image ν_a in $\mathbf{R}$ of μ under the mapping $x \mapsto \langle a, x\rangle$ does not depend on a when a ranges over the unit sphere S_{d-1} of $\mathbf{R}^d$.

REMARK. Naturally, if μ is real, then $\widehat{\mu}(t) = \overline{\widehat{\mu}(-t)}$ implies that φ is real. But $\mu \geq 0$ does not imply that $\varphi \geq 0$. Thus, if σ is the uniform probability measure on S_2, the unit sphere in $\mathbf{R}^3$, $\widehat{\sigma}(t) = \frac{\sin \|t\|}{\|t\|}$.

Problem III-2. Let T be a compact space, let G be a compact topological group, and let $(g, t) \mapsto gt$ be a continuous map from $G \times T$ to T such that $g \mapsto \{(g, t) \mapsto gt\}$ is a homomorphism from G to the group of bijections of T. Finally, suppose that (G, T) is a homogeneous space; that is, for every t_1 and t_2 in T there exists g such that $gt_1 = t_2$. Let dg denote the unique measure of total mass 1 on G which is invariant under left and right multiplication. (We accept without proof the existence and uniqueness of dg.)

(1) If f is continuous on T, show that $t \mapsto \int_G f(g^{-1}t)dg$ is a constant $\sigma[f]$. Conclude that $\sigma[f]$ defines a probability measure on T which is invariant under the action of G.
(2) If μ is a probability measure on T which is invariant under the action of G, show that $g \mapsto \int_T f[g^{-1}t]\mu(dt)$ is a constant. Integrate with respect to dg and conclude that $\mu = \sigma$.
(3) If $(X, \mathcal{A})$ is an arbitrary measurable space and T is equipped with its Borel algebra, let $T \times X$ be given the product σ-algebra. Suppose that G acts on $T \times X$ by $g(t, x) = (gt, x)$. Show that every positive measure μ on $T \times X$ which is invariant under the action of G has the form $\sigma(dt) \otimes \nu(dx)$, where ν is a measure ≥ 0 on $(X, \mathcal{A})$. Converse?

METHOD. If $A \in \mathcal{A}$ is such that $\mu(T \times A) \in (0, +\infty)$, show that $\mu_A(B) = \frac{\mu(B\times A)}{\mu(T\times A)}$ defines a probability measure on T which is invariant under G.
(4) Apply the preceding results when $T = S_d$ is the unit sphere of the Euclidean space $\mathbf{R}^{d+1}$, where $G = O_{d+1}$ is the group of $(d+1) \times (d+1)$ orthogonal matrices and $X = (0, +\infty)$. Conclude that a probability measure P on $\mathbf{R}^{d+1} \setminus 0$ is invariant under G if and only if $\frac{x}{\|x\|}$ and $\|x\|$ are independent and $\frac{x}{\|x\|}$ has the uniform distribution on S_d.

Problem III-3. In the Euclidean space $\mathbf{R}^d$ equipped with the norm $\|x\|$, let m be Lebesgue measure.
(1) If ν_0 and ν_1 are the images of m in $[0, +\infty)$ under the mappings $x \mapsto \|x\|$ and $x \mapsto \frac{\|x\|^2}{2}$ (see Problem I-14), show that

$$\nu_1(d\gamma) = \frac{(\sqrt{2\pi})^d}{\Gamma(\frac{d}{2})}\gamma^{\frac{d}{2}-1}d\gamma,$$

where Γ is the usual Euler function (see, for example, Problem IV-11). Use this to find $\nu_0(d\rho)$.

METHOD. Use the formula

$$\frac{1}{\sigma^d(\sqrt{2\pi})^d}\int_{\mathbf{R}^d} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) dx = 1,$$

which holds for all $\sigma > 0$, to calculate the Laplace transform $(L\nu_1)(s)$ defined in Problem II-19.
(2) Keep the same notation m and ν_0 for the restrictions of m and ν_0 to $\mathbf{R}^d \setminus \{0\}$ and $(0, +\infty)$. If μ is a measure ≥ 0 on $\mathbf{R}^d \setminus \{0\}$ which has density f with respect to m, use Problem III-2 to show that the image of μ on $(0, +\infty)$ under the map $x \mapsto \|x\|$ is of the form $f_1(\rho)\nu_0(d\rho)$ and calculate the function f_1 in terms of f. If μ is rotation invariant, show that there exists a function $f_1 : (0, +\infty) \to [0, +\infty)$ such that $f_1(\|x\|) = f(x)$ m-a.e.

Problem III-4. Euclidean space $\mathbf{R}^d$ is equipped with the scalar product $\langle x, t\rangle$ and the norm $\|t\|$. Γ is the usual Euler function.

(1) Use Problem III-3 to evaluate $I = \displaystyle\int_{\mathbf{R}^d} \frac{dx}{(1+\|x\|^2)^{\frac{d+1}{2}}}$. If a and t are in $\mathbf{R}^d$ and $p > 0$, use Problem II-18 to conclude that

$$\frac{\Gamma(\frac{d+1}{2})}{\pi^{\frac{d+1}{2}}} \int_{\mathbf{R}^d} \mathrm{e}^{i\langle x,t\rangle} \frac{p\,dx}{(p^2+\|x-a\|^2)^{\frac{d+1}{2}}} = \mathrm{e}^{-p\|t\|+i\langle a,t\rangle}.$$

(2) Show that, if $x \in \mathbf{R}^d$ and $p > 0$,

$$2^d(\frac{d+1}{2})\pi^{\frac{d-1}{2}} \frac{p}{(p^2+\|x\|^2)^{\frac{d+1}{2}}} = \int_{\mathbf{R}^d} \mathrm{e}^{-p\|t\|+i\langle x,t\rangle} dt.$$

Problem III-5. Let k be a positive integer. In the Euclidean space $\mathbf{R}^{2k-1}$, the norm is written $\|t\|$ and the scalar product $\langle x,t\rangle$. Consider the map $\varphi : \mathbf{R}^{2k-1} \mapsto [0,1]$ defined by

$$\varphi(t) = \left[(1-\|t\|)^+\right]^k.$$

(1) Using Problem III-1, show that there exists a continuous function $f : [0,+\infty) \to \mathbf{R}$ such that

$$f(\|x\|) = \int_{\mathbf{R}^{2k-1}} \exp(i\langle x,t\rangle)\varphi(t)dt.$$

(2) Use Problems III-3 and III-4 to show that, for every $s > 0$,

$$I = \int_0^\infty \mathrm{e}^{-su} u^{3k-1} f(u)du = C_k \left[\int_0^\infty \mathrm{e}^{-su}(1-\cos u)du\right]^k,$$

where C_k is a constant.
(3) Show that $f \geq 0$ and that $\int_{\mathbf{R}^{2k-1}} f(x)dx < \infty$ by using Problem II-19 and the sequence of functions $f_n : [0,+\infty) \to \mathbf{R}$ defined by $f_1(u) = 1-\cos u$ and $f_{n+1}(u) = \int_0^u f_n(u-\rho)f_1(\rho)d\rho$.

Conclude that φ is the Fourier transform of a probability measure on $\mathbf{R}^{2k-1}$. Compute it for $k=1$ and $k=2$.
(4) Suppose that $g : [0,+\infty) \to \mathbf{R}$ is continuous and satisfies the following conditions: (i) $g(0) = 1$; (ii) $(-1)^{k-1}g^{(k-1)}(x)$ exists and is convex on $(0,+\infty)$; and (iii) $\lim_{x\to+\infty} g(x) = \lim_{x\to+\infty} g^{(k-1)}(x) = 0$. Use Problem II-6 to show that $g(\|t\|)$ is the Fourier transform of a probability measure on $\mathbf{R}^{2k-1}$.

REMARK. The result of (4) for $k = 1$ is due to G. Polya (1923), and the general case to R. Askey (1972).

Problem III-6. Let $\mathbf{C}$ denote the complex numbers. A function $p : \mathbf{C} \to [0, +\infty)$ is called a *seminorm* if

(i) $p(\lambda z) = |\lambda| p(z)$ for all $\lambda \in \mathbf{R}$ and $z \in \mathbf{C}$, and
(ii) $p(z_1 + z_2) \leq p(z_1) + p(z_2)$ for all z_1 and z_2 in $\mathbf{C}$.

(1) Let $p : \mathbf{C} \to [0, +\infty)$ satisfy (i). Prove the equivalence of the following properties:
(a) p is a seminorm.
(b) $\{z : p(z) \leq 1\}$ is a convex subset of $\mathbf{C} = \mathbf{R}^2$.
(c) For all α_1, α_2, α_3 such that $\alpha_1 < \alpha_2 < \alpha_3$ and $\alpha_3 - \alpha_1 < \pi$,

$$(iii) \quad p(\mathrm{e}^{i\alpha_3}) \sin(\alpha_2 - \alpha_1) + p(\mathrm{e}^{i\alpha_1}) \sin(\alpha_3 - \alpha_2) - p(\mathrm{e}^{i\alpha_2}) \sin(\alpha_3 - \alpha_1) \geq 0.$$

(2) Let μ be a bounded positive measure on $[0, \pi)$. Show that

$$(iv) \qquad p_\mu(x + iy) = \int_0^\pi |x \sin\alpha - y \cos\alpha| \mu(d\alpha)$$

defines a seminorm. Show that $p_\mu = p_{\mu_1}$ implies $\mu = \mu_1$.

METHOD. Observe that $p(\mathrm{e}^{i\theta})$ is the convolution of μ and $|\sin\theta|$ in the group $\mathbf{R}/\pi\mathbf{Z}$. (See III-1.8.)
(3) Let $0 \leq \alpha_1 < \alpha_2 < \ldots \alpha_n < \pi$, with the convention that $\alpha_0 = \alpha_n - \pi$ and $\alpha_{n+1} = \alpha_1 + \pi$. The matrices $A = (a_{ij})_{i,j=1}^n$, $B = (b_{ij})_{i,j=1}^n$, and $D = (d_{ij})_{i,j=1}^n$ are defined as follows:
$a_{ij} = |\sin(\alpha_i - \alpha_j)|$ for all $i, j = 1, \ldots, n$.
$b_{ii} = -\sin(\alpha_{i+1} - \alpha_{i-1})$, $b_{i,i+1} = \sin(\alpha_i - \alpha_{i-1})$ (with the convention that $b_{n,n+1} = b_{n,1}$), $b_{i,i-1} = \sin(\alpha_{i+1} - \alpha_i)$ (with the convention that $b_{1,0} = b_{1,n}$) for $i = 1, \ldots, n$, and $b_{ij} = 0$ otherwise.
$d_{ii} = 2\sin(\alpha_{i+1} - \alpha_i)\sin(\alpha_i - \alpha_{i-1})$ for $i = 1, \ldots, n$ and $d_{ij} = 0$ otherwise.
Verify that $AB = D$. If $\mu = \sum_{j=1}^n m_j \delta_{\alpha_j}$, where $m_j > 0$ and δ_{α_j} is the Dirac measure at α_j for $j = 1, 2, \ldots, n$, calculate $p(\mathrm{e}^{i\theta})$ and verify that

$$(v) \qquad [m_1, m_2, \ldots, m_n] A = \left[p_\mu(\mathrm{e}^{i\alpha_1}), p_\mu(\mathrm{e}^{i\alpha_2}), \ldots, p_\mu(\mathrm{e}^{i\alpha_n})\right].$$

(4) If p is a seminorm, show that there exists a bounded positive measure μ on $[0, \pi)$ such that $p = p_\mu$.

METHOD. Let $T = \{\alpha_1, \ldots, \alpha_n\}$ with $\alpha_0 = \alpha_n - \pi < 0 \leq \alpha_1 < \ldots < \alpha_n < \pi < \alpha_{n+1} = \alpha_1 + \pi$. Show that there exists a seminorm p_T such that, if $0 \leq \lambda \leq 1$ and $j = 1, \ldots, n$,

$$(vi) \qquad p_T\left[\lambda \mathrm{e}^{i\alpha_j} + (1 - \lambda)\mathrm{e}^{i\alpha_{j+1}}\right] = \lambda p_T\left[\mathrm{e}^{i\alpha_j}\right] + (1 - \lambda) p_T\left[\mathrm{e}^{i\alpha_{j+1}}\right],$$

and show by using (3) that there exists μ_T concentrated on T such that $p_T = p_{\mu_T}$.

Next, let $\alpha_j = \frac{(j-1)\pi}{n}$ and set $p_n = p_T$ and $\mu_n = \mu_T$. Show that $p = \lim_{n\to\infty} p_n$ and that there exists a bounded positive measure μ on $[0,\pi)$ such that μ_n converges vaguely to μ as $n \to \infty$.

REMARKS. A consequence of (4) is that every seminorm on $\mathbf{R}^2$ can be approximated by finite sums of the type $\sum_j |a_j x + b_j y|$, and not only by $\sup_j |a_j x + b_j y|$. For $\mathbf{R}^n$ with $n > 2$ this is false; in general, a seminorm can be approximated only by suprema of absolute values of linear functionals.

Problem III-7. Let $\mathbf{C}$ be the set of complex numbers, identified with $\mathbf{R}^2$, and let p be a seminorm on $\mathbf{C}$. Show that $\exp(-p(t))$ is the Fourier transform of a probability measure on $\mathbf{R}^2$.

METHOD. Use the fact, proved in Problem III-6, that there exists a sequence of measures $\mu_n \geq 0$ on $[0,\pi)$, concentrated at a finite number of points, such that

$$p(x+iy) = \lim_{n\to\infty} \int_0^\pi |x\sin\alpha - y\cos\alpha| \mu_n(d\alpha).$$

Also use the formula $\mathrm{e}^{-|t|} = \int_{-\infty}^{+\infty} \mathrm{e}^{itx} \frac{dx}{\pi(1+x^2)}$, which appeared in Problem III-4.

REMARKS. This result is due to T. Ferguson (1962). It is false in higher dimensions; only for certain norms (like the Euclidean norm) is $\exp(-p(t))$ the Fourier transform of a probability measure. See Problem III-8 for a counterexample.

Problem III-8. (1) What is the image ν in $\mathbf{R}$, under the projection $(x_0,\ldots,x_n) \mapsto x_0$, of the measure $\exp(-\max_{j=0,\ldots,n}|x_j|)dx_0 dx_1 \ldots dx_n$ in $\mathbf{R}^n$? (See Problem I-14.)
(2) Compute the Fourier transform of ν.

METHOD. Show that $k!(1-it)^{-(k+1)} = \int_0^\infty x^k \exp(-x+itx)dx$ for t real and k a nonnegative integer.
(3) Conclude that $\varphi_{n+1}(t) = \exp(-\max_{j=0,\ldots,n}|t_j|)$ is not the Fourier transform of a probability measure on $\mathbf{R}^{n+1}$ when $n \geq 2$.

REMARK. (3) is due to C. Herz (1963).

Problem III-9. Let E be n-dimensional Euclidean space.
(1) If $\alpha > 0$, $\beta > 0$, and $\alpha + \beta < n$, show that there exists a constant $K(\alpha,\beta)$ such that $I(y) = \int_E \|x\|^{\alpha-n}\|y-x\|^{\beta-n}dx = K(\alpha,\beta)\|y\|^{\alpha+\beta-n}$.

METHOD. Use Problem III-3.

(2) Let $0 < \gamma < n$ and let M_γ be the set of positive measures μ, not necessarily bounded, such that $f(\mu) = \int_{E\times E} \|x-y\|^{\gamma-n}\mu(dx)\mu(dy) < \infty$. Show that, if μ and ν are in M_γ,

$$\left|\int_{E\times E} \|x-y\|^{\gamma-n}\mu(dx)\nu(dy)\right| \leq \sqrt{f(\mu)f(\nu)}.$$

Problem III-10. Let M be the space of *real* Radon measures on $\mathbf{U} = \{z : z \in \mathbf{C} \text{ and } |z| = 1\}$ and let F^+ (respectively F^-) be the vector space over $\mathbf{R}$ of complex functions defined in $\{z : |z| > 1\} = D^+$ (resp. in $\{z : |z| < 1\} = D^-$). For $\mu \in M$, we define

$$f_\mu^+(z) = \int_{\mathbf{U}} (\mathrm{e}^{i\theta} - z)^{-1} d\mu(\mathrm{e}^{i\theta}) \quad \text{for } z \in D^+,$$
$$f_\mu^-(z) = \int_{\mathbf{U}} (\mathrm{e}^{i\theta} - z)^{-1} d\mu(\mathrm{e}^{i\theta}) \quad \text{for } z \in D^-.$$

(1) Show that the linear mapping $\mu \mapsto f_\mu^+$ from M to F^+ is injective.

METHOD. Expand f^+ in a power series in $1/z$.
(2) Find the kernel of the linear mapping $\mu \mapsto f_\mu^-$ from M to F^-.

REMARKS. 1. Although f_μ^+ determines μ, f_μ^- does not.
2. The situation is completely different if μ is complex, since there exist complex measures, like $d\mu(\mathrm{e}^{i\theta}) = \mathrm{e}^{-i\theta}d\theta$, for which $\widehat{\mu}(n) = 0$ for all $n \geq 0$.

Problem III-11. Let $P(x_1, \ldots, x_n) = P(x)$ be a homogeneous polynomial of degree m in n variables which is harmonic; that is, $\sum_{k=1}^n \frac{\partial^2 P}{\partial x_k^2}(x) = 0$ for all x in $\mathbf{R}^n$. For a fixed $\sigma < 0$, let

$$f(x) = (\sigma\sqrt{2\pi})^{-n} \exp\left(\frac{-\|x\|^2}{2\sigma^2}\right) P(x), \quad \text{with} \quad \|x\|^2 = \textstyle\sum_{k=1}^n x_k^2.$$

Show by induction on m that there exists a number $K_m(\sigma)$ such that

$$\widehat{f}(t) = K_m(\sigma)P(t)\exp\left(\frac{-\sigma^2\|t\|^2}{2}\right).$$

METHOD. $mP = \sum_k x_k \frac{\partial P}{\partial x_k}$.

Problem III-12. The goal of this problem is to prove the following inequality of S. Bernstein: If μ is a complex measure on $[-a, +a]$, then $|\widehat{\mu}'(t)| \leq a \sup_{s\in\mathbf{R}} |\widehat{\mu}(s)|$.
(1) Consider the *odd* function $h(\theta)$ of period 2π defined by $h(\theta) = \theta$ if $0 \leq \theta \leq \pi/2$ and $h(\theta) = \pi - \theta$ if $\pi/2 \leq \theta \leq \pi$.

(a) Compute $\nu_n = (2i\pi)^{-1}\int_{-\pi}^{+\pi} h(\theta)\exp(-in\theta)d\theta$ for n in $\mathbf{Z}$.

(b) If ν is the measure defined on $\mathbf{R}$ by $\nu = \sum_{n=-\infty}^{\infty} \nu_n\delta_n$, where δ_n is the Dirac measure at n, show that ν is bounded and that $h(\theta) = i\int_{-\infty}^{+\infty} \exp(ix\theta)\nu(dx)$.

(2) If μ is a complex measure on $[-\pi/2, \pi/2]$, let

$$f(t) = \int_{-\pi/2}^{\pi/2} \exp(it\theta)\mu(d\theta).$$

(a) Show that $f(t) = (f * \nu)(t)$ for all real t.

(b) If $\mu = \mu_0 = (2i)^{-1}(\delta_{\frac{\pi}{2}} - \delta_{-\frac{\pi}{2}})$, deduce from (a) that $\sum_{k=-\infty}^{+\infty}(2k-1)^{-2} = \pi^2/4$.

(c) Returning to the general case, deduce from (a) and (b) that

$$|f'(t)| \leq \frac{\pi}{2} \sup_{s\in\mathbf{R}} |f(s)| \quad \text{for all } t \text{ in } \mathbf{R}.$$

Show that equality holds if and only if μ is concentrated at the points $\pm\pi/2$.

(3) Prove Bernstein's inequality and discuss in detail the case of equality.

Problem III-13. Let $f : (0, +\infty) \to \mathbf{R}$ be measurable and satisfy

$$f(x+y) = f(x) + f(y) \quad \text{for all } x \text{ and } y > 0.$$

(1) If $\varphi(t) = \int_0^1 \exp[itf(x)]dx$ for $t \in \mathbf{R}$, show that $y \mapsto \varphi(t)\exp[itf(y)]$ is continuous on $(0, +\infty)$ and conclude that f is continuous.

(2) Show that $f(x) = xf(1)$ for $x > 0$.

Problem III-14. Let E be a real vector space of finite dimension n and let $\widehat{E}$ be its dual. Let $e_1, \ldots, e_n$ be a basis of E. The dual basis $e_1^*, \ldots, e_n^*$ of $\widehat{E}$ is defined by $\langle e_j, e_i^*\rangle = 0$ if $j \neq i$ and 1 if $j = i$, where $\langle\ ,\ \rangle$ is the canonical bilinear form on $E \times \widehat{E}$. E and $\widehat{E}$ are equipped with Lebesgue measures dx and dt, respectively, such that, if $f \in L^1(E, dx)$ implies $\widehat{f} \in L^1(\widehat{E})$, where $\widehat{f}(t) = \int_E \exp(i\langle x, t\rangle)f(x)dx$, then $f(x) = (2\pi)^{-n}\int_{\widehat{E}} \exp(-i\langle x, t\rangle)\widehat{f}(t)dt$. Let Z denote the set of points $z = \sum_{i=1}^n z_i e_i$ of E such that the z_i are integers and let Z^* denote the set of points $\zeta = \sum_{i=1}^n \zeta_i e_i^*$ of $\widehat{E}$ such that the ζ_i are integers.

Prove Poisson's formula:

> If f is in the space $\mathcal{S}$ of infinitely differentiable functions of rapid decrease, then for every t in $\widehat{E}$
>
> $$\sum_{\zeta\in Z^*} \widehat{f}(2\pi\zeta + t) = [\mathrm{vol}(e_1^*, \ldots, e_n^*)]^{-1} \sum_{z\in Z} f(z)e^{i\langle z,t\rangle}.$$

METHOD. Show that $\sum_{z\in Z}|f(z)| < \infty$ and use Theorem III-4.2 to see that the left-hand side $\psi(t)$ of the equation exists. Observing that the set of periods of ψ contains $2\pi Z^*$, compute the Fourier coefficients of ψ.

REMARKS. 1. With the above hypotheses on the choice of dt on $\widehat{E}$, it can be shown that

$$\mathrm{vol}(e_1, \ldots, e_n) \times \mathrm{vol}(e_1^*, \ldots, e_n^*) = 1.$$

Without loss of generality we may assume that $\operatorname{vol}(e_1, \dots, e_n) = 1$. Let E be given the Euclidean structure such that $(e_1, \dots, e_n)$ is orthonormal; then $\widehat{E}$ can be identified canonically with E, $e_j^* = e_j$, and dx and dt are identical.
2. Poisson's formula is also valid in some situations that differ slightly from that where $f \in \mathcal{S}(E)$. One of these occurs when $f \in L^1(E)$, $f \geq 0$, and $\widehat{f}$ has compact support.
3. A striking application of Poisson's formula is that if

$$g(\sigma) = \sqrt{\sigma} \sum_{n=-\infty}^{+\infty} \exp(-\sigma^2 \pi n^2),$$

then $g(\sigma) = g(\sigma^{-1})$. To prove this, it suffices to take $E = \mathbf{R}$, $e_1 = 1$, and $f(x) = \exp(-2\pi^2 x^2/\sigma^2)$.

Problem III-15. Let E be a real vector space of dimension $n > 0$, let $\widehat{E}$ be its dual, and let E be equipped with Lebesgue measure dx. It is always true that $\widehat{\widehat{E}} = E$. The canonical linear form on $E \times \widehat{E}$ is written $\langle\ ,\ \rangle$. We consider the following operators, where $a \in E$, $b \in \widehat{E}$, c (respectively d) is an invertible linear mapping from E into E (resp. from $\widehat{E}$ into $\widehat{E}$), and ${}^t c$ (resp. ${}^t d$) is the transpose of c (resp. d).

For $f \in L^2(E)$,

$$T_a f(x) = f(x-a), \quad M_b f(x) = \mathrm{e}^{i\langle x,b\rangle} f(x), \quad H_c f(x) = f(c^{-1}x),$$

and $Uf \in L^2(\widehat{E})$ is the Fourier-Plancherel transform described in III-2.4.9.

For $g \in L^2(\widehat{E})$,

$$T_b g(t) = g(t-b), \quad M_a g(t) = \mathrm{e}^{i\langle a,t\rangle} g(t), \quad H_d g(T) = g(d^{-1}t),$$

and $Vg \in L^2(E)$ is the Fourier-Plancherel transform.

Prove the following formulas:

$$\begin{array}{llll}
(1) & UT_a = M_a U & (1') & VT_b = M_b V \\
(2) & UM_b = T_b U & (2') & VM_a = T_a V \\
(3) & UH_c = |\det c| H_{({}^t c)^{-1}} U & (3') & VH_d = |\det d| H_{({}^t d)^{-1}} V \\
(4) & (H_{-\mathbf{1}_E} U)(f) = \overline{U(\overline{f})} & (4') & (H_{-\mathbf{1}_{\widehat{E}}} V)(g) = \overline{V(\overline{g})} \\
(5) & U^{-1} = (2\pi)^{-n} H_{-\mathbf{1}_{\widehat{E}}} V & (5') & V^{-1} = (2\pi)^{-n} H_{-\mathbf{1}_E} U.
\end{array}$$

(Here $\mathbf{1}_E$ and $\mathbf{1}_{\widehat{E}}$ are the identity operators on E and $\widehat{E}$, respectively.

Problem III-16. Use the result of Problem IV-12,

$$\int_0^\infty x^{\alpha-1} \mathrm{e}^{-x+ixt} \frac{dx}{\Gamma(\alpha)} = (1-it)^{-\alpha} \quad \text{for } t \in \mathbf{R} \text{ and } \alpha > 0,$$

with the convention for z^α with $\mathrm{Re} z > 0$ made in Problem IV-12, to compute the Fourier-Plancherel transforms of the following functions in $L^2(\mathbf{R})$:
(1) $|x|^{\alpha-1}\mathrm{e}^{-x}\mathbf{1}_{[0,+\infty)}(x)$
(2) $|x|^{\alpha-1}\mathrm{e}^{x}\mathbf{1}_{(-\infty,0]}(x)$
(3) $|x|^{\alpha-1}\mathrm{e}^{-x}$
(4) $-i\cdot\mathrm{sign}(x)|x|^{\alpha-1}\mathrm{e}^{-|x|}$
(5) $(x-a-ib)^{-n}$, with n a positive integer, a and b real, and $b\neq 0$
(6) $b(x^2+b^2)^{-1}$
(7) $x(x^2+b^2)^{-1}$
(8) $f(x)$, where $f(x)$ is a rational function with no real poles and without entire part.

METHOD. For (5), use (1) and problem III-15.

Problem III-17. Compute the Fourier-Plancherel transforms of the following functions:
(1) $\mathbf{1}_{[-1,+1]}(x)$
(2) $\mathbf{1}_{[\alpha,\beta]}(x)$
(3) $\sin x/x$
(4) $\sin^2 x/x^2$
(5) $(1-|x|)^+$
(Here $a^+=\max\{0,a\}$.)

Problem III-18. If $f\in L^2(\mathbf{R})$ and $(U_af)(t)=\int_{-a}^{a}\mathrm{e}^{ixt}f(x)dx$, show that $\lim_{a\to\infty}U_a(f)=U(f)$, where U denotes the Fourier-Plancherel transform of f.

Problem III-19. If f and g are in $L^2(\mathbf{R})$, show that

$$\int f(x)\widehat{g}(x)dx=\int \widehat{f}(x)g(x)dx.$$

METHOD. Use the fact that $L^1\cap L^2(\mathbf{R})$ and $A(\mathbf{R})$ are dense in $L^2(\mathbf{R})$. (See III-2.4.7.)

Problem III-20. Let $g_b(x)=i(\mathrm{sign}x)\mathrm{e}^{-b|x|}$ for $b>0$, let U be the Fourier-Plancherel transform in $L^2(\mathbf{R})$, and let M_{g_b} be the operator on $L^2(\mathbf{R})$ defined by $M_{g_b}f(x)=g_b(x)f(x)$. Set

$$\mathcal{H}_b=U^{-1}M_{g_b}U.$$

(1) Show that $\mathcal{H}_bf(x)=\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{y}{y^2+b^2}f(x-y)dy$ for almost every x, if $f\in L^2(\mathbf{R})$ and $b>0$.

METHOD. Use Problem III-16(7) to compute $U(g_b)$, then apply Problem III-19.
(2) If $f\in L^2(\mathbf{R})$, show that $\mathcal{H}_0f=\lim_{b\downarrow 0}\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{y}{y^2+b^2}f(x-y)dy$ exists in L^2 and give its Fourier transform. Also calculate $\mathcal{H}_0^2f$.

REMARK. $\mathcal{H}_0 f$ is called the *Hilbert transform* of f.

Problem III-21. Suppose that $f \in L^2(\mathbf{R})$ and $g \in L^1(\mathbf{R})$. Show that

$$h(x) = \int_{-\infty}^{+\infty} f(x-y)g(y)dy$$

exists for almost every x and defines a function h in $L^2(\mathbf{R})$ such that $\|h\|_{L^2} \leq \|f\|_{L^2}\|g\|_{L^1}$ and $\widehat{h} = \widehat{g}\widehat{f}$ (where $\widehat{g}$ is the Fourier transform of $g \in L^1$ and $\widehat{h}$ and $\widehat{f}$ are the Fourier-Plancherel transforms in L^2).

METHOD. Apply the Cauchy-Schwarz inequality to $|f(x-y)|\, |g(y)|^{1/2}$ (considered as a function of y) and $|g(y)|^{1/2}$ and use Problem III-18.

Problem III-22. Let $0 < \epsilon < a$ and let $g_{\epsilon,a}(y) = (\pi y)^{-1}\mathbf{1}_{\{\epsilon \leq |y| \leq a\}}(y)$.
(1) Compute $\lim_{\epsilon\to 0}\lim_{a\to+\infty}\widehat{g}_{\epsilon,a}(t)$, where $\widehat{g}_{\epsilon,a}$ is the Fourier transform on $L^1(\mathbf{R})$ of $g_{\epsilon,a}$. (Use Problem II-16.)
(2) For $f \in L^2(\mathbf{R})$, we set

$$\mathcal{H}_{\epsilon,a}(f) = \int_{\epsilon \leq |y-x| \leq a} \frac{f(y)dy}{x-y}.$$

(This equals $f * g_{\epsilon,a}$ in the sense of Problem III-21.) Using Problems III-20 and III-21, show that $\lim_{\epsilon\to 0}\lim_{a\to+\infty}\mathcal{H}_{\epsilon,a}(f)$ exists and coincides with the Hilbert transform of f (Problem III-20).

Problem III-23. A function f in $L^2(\mathbf{R})$ is called *hermitian* if $f(x) = \overline{f(-x)}$ and *skew hermitian* if $f(x) + \overline{f(-x)} = 0$. Let $\widehat{f}$ denote the Fourier-Plancherel transform of f and let $\mathcal{H}_0 f$ denote the Hilbert transform of f. (See Problems III-20 and III-22.) Prove the following statements:

f is iff	Hermitian	Skew-hermitian	Real	Purely imaginary	Even	Odd
$\widehat{f}$ is iff	Real	Purely imaginary	Hermitian	Skew-hermitian	Even	Odd
$(\mathcal{H}_0 f)^\wedge$ is iff	Purely imaginary	Real	Hermitian	Skew-hermitian	Odd	Even
$(\mathcal{H}_0 f)$ is	Skew-hermitian	Hermitian	Real	Purely imaginary	Odd	Even

Problem III-24. Compute the Hilbert transform (see Problem III-20) of each of the following functions:

$$\begin{aligned} f_1(x) &= \tfrac{1}{\pi}\tfrac{1}{1+x^2}, & f_2(x) &= \tfrac{x}{\pi(1+x^2)}, \\ f_3(x) &= \tfrac{1}{2}\mathbf{1}_{[-1,1]}(x), & f_4(x) &= \tfrac{1}{\pi}\log\left|\tfrac{x+1}{x-1}\right|, \\ f_5(x) &= (1-|x|)^+, & f_6(x) &= \tfrac{1}{\pi}\log\left|\tfrac{x+1}{x-1}\right| + \tfrac{x}{\pi}\log\left|\tfrac{x^2-1}{x^2}\right|. \end{aligned}$$

METHOD. Use Problem III-16 for f_2 and Problem III-22 for f_6.
For f_5, $a^+ = \max\{0, a\}$.

Problem III-25. Let $\mathcal{S}$ be the vector space of C^∞ functions on $\mathbf{R}$ which, together with all their derivatives, are of rapid decrease.
(1) Show that if $f \in \mathcal{S}$, then $\lim_{\epsilon\to 0}\int_{|x|\geq\epsilon}\frac{f(x)}{x}dx$ exists and defines a continuous linear functional (or "tempered distribution") on $\mathcal{S}$.
(2) Show that the Fourier transform of the distribution defined in (1) is the Radon measure $\mu(dt) = i\pi(\text{sign } t)dt$.

METHOD. Split the first integral into $\{\epsilon \leq |x| \leq 1\} \cup \{|x| > 1\}$. Also use the fact, proved in Problem II-16, that $\int_0^\infty \frac{\sin x}{x}dx = \frac{\pi}{2}$.

Problem III-26. Let $I = (a, b)$ and let $f \in L^1(I)$.
(1) If $F(x) = \int_a^x f(t)dt$ for $x \in I$, show that $F'(x) = f(x)$ in the weak sense (III-3.3.1).
(2) If $F \in L^1(I)$ and $F' = f$ in the weak sense, show that, for $a < \alpha < \beta < b$,

$$\int_\alpha^\beta f(t)dt = F(\beta) - F(\alpha).$$

(3) Let s be a positive integer. Show that F is in H^s_{loc}, the local Sobolev space (see III-3.5.6), if and only if there exists $f \in L^2_{loc}(I)$ such that the weak derivative of order $s-1$ of F exists in the ordinary sense and satisfies

$$F^{(s-1)}(x) = F^{(s-1)}(a) + \int_\alpha^x f(t)dt$$

for all α and x in I.

Problem III-27. Let $f \in L^2(\mathbf{R})$, with Fourier-Plancherel transform $\widehat{f}$. Prove Hermann Weyl's inequality,

$$\left[\int_{-\infty}^{+\infty}|f(x)|^2dx\right]^2 \leq \frac{2}{\pi}\int_{-\infty}^{+\infty}x^2|f(x)|^2dx \times \int_{-\infty}^{+\infty}t^2|\widehat{f}(t)|^2dt,$$

and analyze the case of equality.

METHOD. Without loss of generality, assume that f is in the Sobolev space $H^1(\mathbf{R})$. Show that $\int_{-\infty}^{+\infty}|f(x)|^2dx = -2\text{Re}\int_{-\infty}^{+\infty}xf(x)\overline{f'(x)}dx$, with the help of Problem I-15(2). Conclude by using the Cauchy-Schwarz inequality (Problem I-12).

REMARK. This inequality has an interpretation in quantum mechanics, where it is known as Heisenberg's uncertainty principle. [3]

---

[3] H. Weyl, *The Theory of Groups and Quantum Mechanics* (London: Dover, 1931).

Exercises for Chapter IV

Problem IV-1. The points marked on the faces of two dice are, respectively, for the first: 1, 2, 2, 3, 3, 4; for the second: 1, 3, 4, 5, 6, 8.
If X is the sum of the points obtained by throwing the two dice, compute $P[X = k]$ for integer k. Answer the same question for ordinary dice.

Problem IV-2. The random variable X is called a geometric distribution with parameter p, $0 < p < 1$, if

$$P[X = k] = (1-p)^{k-1}p, \quad k = 1,\ 2,\ 3, \ldots$$

Compute $\mathbf{E}(X)$ by using Problem I-6(1).

Problem IV-3. Suppose that δ_a is the Dirac measure at a, $p \in (0,1)$, and $\lambda > 0$. Consider the following two probability measures on $\mathbf{N}$:

$$\begin{aligned} \nu_p &= (1-p)\delta_0 + p\delta_1 && \text{(Bernouilli distribution with parameter } p\text{)} \\ \mu_\lambda &= \textstyle\sum_{k=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \delta_k && \text{(Poisson distribution with parameter } \lambda\text{)} \end{aligned}$$

(1) Show that the vague limit of the sequence $\{\nu_{\frac{\lambda}{n}}^{*n}\}_{n>\lambda}$ is μ_λ and that $\mu_{\lambda_1} * \mu_{\lambda_2} = \mu_{\lambda_1+\lambda_2}$.
(2) Let $0 < p < 1$. Consider the measure m_p on $\mathbf{N}^2$ concentrated at the points $(0,0)$, $(0,1)$, $(1,1)$, and $(k,0)$ with $k \geq 2$ (note the absence of $(1,0)$), such that X has distribution μ_p and Y has distribution ν_p if (X,Y) has distribution m_p. Compute m_p and conclude that $P(X \neq Y) \leq 2p^2$. (Use the fact that $\mathrm{e}^{-p} \geq 1-p$.)

(3) If (X, Y) is an arbitrary variable in $\mathbf{N}^2$ and $A \subset \mathbf{N}$, show that

$$|P(X \in A) - P(Y \in A)| \leq P(X \neq Y).$$

(4) Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be n independent random variables with values in $\mathbf{N}^2$ and with distributions $m_{p_1}, m_{p_2}, \ldots, m_{p_n}$. Let $A \subset \mathbf{N}$. Use (2) and (3) to show that

$$|P(X_1 + \cdots X_n \in A) - P(Y_1 + \cdots + Y_n \in A)| \leq 2 \sum_{j=1}^{n} p_j^2.$$

(5) If $n > \lambda$ and $A \subset \mathbf{N}$, show that

$$|\nu_{\lambda/n}^{*n}(A) - \mu_\lambda(A)| \leq \frac{2\lambda^2}{n}.$$

REMARKS. The approximation of the binomial distribution by the Poisson distribution is both elementary and essential for applications. (5) gives an upper bound for the error committed by replacing a binomial distribution ν_p^n by a Poisson distribution μ_{np}, and (4) treats the case of experiments that are independent but not identical. This result is due to J.L. Hodges and L. Lecam (1960).

Problem IV-4. On a probability space $(\Omega, \mathcal{A}, P)$, we define a random variable N with positive integer values and random variables $\{X_n\}_{n \geq 1}$, with values in a measurable space $(I, \mathcal{B})$, such that the X_n all have the same distribution m but are not necessarily independent.
(1) Show that the distribution μ of X_N is absolutely continuous with respect to m.
(2) If $f(x) = \frac{d\mu}{dm}(x)$ and $\alpha > 0$, show that

$$\mathbf{E}(N^\alpha) \geq \frac{1}{1+\alpha} \int_I f^{\alpha+1}(x) dm(x).$$

METHOD. If $B(y) = \{x \in I : f(x) > y\}$, show that

$$(i) \qquad \mu(B(y)) \leq \sum_{n \leq y} P[X_n \in B(y)] + P[N > y]$$

and use Problem I-6.
(3) Show that $1 + \mathbf{E}(\log N) \geq \int f(x) \log f(x) dm(x)$ by letting $\alpha \downarrow 0$ in (2) and using the monotone convergence theorem.

Problem IV-5. With the notation of Problem IV-4, we take $I = [0, 1]$, $\mathcal{B}$ = the Borel algebra, and m = Lebesgue measure, and we assume that the $\{X_n\}_{n \geq 1}$ are independent. Let $f : I \to [0, +\infty)$ be a nonnegative measurable function, bounded by a number $b > 1$, which satisfies $\int_0^1 f(x)dx = 1$. Let

$$N = \inf\{2n : bX_{2n-1} \leq f(X_{2n})\}.$$

Show that X_N has density $\frac{d\mu}{dm} = f$.

REMARK. This procedure for constructing a random variable of given density f on $[0,1]$ from uniform random variables was invented by J. Von Neumann in 1951.

Problem IV-6. (1) Let Y be a positive random variable. Show that for all $y > 0$

$$P(Y \geq y) \leq \frac{1}{y}\mathbf{E}(Y) \quad \text{(Chebyshev's inequality)}.$$

(2) Let X be a real random variable such that $\mathbf{E}(X^2) < \infty$. If $m = \mathbf{E}(X)$, show that for all $t > 0$

$$P(|X - m| \geq t) \leq \frac{1}{t^2}\mathbf{E}((X - m)^2) \quad \text{(Bienaimé's inequality)}.$$

(3) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent real random variables with the same distribution and such that $\mathbf{E}(X_1^2) < \infty$. If $m = \mathbf{E}(X_1)$, show that for all $\epsilon > 0$ and for all $\alpha \in [0, \frac{1}{2})$

$$P\left[\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \frac{\epsilon}{n^\alpha}\right] \to 0 \text{ as } n \to 0$$

(weak law of large numbers).

(4) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent real random variables with the same distribution, for which there exists $k > 0$ such that $\mathbf{E}[\exp k|X_1|] < \infty$. If $m = \mathbf{E}(X_1)$, show that for every $\epsilon > 0$ there exists q in $(0,1)$ such that

$$P\left[\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \epsilon\right] \leq 2q^n.$$

Conclude that $\frac{1}{n}(X_1 + \cdots + X_n) \to m$ almost surely as $n \to \infty$ (strong law of large numbers).

METHOD. Show that $m = \frac{d}{ds}[\mathbf{E}(\exp(sX_1))]_{s=0}$ and apply Chebyshev's inequality to $Y = \exp(s(X_1 + \cdots X_n))$.

Problem IV-7. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of nonnegative real random variables with the same distribution, such that X_j and X_n are independent for every pair (j, n) with $j \neq n$. Assume that $\mathbf{E}(X_1) < \infty$. Set $S_n = \sum_{j=1}^{n} X_j$, $Y_n = X_n \mathbf{1}_{\{X_n \leq n\}}$, and $S_n^* = \sum_{j=1}^{n} Y_j$. The goal of this problem is to prove the law of large numbers:

$$P\left[\lim_{n\to\infty} \frac{S_n}{n} = \mathbf{E}(X_1)\right] = 1.$$

(1) Using Problem I-6, show that $\mathbf{E}(X_1) < \infty$ implies $\sum_{n=1}^{\infty} P[X_n \neq Y_n] < \infty$. Using the Borel-Cantelli lemma (I-5.2.8), conclude that $\lim_{n\to\infty}(S_n - S_n^*)$ exists with probability 1.

(2) Show that $\lim_{n\to\infty} \frac{1}{n}\mathbf{E}(S_n^*) = \mathbf{E}(X_1)$.

(3) Let α be a real number greater than 1 and let k_n be the integer part of α^n. Prove the existence of a constant C_1 such that $\sum\{k_n^{-2} : n \text{ such that } k_n \geq j\} \leq C_1 j^{-2}$. With the help of Bienaimé's inequality, conclude that

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{S_{k_n}^* - \mathbf{E}(S_{k_n}^*)}{k_n}\right| \geq \epsilon\right\} \leq \frac{C_1}{\epsilon^2}\sum_{j=1}^{\infty}\frac{1}{j^2}\mathbf{E}(Y_j^2).$$

Then prove that $\sum_{j=1}^{\infty}\frac{1}{j^2}\mathbf{E}(Y_j^2) < \infty$.
(4) Deduce from (2), (3), and the Borel-Cantelli lemma that

$$P\left[\lim_{n\to\infty}\frac{S_{k_n}^*}{k_n} = \mathbf{E}(X_1)\right] = 1,$$

then from (1) that

$$P\left[\lim_{n\to\infty}\frac{S_{k_n}}{k_n} = \mathbf{E}(X_1)\right] = 1.$$

(5) Prove that, for every $\alpha > 1$,

$$P\left[\alpha^{-1}\mathbf{E}(X_1) \leq \liminf_{n\to\infty}\frac{S_n}{n} \leq \limsup_{n\to\infty}\frac{S_n}{n} \leq \alpha\mathbf{E}(X_1)\right] = 1.$$

Deduce the law of large numbers from this.

REMARKS. The elementary proof whose outline is sketched here is due to N. Etemadi (1981).

Problem IV-8. Let $\{X_m\}_{n=1}^{\infty}$ be independent real random variables with the same distribution and such that $\mathbf{E}(X_1) = 0$ and $0 < \mathbf{E}(X_1^2) < \infty$. Let $S_n = X_1 + \cdots X_n$.
(1) Show that $\lim_{n\to\infty} P(S_n \geq 0) = \frac{1}{2}$ by using Laplace's theorem (IV-4.3.1) and Problem II-22.
(2) Use the preceding result and the weak law of large numbers proved in Problem IV-6(3) (that $\lim_{n\to\infty} P[|\frac{S_n}{n^{\frac{3}{4}}}| \geq \epsilon] = 0$ for all $\epsilon > 0$) to show that $\lim_{n\to\infty}[\mathbf{E}(\exp(-S_n)\mathbf{1}_{\{S_n\geq 0\}})]^{1/n} = 1$.

Problem IV-9. (1) If X and Y are independent real random variables, show that $P(X+Y \geq a+b) \geq P(X \geq a)P(Y \geq b)$ for all real a and b.
(2) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of real independent random variables with the same distribution, and set $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$. Let s be a fixed real number. Set $p_n = P[S_n \geq ns]$. Show that $p_{n+m} \geq p_n p_m$ for all m, $n \geq 0$ and that, for $n > 0$, $p_n = 0$ if and only if $p_1 = 0$.
(3) If the sequence $\{a_n\}_{n=0}^{\infty}$ of nonnegative real numbers is such that $a_{n+m} \geq a_n + a_m$ for all m, $n \geq 0$, show that $\lim_{n\to\infty}\frac{a_n}{n} = \inf_{d>0}\frac{a_d}{d}$. Conclude that $\lim_{n\to\infty}\sqrt[n]{p_n} = \alpha(s)$ exists.

Problem IV-10. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent real variables with the same distribution. Suppose that $\varphi(t) = \mathbf{E}(\exp tX_1)$ exists for all t in an open interval I containing 0 and fix a real number $s > \mathbf{E}(X)$ such that $t \mapsto \mathrm{e}^{-ts}\varphi(t)$ attains its minimum $\alpha(s)$ at a point τ of I. Let $S_n = X_1 + \cdots + X_n$.

(1) Show that $\log \varphi(t)$ is convex on I and that $\tau > 0$. Conclude, using Chebyshev's inequality (Problem IV-6), that

$$\left[P\left(\frac{S_n}{n} \geq s\right)\right]^{1/n} \leq \alpha(s).$$

(2) Let μ_1 be the distribution of $X_1 - s$ and let $\nu(dx) = \frac{\mathrm{e}^{\tau x}}{\alpha(s)}\mu_1(dx)$. Prove that ν is a probability measure, that $\int x\nu(dx) = 0$, and that $\int x^2\nu(dx) < \infty$.

(3) Let $\{Z_n\}_{n=1}^{\infty}$ be a sequence of independent random variables with the same distribution ν. Show that

$$P\left[\frac{S_n}{n} \geq s\right] = (\alpha(s))^n \mathbf{E}[\exp(-\tau(Z_1 + \cdots + Z_n))\mathbf{1}_{\{Z_1+\cdots+Z_n \geq 0\}}].$$

Conclude from Problem IV-8 that

$$\alpha(s) = \lim_{n\to\infty}\left[P\left(\frac{S_n}{n} \geq s\right)\right]^{1/n}.$$

(4) Compute $\alpha(s)$ in the following cases:

(a) $\varphi(t) = \exp(t^2/2)$ (normal distribution)
(b) $\varphi(t) = \cosh(t)$ (Bernouilli distribution)
(c) $\varphi(t) = (1-t)^{-\alpha}$, $t < 1$, $\alpha > 0$ (gamma distribution)
(d) $\varphi(t) = \exp \lambda(\mathrm{e}^t - 1)$, $\lambda > 0$ (Poisson distribution)
(e) $\varphi(t) = (\frac{p}{1-q\mathrm{e}^t})$, $p + q = 1$, $0 < p < 1$, $t < -\log q$, $\alpha > 0$ (negative binomial distribution)
(f) $\varphi(t) = \frac{1}{1-t^2}$, $|t| < 1$ (Laplace's first distribution)
(g) $\varphi(t) = \frac{1}{\cos t}$ (logarithm of a Cauchy distribution)

REMARK. It is not known what conditions on a decreasing function α on $\mathbf{R}$ are sufficient for the existence of a distribution μ of the X_n such that

$$\lim_{n\to\infty}\left[P\left(\frac{S_n}{n} \geq s\right)\right]^{1/n} = \alpha(s).$$

Problem IV-11. If z_1 and z_2 are complex numbers with positive real part, set $\Gamma(z_1) = \int_0^{\infty} x^{z_1-1}\mathrm{e}^{-x}dx$ and $B(z_1, z_2) = \int_0^1 x^{z_1-1}(1-x)^{z_2-1}dx$. Assume without proof the formula

$$B(z_1, z_2) = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1 + z_2)}.$$

If a and b are positive, the probability measures

$$\gamma_a(dx) = \mathbf{1}_{(0,+\infty)}(x)x^{a-1}\mathrm{e}^{-x}\frac{dx}{\Gamma(a)}$$

$$\beta_{a,b}(dx) = \mathbf{1}_{(0,1)}(x)x^{a-1}(1-x)^{b-1}\frac{dx}{B(a,b)}$$

$$\beta^{(2)}_{a,b}(dx) = \mathbf{1}_{(0,+\infty)}(x)x^{a-1}(1+x)^{-a-b}\frac{dx}{B(a,b)}$$

are called, respectively, the gamma distribution with parameter a and the beta distributions of the first and second kind with parameters a and b. (1) If μ is a bounded measure on $(0,+\infty)$, its Mellin transform is $(M\mu)(t) = \int_0^\infty x^{it}\mu(dx)$ for t real. (This is the Fourier transform of the image of μ under $x \mapsto \log x$.) Compute $M\gamma_a$, $M\beta_{ab}$, and $M\beta^{(2)}_{ab}$. (2) If X is a random variable with distribution $\beta_{a,b}$, compute the distribution of $X/1-X$. (3) If X and Y are independent r.v. with distributions γ_a and γ_b, compute the distributions of X/Y and $X/(X+Y)$. (4) If X, Y, and Z are independent r.v. with distributions $\beta_{a,b}$, $\beta_{a+b,c}$, and $\beta^{(2)}_{a+b,c}$, compute the distributions of XY and XZ.

Problem IV-12. (1) Let γ_a be the probability measure of Problem IV-11, with $a > 0$. Compute its Fourier transform. If X and Y are independent random variables with distributions γ_a and γ_b, compute the distribution of $X+Y$.
(2) Let X be a Gaussian random variable with density $\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-x^2/2\sigma^2}dx$. Compute $\mathbf{E}\left[\left(\frac{X^2}{2\sigma^2}\right)^{it}\right]$ for t real, and use Problem IV-11 to find the distribution of $\frac{X^2}{2\sigma^2}$.
(3) Let $X_1,\ldots,X_d$, $Y_1,\ldots,Y_m$ be independent random variables with the same distribution as X of (2). Compute the distribution of $\frac{1}{2\sigma^2}[X_1^2+\ldots+X_d^2]$ by using (1) and (2), and the distributions of $\frac{X_1^2+\cdots+X_d^2}{Y_1^2+\cdots+Y_m^2}$ and $\frac{X_1^2+\cdots+X_d^2}{X_1^2+\cdots+X_d^2+Y_1^2+\cdots+Y_m^2}$ by using Problem IV-11(3).

Problem IV-13. In Euclidean space $\mathbf{R}^{d+1}$, consider a random variable $X = (X_0, X_1, \ldots, X_d)$ whose distribution μ is invariant under every orthogonal matrix of $\mathbf{R}^{d+1}$ and satisfies $\mu(\{0\}) = 0$. Let ν denote the distribution of X on $(0,+\infty)$ and let $Y = (\frac{X_1}{X_0}, \frac{X_2}{X_0}, \ldots, \frac{X_d}{X_0})$.
(1) Use Problem III-2 to show that the distribution of Y is independent of ν.
(2) From now on, assume that the $\{X_j\}_{j=0}^n$ are independent, with the same distribution μ and with Fourier transform $\exp(-\frac{t^2}{2})$. Show, using Problem III-1, that μ must be invariant under every orthogonal matrix.

(3) If $a \in \mathbf{R}$, compute the integral

$$I(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\left(x^2 + \frac{a^2}{x^2}\right)\right] dx$$

by using the following fact from Problem II-12(1):

$$\int_{-\infty}^{+\infty} f(x - \frac{|a|}{x})dx = \int_{-\infty}^{+\infty} f(y)dy \quad \text{for every } f \text{ integrable on } \mathbf{R}.$$

(4) By first conditioning with respect to X_0 (see Problem IV-34), compute the Fourier transform of the distribution of Y.
(5) Using Problem IV-11, find the distribution of $\|Y\|^2$. Derive the density of Y from this, by observing that the distribution of Y is invariant under every orthogonal matrix in O_d and using Problem III-3.

Problem IV-14. Let γ_a be the probability measure of Problem IV-11, with $a > 0$.
(1) Use Problem IV-12 to compute $\lim_{a\to\infty} \int_0^\infty \exp[it(\frac{x-a}{\sqrt{a}})]\gamma_a(dx)$.
(2) Using Problem II-14, show that

$$\lim_{a\to+\infty} \int_0^\infty |\frac{x-a}{\sqrt{a}}|\gamma_a(dx) = \int_{-\infty}^{+\infty} |x|e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}.$$

(3) Integrate by parts to compute the integral $\int_0^\infty |\frac{x-a}{\sqrt{a}}|\gamma_a(dx)$ and prove Stirling's formula:

$$\lim_{a\to+\infty} \frac{a^{a-1/2}e^{-a}}{\Gamma(a)} = \frac{1}{\sqrt{2\pi}}.$$

Problem IV-15. (1) Let μ be a probability measure on $\mathbf{R}$ such that $\widehat{\mu}(t) = \widehat{\mu}(t\cos\theta)\widehat{\mu}(t\sin\theta)$ for all real t and θ. Show that there exists $\sigma \geq 0$ such that $\widehat{\mu}(t) = \exp(-\frac{\sigma^2 t^2}{2})$.

METHOD. Show that $\widehat{\mu}(t) \geq 0$, then that $\widehat{\mu}(t) > 0$ for every t. Finally, consider $f(x) = -\log\widehat{\mu}(\sqrt{x})$ for $x \geq 0$.
(2) For positive integers d_1 and d_2, let μ_1 and μ_2 be probability measures on the Euclidean spaces $\mathbf{R}^{d_1}$ and $\mathbf{R}^{d_2}$ such that $\nu = \mu_1 \otimes \mu_2$ is invariant under the group G of orthogonal matrices on $\mathbf{R}^{d_1+d_2}$. Show that there exists $\sigma \geq 0$ such that $\widehat{\mu}_j(t) = \exp(-\frac{\sigma^2}{2}\|t\|_j)$, $j = 1, 2$, where $\|t\|_1$ and $\|t\|_2$ are the norms in $\mathbf{R}^{d_1}$ and $\mathbf{R}^{d_2}$.

METHOD. Use Problem III-1 and part (1) of this problem for the case where $d_1 = d_2 = 1$.

REMARK. The converse of the property in (2) is trivial. This characterization of centered normal distributions is sometimes called Maxwell's theorem.

Problem IV-16. A real random variable Z is called *symmetric* if Z and $-Z$ have the same distribution.
(1) Show that Z has distribution $c(dz) = \frac{dz}{\pi(1+z^2)}$ if and only if Z is symmetric and $|Z|^2$ has distribution $\beta^{(2)}(1/2, 1/2)$. (See Problem IV-11.) Assuming without proof the formula $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z}$ for complex numbers z such that $0 < \mathrm{Re} z < 1$, compute $\mathbf{E}(|Z|^{it})$ for real t in this case.
(2) Let X_1 and X_2 be two real random variables that are independent and symmetric, and have distributions μ_1 and μ_2 such that $\mu_1(\{0\}) = \mu_2(\{0\}) = 0$. Show that $Z = \frac{X_2}{X_1}$ has distribution c in the following cases:

(a) $\mu_1(dx) = \mu_2(dx) = \exp(-x^2/2)dx/\sqrt{2\pi}$
(b) $|X_1|^2$ has distribution $\beta(\frac{1}{2}, b)$ and $|X_2|^2$ has distribution $\beta^{(2)}(\frac{1}{2}, \frac{1}{2}+b)$
(c) $\mu_1(dx) = \mu_2(dx) = \sqrt{2}/\pi dx/(1+x^4)$

(3) With X_1 and X_2 as in (2), deduce from (2a) that $U = \left(\frac{X_1}{\sqrt{X_1^2+X_2^2}}, \frac{X_2}{\sqrt{X_1^2+X_2^2}}\right)$ is uniformly distributed on the unit circle of Euclidean space $\mathbf{R}^2$ if and only if $Z = \frac{X_2}{X_1}$ has distribution c.

REMARKS. Example (2c) is due to Laha (1949). Moreover, if (X_1, X_2) is as in (2) with U uniform, then $(\frac{1}{X_1}, \frac{1}{X_2})$ has the same property.

Problem IV-17. A probability measure ν on a Euclidean space $\mathbf{R}^d$ is called *isotropic* if $\nu(\{0\}) = 0$ and the image of ν under the mapping $x \mapsto \frac{x}{\|x\|}$, in the unit sphere S_{d-1} of $\mathbf{R}^d$, is the unique rotation-invariant probability measure σ_{d-1} on S_{d-1}. It is called *radial* if its image ν_a in $\mathbf{R}$ under the mapping $x \mapsto \langle a, x\rangle$ does not depend on a when a ranges over the unit sphere.
(1) Let μ_1 and μ_2 be probability measures on the Euclidean spaces $\mathbf{R}^{d_1}$ and $\mathbf{R}^{d_2}$, with d_1 and d_2 positive. Show that the probability measure $\nu = \mu_1 \otimes \mu_2$ on the Euclidean space $\mathbf{R}^{d_1+d_2}$ is isotropic if and only if μ_1 and μ_2 are radial and if, for every a_1 in S_{d_1-1} and a_2 in S_{d_2-1}, the image of ν under $(x_1, x_2) \mapsto \frac{\langle a_2, x_2\rangle}{\langle a_1, x_1\rangle}$ is $c(dz) = \frac{dz}{\pi(1+z^2)}$.

METHOD. Prove the assertion first for $d_1 = d_2 = 1$ and use Problem IV-16.
(2) Let (X_1, X_2, X_3) be three independent random variables such that the distribution ν of (X_1, X_2, X_3) in $\mathbf{R}^3$ is isotropic. Show that there exists $\sigma > 0$ such that

$$\mathbf{E}[\exp(itX_j)] = \exp\left(-\frac{\sigma^2 t^2}{2}\right) \quad \text{for } j = 1,\ 2,\ 3, \text{ and } t \in \mathbf{R}.$$

METHOD. Apply (1) to the distributions μ_1 of X_1 and μ_2 of (X_2, X_3) and use Problem IV-14.

REMARKS. The converse of (1) is true but rather lengthy to prove. (2) is true for n independent random variables, $n \geq 3$; this follows easily from the problem. (Problem IV-16 showed that this would be false for $n = 2$.)

This property of the normal distribution is due to I. Kotlarski (1966), who proves it with the additional hypothesis that the X_j are symmetric.

Problem IV-18. Let E be a finite-dimensional real vector space, let E^* be its dual, and let $\langle x,t\rangle$ be the canonical bilinear form on $E \times E^*$. If μ is a probability measure on E, its Fourier transform is defined on E^* by

$$\widehat{\mu}(t) = \int_E \exp(i\langle x,t\rangle)\mu(dx).$$

(1) If there exists $t_0 \neq 0$ such that $|\widehat{\mu}(t_0)| = 1$, show that μ is concentrated on a countable union of affine hyperplanes and determine them.

METHOD. First consider the case where $\widehat{\mu}(t_0) = 1$.
(2) If there exists a probability measure ν on E such that $\widehat{\mu}(t)\widehat{\nu}(t) = 1$ for every t in E^*, show that μ and ν are Dirac measures.

METHOD. First prove this when $\dim E = 1$.

REMARKS. This result can be generalized by replacing E and E^* by a locally compact abelian group and its group $\check{G}$ of continuous characters χ. (See III-1.4.)

Problem IV-19. Let X_1, X_2, Y_1, and Y_2 be independent real random variables such that Y_1 and Y_2 are strictly positive and $\mathbf{E}[\exp(itX_j)] = \exp(-t^2/2)$ for $j = 1,2$ and t real. Let $R = [X_1^2Y_1^2 + X_2^2Y_2^2]^{1/2}$. Using Problems IV-16 and IV-18, find the distributions of Y_1 and Y_2 such that $U = (X_1Y_1/R, X_2Y_2/R)$ is uniformly distributed on the unit circle of $\mathbf{R}^2$.

Problem IV-20. Let σ_{d-1} be the uniform probability measure on the unit sphere S_{d-1} of the Euclidean space $\mathbf{R}^d$, and let ν_d be the image of μ_d under the dilation $x \mapsto \sqrt{d}x$.

Prove that ν_d converges narrowly to $\nu(dx) = \exp(-x^2/2)dx/\sqrt{2\pi}$.

METHOD. If $Y_1, \ldots, Y_d, \ldots$ is a sequence of independent random variables with the same distribution ν and if $R_d = [Y_1^2 + \cdots + Y_d^2]^{1/2}$, use the fact that σ_{d-1} is the distribution of $R_d^{-1}(Y_1, Y_2, \ldots, Y_d)$, the weak law of large numbers of Problem IV-6, and Problem I-10.

REMARK. This property of uniform distributions on spheres is known as Poincaré's lemma.

Problem IV-21. Let $\mathcal{S}_n$ denote the set of probability measures μ on $\mathbf{R}$ such that there exists a probability measure μ_n on the Euclidean space $\mathbf{R}^n$ whose image in $\mathbf{R}$ under $x \mapsto \langle a,x\rangle$ is μ for *every* a in the unit sphere of $\mathbf{R}^n$. Prove that $\mu \in \cap_{n=1}^{\infty}\mathcal{S}_n$ if and only if there exists a probability measure ρ on $[0,+\infty)$ such that the Fourier transform of μ satisfies $\widehat{\mu}(t) = \int_0^\infty \exp(-\frac{y^2t^2}{2})\rho(dy)$. Prove that such a ρ, if it exists, is unique.

METHOD. For the uniqueness of ρ, use Problem II-20. For its existence, use Problems III-1, III-2(4), and IV-20, as well as Paul Lévy's theorem on the convergence of distributions.

REMARK. This property is due to I. Schoenberg (1937) .

Problem IV-22. Let $(X_0, X_1, \ldots, X_d)$ be an $\mathbf{R}^{d+1}$-valued random variable that is radial, i.e. whose distribution is invariant under the group O_{d+1} of $d \times d$ orthogonal matrices. Let $t = (t_1, t_2, \ldots, t_d)$ and $\|t\| = [t_1^2 + \cdots + t_d^2]^{\frac{1}{2}}$. Prove that $\mathbf{E}[\exp(i \sum_{j=1}^d t_j X_j - \|t\| X_0)] = 1$ for every t in $\mathbf{R}^d$ such that $\mathbf{E}[\exp(-\|t\| X_0)] < \infty$.

METHOD. Prove the assertion first for $d = 1$ and μ concentrated on the unit circle.

Problem IV-23. Let $\{(V_n, W_n)\}_{n=1}^\infty$ be a sequence of independent random variables with the same distribution, with values in $\mathbf{R} \times \mathbf{R}^d$ (where $\mathbf{R}^d$ has the Euclidean structure), and satisfying $\mathbf{E}[\log |V_1|] < 0$ and $\mathbf{E}[\log^+ \|W_1\|] < \infty$.
(1) Prove that $\sum_{n=0}^\infty |V_1 \ldots V_n| \; \|W_{n+1}\|$ converges almost surely.

METHOD. Use the Borel-Cantelli lemma to show that $\limsup_{n \to \infty} \|W_n\|^{1/n} \leq 1$, then use the strong law of large numbers. (See Problem IV-7.)
(2) Let μ be the distribution of the $\mathbf{R}^d$-valued random variable which is equal to the sum of the series $\sum_{n=0}^\infty V_1 \ldots V_n W_{n+1}$. Let ν be a distribution on $\mathbf{R}^d$ whose Fourier transform $\widehat{\nu}$ satisfies

$$\widehat{\nu}(t) = \mathbf{E}[\widehat{\nu}(V_1 t) \exp(i \langle W_1, t \rangle)] \quad \text{for every } t \text{ in } \mathbf{R}^d.$$

Show that $\mu = \nu$.
(3) Let $\{U_n\}_{n=0}^\infty$ be a sequence of independent $\mathbf{R}^{d+1}$-valued random variables with the same distribution, the uniform distribution on the unit sphere S_d of $\mathbf{R}^{d+1}$. Let $V_n - 1$ and W_n be the projections of U_n onto $(\mathbf{R}, 0, 0, \ldots)$ and onto its orthogonal complement. Prove that if μ is the distribution of $\sum_{n=0}^\infty V_1 \ldots V_n W_{n+1}$, then $\widehat{\mu}(t) = \exp(-\|t\|)$.

METHOD. Use (2) and Problem IV-22.
(4) Let $\{X_n\}_{n=1}^\infty$ be a sequence of independent random variables with values in $\mathbf{N} = \{0, 1, 2, \ldots\}$ and with the same distribution, such that X_1 satisfies $p_k = P[X_1 = k] < 1$ for every k in $\mathbf{N}$. Set $q_k = P[X_1 < k]$. Show that if μ is the distribution of $\sum_{n=0}^\infty p_{X_1} p_{X_2} \cdots p_{X_n} q_{X_{n+1}}$, then μ is Lebesgue measure on $[0, 1]$.

Problem IV-24. Let X and Y be independent random variables with the same distribution and with values in Euclidean space $\mathbf{R}^d$, $d > 1$, which satisfy the following conditions: (i) $P[X = 0] = 0$; (ii) $\frac{X}{\|X\|}$ and $\|X\|$ are independent; and (iii) $\frac{X}{\|X\|}$ is uniformly distributed on the sphere S_{d-1}. (That is, the distribution of X is "radial" — see Problem III-2(4).) Prove that

$$P[\|2X - Y\| \leq \|Y\|] < \frac{1}{4}$$

and that this inequality is the best possible.

METHOD. Consider $R = \frac{\|X\|}{\|Y\|}$, use the fact that R and R^{-1} have the same distribution on $(0, +\infty)$, and prove the inequality by first conditioning with respect to $|\log R|$. For the second part, take $\|X\|$ with density $\frac{1}{n}x^{(1-n)/n}$ on $(0, 1]$ and show that the distribution ν_n of $\exp(-|\log R|)$ tends vaguely to the Dirac measure at 0.

REMARKS. 1. There is also an explicit expression,

$$P[\|2X - Y\| \leq \|Y\|] = \frac{1}{4}\int_0^\infty G(a)d\nu^*(a),$$

where $G(y) = \frac{1}{B(\frac{1}{2}, \frac{d-1}{2})}\int_y^\infty \frac{dx}{\sqrt{x}(1+x)^{d/2}}$ and $\nu^*(da)$ is the distribution of $A^2/(1 - A^2)$.
2. This inequality is due to A.O. Pittenger, who proves it with the additional hypothesis $P[\|X\| = x] = 0$ for all $x \geq 0$ (1981).
3. Relaxing the hypothesis of the problem to $P[\|X\| = 0] = 0$ easily yields the upper bound

$$P[\|2X - Y\| \leq \|Y\|] < p + (1 - p)^2\frac{1}{4},$$

where $p = P[\|X\| = 0] < 1$, and this again is best possible. Note also that $P[\|2X - Y\| < \|Y\|] = (1 - p)^2/4 < 1/4$ in all cases.

Problem IV-25. Let H be a separable Hilbert space and let p_U denote the orthogonal projection of H onto a subspace U. Define the Boolean algebra $\mathcal{B}$ of subsets B of H for which there exists a finite-dimensional subspace V of H and a Borel set B_V of V such that $B = p_V^{-1}(B_V)$. Let $\sigma(\mathcal{B})$ denote the σ-algebra generated by $\mathcal{B}$.
(1) Show that $\{x : \|x\| \leq r\} \in \sigma(\mathcal{B})$ if $r > 0$.

METHOD. Use the fact that, since H is separable, there exists an increasing sequence $\{V_n\}_{n=1}^\infty$ of finite-dimensional subspaces of H such that $\cup_{n=1}^\infty V_n$ is dense in H.
(2) A *cylindrical probability* on H is given by probabilities μ_V on each finite-dimensional subspace V of H such that, if $V_1 \subset V_2$, the image of μ_{V_2} under p_{V_1} is μ_{V_1}. For $B \in \mathcal{B}$, let E_B denote the set of finite-dimensional subspaces V such that there exists a Borel subset B_V of V with $B = p_V^{-1}(B_V)$. Prove that $V \mapsto \mu_V(B_V)$ is constant on E_B. Denoting this constant by $\mu(B)$, prove that μ is finitely additive on $\mathcal{B}$.
(3) Consider the cylindrical probability defined as follows. Let ρ be a probability measure on $[0, +\infty)$ and let μ_V be defined by its Fourier transform,

$$\widehat{\mu}_V(t) = \int_V \exp(i\langle x, t\rangle)\mu_V(dx) = \int_0^\infty \exp\left(-\frac{y^2\|t\|^2}{2}\right)\rho(dy) \quad \text{for } t \in V.$$

Show that μ is not σ-additive on $\mathcal{B}$ if $\rho(\{0\}) < 1$.

METHOD. Otherwise μ could be extended to a σ-additive probability measure μ on $\sigma(\mathcal{B})$. Use Problems I-10 and IV-6 to show that this would imply $\mu(\{x : \|x\| \leq r\}) = \rho(\{0\})$ for $r > 0$.

Problem IV-26. In Euclidean space $\mathbf{R}^n$, consider the positive quadratic form q defined by $q(x) = \sum_{k=1}^n \lambda_k x_k^2$, where $x = \{x_k\}_{k=1}^n$ and $\lambda_k \geq 0$. Set $\|q\| = \sum_{k=1}^n \lambda_k$.
(1) If X is an $\mathbf{R}^n$-valued random variable such that

$$\mathbf{E}(\exp(i\langle X, t\rangle)) = \exp\left(-\frac{\|t\|^2}{2}\right),$$

show that $P[q(X) \geq r^2] \leq \frac{\|q\|}{R^2}$ for every $r > 0$.

METHOD. Use Chebyshev's inequality, Problem IV-6.

(2) Let μ be a probability measure on $\mathbf{R}^n$ with Fourier transform $\widehat{\mu}(t) = \int_{\mathbf{R}^n} \exp(i\langle x, t\rangle)\mu(dx)$ and let $\epsilon > 0$ be such that $|1 - \widehat{\mu}(t)| \leq \epsilon$ for every t in $\mathbf{R}^n$ with $q(t) \leq 1$. Prove that, for every $r > 0$,

$$\int_{\mathbf{R}^n} \exp\left(-\frac{\|x\|^2}{2r^2}\right)\mu(dx) \geq 1 - \epsilon - \frac{2\|q\|}{r^2}.$$

(3) Prove that, for every r, $R > 0$,

$$\mu(\{x : \|x\| \leq R\} \geq 1 - \epsilon - \frac{2\|q\|}{r^2} - \exp\left(-\frac{R^2}{2r^2}\right).$$

Conclude that there exists a number $R(\|q\|, \epsilon)$ such that

$$\mu(\{x : \|x\| \leq R(\|q\|, \epsilon)\} \geq 1 - 2\epsilon.$$

REMARK. This result is called Minlos's lemma (1959).

Problem IV-27. The notation is that of Problem IV-25 and $\mu = (\mu_V)_V$ is a cylindrical probability on H. A positive quadratic form q on H is a bounded linear mapping $A : H \to H$ such that $q(x) = \langle Ax, x\rangle \geq 0$ for every x. If the dimension of V is n, there exist a basis $b = \{b_1, \ldots, b_n\}$ of V and nonnegative numbers $\lambda_1, \ldots, \lambda_n$ such that, if $\sum_{k=1}^n x_k b_k$ is in V, then $q(x) = \sum_{k=1}^n \lambda_k x_k^2$. Moreover, the distribution of the $\{\lambda_k\}_{k=1}^n$ is independent of b, and we may set $\|q_V\| = \sum_{k=1}^n \lambda_k$. This implies that $\|q_{V_1}\| \leq \|q_{V_2}\|$ if $V_1 \subset V_2$, and we set $\|q\| = \sup_V \|q_V\| \leq +\infty$.
(1) Let $\widehat{\mu}_V(t) = \int_V \exp(i\langle x, t\rangle)\mu_V(dx)$ for $t \in V$. Show that $\widehat{\mu}_{V_1}(t) = \widehat{\mu}_{V_2}(t)$ if $t \in V_1 \cap V_2$.
(2) Set $\widehat{\mu}(t) = \widehat{\mu}_V(t)$ if $t \in V$. Suppose that, for all $\epsilon > 0$, there exists a positive quadratic form q_ϵ on H such that $\|q_\epsilon\| < \infty$ and $|1 - \widehat{\mu}(t)| \leq \epsilon$ for all t such that $q_\epsilon(t) \leq 1$. Deduce from Problem IV-26 that, for all $\epsilon > 0$, there exists $R(\epsilon)$ such that

$$\mu_V(\{x : x \in V \text{ and } \|x\| \leq R(\epsilon)\}) \geq 1 - 2\epsilon \quad \text{for every } V.$$

(3) With the preceding hypotheses, prove that μ is a σ-additive probability measure on the Boolean algebra $\mathcal{B}$ by showing that, if $A_n \in \mathcal{B}$, $A_n \supset A_{n+1}$, and $\mu(A_n) \geq \delta > 0$ for every n, then $\cap_{n=1}^{\infty} A_n \neq \emptyset$.

METHOD. Let V_n be a finite-dimensional subspace of H containing a Borel set A'_n such that $A_n = p_{V_n}^{-1}(A'_n)$ and let $B'_n(R)$ be the closed ball of radius R in V_n. We may assume that $V_n \subset V_{n+1}$. Construct compact sets K'_n of V_n contained in $A'_n \cap B'_n(R)$, introduce $K_n = p_{V_n}^{-1}(K'_n)$, and use the fact that $C_n = K_n \cap \ldots \cap K_n \cap \{x : \|x\| \leq R\}$ is a decreasing sequence of compact sets in the weak topology on H.
REMARK. This result is due to Minlos (1959).

Problem IV-28. Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with the same distribution defined by $P[X_n = 1] = P[X_n = -1] = 1/2$. Compute the limiting distribution as $n \to \infty$ of

$$Y_n = [1 + 4 + 9 + \ldots + n^2]^{-1/2}[X_1 + 2X_2 + 3X_3 + \ldots + nX_n].$$

METHOD. Consider the characteristic function of Y_n.
REMARK. This is a simple special case of Lindeberg's theorem, which is a significant generalization of Laplace's theorem, IV-4.3.1 (also often called the central limit theorem). Lindeberg's theorem is stated as follows: If (i) the real random variables $\{X_n\}_{n=1}^{\infty}$ are independent (but do not necessarily have the same distribution); (ii) for every n, $\mathbf{E}(X_n) = 0$ and $\sigma_n^2 = \mathbf{E}[(X_1 + \cdots + X_n)^2] < \infty$; and (iii) for every ϵ,

$$\mathbf{E}\left[\sum_{k=1}^{n} f_\epsilon\left(\frac{X_k}{\sigma_n}\right)\right] \to 0, \quad \text{where } f_\epsilon(x) = x^2 \mathbf{1}_{[\epsilon,+\infty)}(x),$$

then the distribution of $\frac{1}{\sigma_n}(X_1 + \cdots + X_n)$ tends to the Gaussian distribution $N(0,1)$ as above.

Problem IV-29. On the real line, consider the Gaussian distribution $\mu(dx) = \frac{1}{2\pi}\exp(\frac{x^2}{2})dx$. Let $L^2(\mu)$ be the Hilbert space of functions which are square integrable with respect to μ, with the scalar product

$$\langle f, g\rangle = \int_{-\infty}^{+\infty} f(x)\overline{g(x)}\mu(dx).$$

The Hermite polynomials $\{H_n(x)\}_{n=0}^{\infty}$ are defined by

$$\sum_{n=0}^{\infty} H_n(x)(it)^n = \exp(itx + \frac{t^2}{2}) = \varphi(t,x) \quad \forall t \in \mathbf{C}.$$

Assume without proof that this implies

$$(*) \qquad \sum_{n=0}^{\infty} |H_n(x)|\ |t|^n \leq \exp\left(|t|\ |x| + \frac{|t|^2}{2}\right).$$

(1) By computing $\langle \varphi(t,.), \varphi(s,.)\rangle$ in two different ways, show that $\langle H_n, H_m\rangle = 0$ if $n \neq m$ and that $\langle H_n, H_n\rangle = \frac{1}{n!}$. Use the uniqueness of the Fourier transform to show that if f in $L^2(\mu)$ satisfies $\langle f, H_n\rangle = 0$ for every n, then $f = 0$.
(2) Show that $H'_{n-1}(x) = H_n(x)$ and that $(n+1)H_{n+1}(x) = xH_n(x) - H_{n-1}(x)$ if $n \geq 1$.
(3) Let $f \in L^2(\mu)$ and let $f_n = n!\langle f, H_n\rangle$. Show that $f = \sum_{n=0}^{\infty} f_n H_n$ (where the convergence of the series is in the $L^2(\mu)$ sense). If, moreover, f' exists (in the sense that $F(x) = f(0) + \int_0^x f'(t)dt$ for every x) and belongs to $L^2(\mu)$, show that $f' = \sum_{n=0}^{\infty} f_{n+1}H_n$.

METHOD. Compute $\langle f', H_n\rangle$ by means of an integration by parts and (2).
(4) Prove H. Chernoff's inequality: If X is a Gaussian random variable with distribution μ and if f is a real-valued function such that both $f'\mathbf{E}[|f'(X)|^2]$ and $\mathbf{E}[|f(X)|^2]$ exist, then $\mathbf{E}[|f'(X)|^2] \geq$ variance of $f(X)$. Analyze the case of equality.

Problem IV-30. Let (X,Y) be a Gaussian random variable with values in $\mathbf{R}^2$ such that X and Y have distribution $\mu(dx) = (2\pi)^{-1/2}\exp(-\frac{x^2}{2})dx$.
(1) For the Hermite polynomials defined in Problem IV-29, prove that

$$H_n(y\cos\theta + z\sin\theta) = \sum_{k=0}^{n} H_k(y)\cos^k\theta H_{n-k}(z)\sin^{n-k}\theta.$$

(2) Assume that $\cos\theta = \mathbf{E}(XY) \neq \pm 1$ and define the random variable $Z = \dfrac{X - Y\cos\theta}{\sin\theta}$. Verify that Y and Z are independent and use (1) to prove that $\mathbf{E}[H_n(X)|Y] = H_n(Y)(\mathbf{E}(XY))^n$.
(3) Prove Gebelein's inequality: If $f \in L^2(\mu)$ with $\mathbf{E}(f(X)) = 0$, then

$$\mathbf{E}[\mathbf{E}[f(X)|Y])^2] \leq (\mathbf{E}(XY))^2\mathbf{E}(f^2(X)).$$

Analyze the case of equality.

METHOD. Write $f = \sum_{n=1}^{\infty} f_n H_n$ as in Problem IV-29.

Problem IV-31. Let H_n be the nth Hermite polynomial described in Problem IV-29 and compute

$$\int_{-\infty}^{+\infty} \mathrm{e}^{ixt}H_n(x)\mathrm{e}^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = \widehat{H_n\mu}(t).$$

Use this to find

$$\int_{-\infty}^{+\infty} \mathrm{e}^{ixt}x^n\mathrm{e}^{-x^2/2}\frac{dx}{\sqrt{2\pi}}.$$

Problem IV-32. Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\mathcal{B}$ be a sub-

σ-algebra of $\mathcal{A}$. We would like to show that if $X \in L^1(\mathcal{A})$, then

$$(*) \qquad \int_B X dP = \int_B \mathbf{E}[X|\mathcal{B}]dP \quad \text{for all } B \in \mathcal{B},$$

and that $(*)$ characterizes $\mathbf{E}[X|\mathcal{B}]$.
(1) Show that $(*)$ holds if $X \in L^2(\mathcal{A})$.
(2) If $X > 0$, let $L(X) = \lim_{n\to+\infty} \mathbf{E}[\min(X,n)|\mathcal{B}]$. If $X \in L^1(\mathcal{A})$, let $L(X) = L(X^+) - L(X^-)$, where $X^+ = \max(X,0)$ and $X^- = \max(-X,0)$. Show that $L(X) \in L^1(\mathcal{B})$ and that $\int_B (X - L(X))dP = 0$ for all B in $\mathcal{B}$.
(3) Show that if f, $g \in L^1(\mathcal{B})$ are such that $\int_B (f-g)dP = 0$ for every B in $\mathcal{B}$, then $f = g$.
(4) Show that $L(X)$ is a bounded linear operator from $L^1(\mathcal{A})$ to $L^1(\mathcal{B})$ and infer that $L(X) = \mathbf{E}(X|B)$.

REMARK. This characterization of conditional expectation is often taken as a definition in the literature.

Problem IV-33. Suppose that, for every $n \geq 0$, $X_n \in L^1(\mathcal{A})$ and $X_n \geq 0$. Use the preceding problem to show that if $X_n \uparrow X_0$, then

$$Y_n = \mathbf{E}[X_n|\mathcal{B}] \uparrow \mathbf{E}[X_0|\mathcal{B}].$$

Problem IV-34. Suppose that $(\Omega, \mathcal{A}, P)$ is a probability space, $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{A}$, Y is a $\mathcal{B}$-measurable random variable, and X is a random variable independent of $\mathcal{B}$. Consider $f : \mathbf{R}^2 \to \mathbf{R}$ such that $f(X,Y)$ is integrable. The goal of this problem is to show that if μ is the distribution of X, then

$$(*) \qquad \mathbf{E}[f(X,Y)|\mathcal{B}] = \int_{-\infty}^{+\infty} f(x,Y)\mu(dx).$$

(1) Show that $(*)$ holds if $f(x,y) = \mathbf{1}_I(x)\mathbf{1}_J(y)$, where I and J are Borel subsets of $\mathbf{R}$.
(2) Let $\mathcal{P}$ be the Boolean algebra on $\mathbf{R}^2$ consisting of sets of the form $E = \cup_{p=1}^q I_p \times J_p$, where I_p and J_p are Borel subsets of $\mathbf{R}$. Show that $(*)$ holds if $f(x,y) = \mathbf{1}_E(x,y)$ with $E \in \mathcal{P}$.
(3) Let $\mathcal{M}$ be the family of Borel subsets M of $\mathbf{R}^2$ such that $f(x,y) = \mathbf{1}_M(x,y)$ satisfies $(*)$. Show that $\mathcal{M}$ is a monotone class.
(4) Prove $(*)$ successively for the following cases: (a) f is a simple function on $\mathbf{R}^2$; (b) f is a positive measurable function with $f(X,Y)$ integrable; and (c) the general case.

Problem IV-35. On a probability space $(\Omega, \mathcal{A}, P)$, consider an integrable random variable X and a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$, both independent of another sub-σ-algebra $\mathcal{C}$ of $\mathcal{A}$. Prove that if $\mathcal{D}$ is the σ-algebra generated by $\mathcal{B} \cup \mathcal{C}$, then

$$\mathbf{E}[X|\mathcal{D}] = \mathbf{E}[X|\mathcal{B}].$$

METHOD. Prove the assertion first for square integrable X.

Problem IV-36. If X and Y are integrable random variables such that $\mathbf{E}[X|Y] = Y$ and $\mathbf{E}[Y|X] = X$, show that $X = Y$ a.s.

METHOD. Show that, for fixed x,

$$(i) \qquad 0 \le \int_{Y \le x \le X} (X - Y)dP = \int_{x<X \text{ and } x<Y} (Y - X)dP,$$

and conclude by symmetry that both sides of the equation are zero. Then use Problem I-13.

Problem IV-37. Suppose that $(\Omega, \mathcal{A}, P)$ is a probability space, X and Y are integrable random variables, and $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{A}$ such that X is $\mathcal{B}$-measurable.
(1) Show that $\mathbf{E}[Y|\mathcal{B}] = X$ implies $\mathbf{E}[Y|X] = X$.
(2) Show by a counterexample that $\mathbf{E}[Y|X] = X$ does not imply that $\mathbf{E}[Y|\mathcal{B}] = X$.

REMARK. If $\{\mathcal{A}_n\}_{n\ge0}$ is a filtration of $(\Omega, \mathcal{A}, P)$, $\{X_n\}_{n\ge0}$ a sequence adapted to this filtration, and $\mathcal{B}_n$ the σ-algebra generated by $X_0, \ldots, X_n$, then $\{X_n, \mathcal{B}_n\}_{n\ge0}$ is a martingale if $\{X_n, \mathcal{A}_n\}_{n\ge0}$ is. The converse is false.

Problem IV-38. Let $(Y_0, Y_1, \ldots, Y_n)$ be an $(n+1)$-tuple of real random variables defined on a probability space $(\Omega, \mathcal{E}, P)$. Let $\mathcal{F}$ denote the sub-σ-algebra of $\mathcal{E}$ generated by $(Y_1^{(\omega)}, \ldots, Y_n^{(\omega)}) = f(\omega)$ and assume that $\mathbf{E}(|Y_0|) < \infty$.
(1) By applying Theorem IV-6.5.1 to f, show that there exists a Borel-measurable function $g : \mathbf{R}^n \to \mathbf{R}$ such that

$$\mathbf{E}[Y_0|\mathcal{F}] = g(Y_1, Y_2, \ldots, Y_n) \quad P\text{-almost everywhere.}$$

(2) Assume that the distribution of $(Y_0, Y_1, \ldots Y_n)$ in $\mathbf{R}^{n+1}$ is absolutely continuous with respect to Lebesgue measure dy_0, $dy_1, \ldots, dy_n$, and let $d(y_0, y_1, \ldots, y_n)$ denote its density. Prove that

$$\mathbf{E}(Y_0|\mathcal{F}) = [K(Y_1, Y_2, \ldots, Y_n)]^{-1} \int_{-\infty}^{+\infty} y_0 \; d(y_0, Y_1, \ldots, Y_n)dy_0,$$

where $K(y_1, \ldots, y_n) = \int_{-\infty}^{+\infty} d(y_0, y_1, \ldots, y_n)dy_0$. Prove that if A is a Borel subset of $\mathbf{R}$, then

$$\begin{aligned} P[Y_0 \in A|\mathcal{F}] &= \mathbf{E}[\mathbf{1}_{Y_0 \in A}|\mathcal{B}] \\ &= [K(Y_1, \ldots, Y_n)]^{-1} \int_A d(y_0, Y_1, \ldots, Y_n)dy_0. \end{aligned}$$

(3) Assume that the distribution of $(Y_0, Y_1, \ldots, Y_n)$ in $\mathbf{R}^{n+1}$ is Gaussian (with the definition in IV-4.3.4, which implies that $\mathbf{E}(Y_j) = 0$ for $j =$

$0, \ldots, n)$. Use the observation that if $(X, Y_1, \ldots, Y_n)$ is Gaussian in $\mathbf{R}^{n+1}$, then X is independent of $(Y_1, \ldots, Y_n)$ if and only if $\mathbf{E}(XY_j) = 0 \ \forall j = 1, \ldots, n$, to show that there exist real numbers $\lambda_1, \ldots, \lambda_n$ such that $\mathbf{E}[Y_0|\mathcal{F}] = \lambda_1 Y_1 + \cdots + \lambda_n Y_n$.

Problem IV-39. Let $\{X_n\}$ be a sequence of independent real random variables with the same distribution and let $\mathcal{F}_n$ be the σ-algebra generated by $X_1, \ldots, X_n$. Set $S_n = X_1 + \cdots + X_n$ for $n > 0$ and set $S_0 = 0$. Which of the following processes are martingales relative to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$?
(1) S_n, if $\mathbf{E}(|X_1|) < \infty$.
(2) $X_1^2 + \cdots + X_n^2 - n\lambda$, if $\mathbf{E}(X_1^2) < \infty$ and λ is real.
(3) $\exp(\alpha S_n - n\lambda)$, if $\varphi(\alpha) = \mathbf{E}(\exp(\alpha X_1)) < \infty$ and α and λ are real.
(4) $Y_n = |S_{\min(n,T)}|$, where $T = \inf\{n > 0 : S_n = 0\}$, and we assume that $P[X_1 = 1] = P[X_1 = -1] = \frac{1}{2}$.

Problem IV-40. Let $Y_1, \ldots, Y_n, \ldots$ be independent real random variables with the same distribution and such that $\mathbf{E}[|Y_1|] < \infty$. Set $S_n = Y_1 + \cdots + Y_n$.
(1) Show that $\mathbf{E}[Y_k|S_n] = S_n/n$ if $1 \le k \le n$.
(2) If m is fixed and $X_k = S_{m-k}/(m-k)$ for $0 \le k \le m-1$, show that $(X_0, \ldots, X_{m-1})$ is a martingale. (Apply Problem IV-35.)

Problem IV-41. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent random variables with the same distribution defined by $P[X_n = k] = 2^{-k}$ for $k = 1, 2, \ldots$. Random variables Z_n are defined by letting Z_0 be a positive constant and setting $Z_n = (3Z_{n-1})/2^{X_n}$ for $n = 1, 2, \ldots$.
(1) Prove that $\{Z_n\}_{n=0}^{\infty}$ is a martingale relative to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$, where $\mathcal{F}_n$ is the σ-algebra generated by $X_1, \ldots, X_n$.
(2) Use the law of large numbers (see Problem IV-6) to prove that $Z_n \to 0$ almost surely as $n \to \infty$.

REMARK. This gives a heuristic confirmation of the following unproved conjecture in number theory. If n is an odd positive integer, let $f(n) = (3n+1)2^{-\nu(3n+1)}$, where $2^{\nu(3n+1)}$ denotes the greatest power of 2 that divides the integer $3n+1$. The conjecture asserts that, for every n, there exists an integer k such that the kth iterate of f satisfies $f^{(k)}(n) = 1$. If n is very large, $\nu(3n+1)$ appears to behave like the variable X_1 of the problem, and $\{Z_k\}_{k=1}^{\infty}$ like the sequence $\{f_k(n)\}_{k=1}^{\infty}$.

Problem IV-42. Let $H \subset L^1(\Omega, \mathcal{A}, P)$, where $(\Omega, \mathcal{A}, P)$ is a probability space.
(1) If F is a positive function on $(0, +\infty)$ such that $F(x)/x$ is increasing and $\to +\infty$ as $n \to \infty$, and if

$$\sup_{h \in H} \mathbf{E}(F|h) = M < \infty,$$

show that H is uniformly integrable.

METHOD. Use Proposition IV-5.7.2.
(2) If H is a bounded subset of $L^p(\Omega, \mathcal{A}, P)$ with $p > 1$, show that H is uniformly integrable.

Problem IV-43. Let $\{X_{ij}\}_{i,j=1}^{\infty}$ be independent random variables with values in $\mathbf{N}$ and with the same distribution. Assume that $0 < m = \mathbf{E}(X_{11}) < \infty$ and that $\sigma^2 = \mathbf{E}((X_{11} - m)^2) < \infty$. Consider the sequence of random variables defined by

$$\begin{array}{lcllr} Z_0 & = & 1 & & \\ Z_{n+1} & = & 0 & \text{if} & Z_n = 0 \\ Z_{n+1} & = & \sum_{i=1}^{Z_n} X_{i,n+1} & \text{if} & Z_n > 0. \end{array}$$

$\mathcal{F}_n$ is the σ-algebra generated by $\{X_{i,j} : 1 \leq i < \infty,\ 1 \leq j \leq n\}$.
(1) Show that $\{Z_n/m^n, \mathcal{F}_n\}_{n=1}^{\infty}$ is a martingale.
(2) Show that $\mathbf{E}\left(Z_{n+1}^2/m^{2(n+1)}\right) = \mathbf{E}\left(Z_n^2/m^{2n}\right) + \sigma^2/m^{2n+1}$.

Conclude that, if $m > 1$, the martingale is regular. (Use Problem IV-42 and Theorem IV-5.8.1.)

REMARK. $\{Z_n\}_{n=0}^{\infty}$ is sometimes called the Galton-Watson process, and serves as a model in genetics. ($X_{i,j}$ is the number of offspring of the individual i of the jth generation, which has total size Z_j.)

Exercises for Chapter V

Problem V-1. Let E be the set of compactly supported C^∞ functions on $\mathbf{R}$, and let d and δ be the operators on E defined by

$$(d\varphi)(x) = \varphi'(x) \quad \text{and} \quad (\delta\varphi)(x) = -\varphi'(x) + x\varphi(x).$$

(1) Prove by induction on n that

$$d^n\delta - \delta d^n = nd^{n-1}.$$

(2) Let p be a norm on E. Let B be the algebra of operators on E which are continuous with respect to this norm, that is the set of endomorphisms a of E such that

$$\|a\| = \sup\{p(a(\varphi)) : p(\varphi) \leq 1\}$$

is finite. Assume that d and δ are in B. Use (1) to prove that, for all $n \geq 1$,

$$\|nd^{n-1}\| \leq 2\|d^{n-1}\| \; \|d\| \; \|\delta\|.$$

(3) Deduce from (2) that d and δ are never simultaneously continuous.

REMARK. This result is due to Aurel Wintner (1947).

Problem V-2. Let $\{H_n\}_{n=0}^{\infty}$ be the sequence of Hermite polynomials defined in V-1.3.
(1) Use Proposition V-1.3.4 to show that, for $n \geq 1$,

$$H_{n+1} + nH_{n-1} = xH_n.$$

If $\widetilde{H}_n = \frac{H_n}{n!}$ (compare with Problem IV-29), show that

$$\widetilde{H}_{n+1} = \frac{x}{n+1}\widetilde{H}_{n+1} - \frac{1}{n+1}\widetilde{H}_{n-1}.$$

(2) Conclude from (1) that the radius of convergence $R(x)$ of $\sum_{n=0}^{\infty} \frac{z^n}{n!} H_n(x)$ is $+\infty$ for every complex number x.

METHOD. For (2), show that for every $\epsilon > 0$ there exist an integer $N(\epsilon)$ and a sequence $\{x_n\}_{n=N(\epsilon)-1}^{\infty}$ such that $|\widetilde{H}_n| \leq x_n$ and $x_{n+1} = \epsilon x_n + \epsilon x_{n-1}$.

Problem V-3. Let $\{H_n\}_{n=0}^{\infty}$, d, and δ be defined as in V-1.3. For nonnegative integers n, consider

$$F_n(x) = H_n(ix)(-i)^n.$$

Let $\lambda \in \mathbf{C}$ and define ρ by $\rho = \delta + \lambda d$.
(1) For $n \geq 1$, prove that $d^n \rho = \rho d^n + n d^{n-1}$ and $F_{n+1} = xF_n + nF_{n-1}$.
(2) Prove by induction on n that

$$(d+\rho)^n = \sum_{k=0}^{n} C_n^k H_k(\rho) d^{n-k},$$

where C_n^k denotes the binomial coefficient.
(3) If φ is a polynomial and t is real, let $\tau_t(\varphi)(x) = \varphi(x+t)$. Prove that $(\exp(td))(\varphi) = \tau_t(\varphi)$ and that

$$(\exp\ t(d+\rho))(\varphi) = \left(\exp \frac{t^2}{2} \exp(t\rho)\tau_t\right)(\varphi).$$

In particular, if $\lambda = 1$ (that is, if $\rho(\varphi)(x) = x\varphi(x)$), compute $(\exp\ t(d + x))(\varphi)$.

REMARK. The result of (2) is due to Viskov[4]; that of (3) is due to Ville.[5]

Problem V-4. Let X and Y be independent random variables with the same distribution $\nu_1(dx) = \exp(-x^2/2)dx/\sqrt{2\pi}$. Let $g : \mathbf{R} \to [0, +\infty)$ be a measurable function and let $Z = X + Y\sqrt{g(X)}$. Assume that Z has a normal distribution. Cantelli conjectured in 1917 that g is then constant almost everywhere; this is still unproved in 1994.
(1) Let $g_0 = \mathbf{E}(g(X))$. For all real t, compute $\mathbf{E}(\exp\ tZ)$ as a function of g_0. Prove that $\exp(\alpha g) \in L^2(\nu_1)$ for all $\alpha > 0$.

METHOD. Use the Cauchy-Schwarz inequality.

---

[4]O. Viskov, *Theory of Probability and Its Applications*, Vol. 30, n. 1 (1984), 141–143.
[5]J. Ville, *Comptes Rendus Acad. des Sc. 221* (1945), 529–539.

(2) Let $\{g_n\}_{n=0}^{\infty}$ be the sequence of real numbers such that $g(x) = \sum_{n=0}^{\infty} g_n \frac{H_n(x)}{n!}$ in the $L^2(\nu_1)$ sense. By considering $\mathbf{E}(Z^3)$ and $\mathbf{E}(Z^4)$, show that $g_1 = 0$ and $-2g_2 = \sum_{n=2}^{\infty} g_n^2/n!$.
(3) Prove that $g(x) \leq g_0 + 1$ almost everywhere.

METHOD. If $\epsilon > 0$, let $A_\epsilon = \{x : g(x) \geq \epsilon + g_0 + 1\}$, and let a be a real number such that $A'_\epsilon = A_\epsilon \cap [a, +\infty)$ has positive measure. Consider

$$\int_{A'_\epsilon} \exp[tx + \frac{t^2}{2}(g(x) - 1 - g_0)]\nu_1(dx).$$

Problem V-5. As usual, we denote by $\{H_n\}_{n\geq 0}$ the sequence of Hermite polynomials and by ν_1 the normal distribution on $\mathbf{R}$. Let μ be a probability distribution on $\mathbf{R}^2$ such that if (X, Y) has distribution μ, then X and Y have distribution ν_1 and there exists a real sequence $\{C_n\}_{n\geq 0}$ with

$$\mathbf{E}(H_n(X)|Y) = C_n H_n(Y).$$

(1) Prove that $C_n = \mathbf{E}(H_n(X)H_n(Y))$ and $-1 \leq C_n \leq 1$ for all n in $\mathbf{N}$.
(2) Prove that if $\sum_{n\geq 1} C_n^2 < +\infty$, then μ is absolutely continuous with respect to $\nu_1(dx)\nu_1(dy)$ and its density is

$$f(x, y) = \sum_{n\geq 0} \frac{C_n}{n!} H_n(x)H_n(y).$$

METHOD. For (2), write $\mu(dx, dy) = \nu_1(dy)K(y, dx)$. Show that the function $x \mapsto f(x, y)$ is in $L^2(\nu_1)$ y-almost everywhere and that, for every $\theta \in \mathbf{C}$,

$$\int \exp(\theta x)(f(x, y)\nu_1(dx) - K(y, dx)) = 0 \quad y\text{-a.e.}$$

Problem V-6. We keep the notation of Problem V-5 and denote by $\mathcal{C}$ the set of probability measures μ on $\mathbf{R}^2$ described there. Let μ be a fixed element of $\mathcal{C}$.
(1) Define $\{b_{n,k}\}_{0\leq k\leq n}$ by

$$x^n = \sum_{k=0}^{n} b_{n,k} H_k(x)$$

and let

$$P_n(y) = \sum_{k=0}^{n} b_{n,k} C_k H_k(y).$$

Show that $\int x^n K(y, dx) = P_n(y)$ y-a.e. and that $\lim_{y\to\infty} y^{-n} P_n(y) = C_n$.
(2) Let $\sigma(y, dt)$ be the image of $K(y, dx)$ under the mapping $x \mapsto x/y$. For $\theta \in \mathbf{C}$, show that

$$\int_{-\infty}^{+\infty} \exp(\theta t)\sigma(y, dt) = \exp\left(\frac{\theta^2}{2y^2}\right) \sum_{k=0}^{\infty} \frac{C_k}{k!} \theta^k y^{-k} H_k(y)$$

and

$$\lim_{y \to \infty} \int_{-\infty}^{+\infty} \exp(\theta t) \sigma(y, dt) = \sum_{k=0}^{\infty} \frac{C_k}{k!} \theta^k.$$

(3) Show that the probability measure $\sigma(dt) = \lim_{y \to \infty} \sigma(y, dt)$ exists and that

$$C_n = \int t^n \sigma(dt).$$

From the fact that $|C_n| \leq 1$, conclude that $\sigma(\mathbf{R} \setminus [-1, 1]) = 0$.
(4) Show that σ is the unique probability measure on $[-1, 1]$ such that $C_n = \int_{-1}^{1} t^n \sigma(dt)$.
(5) Show that the mapping $\mu \mapsto \sigma$, from $\mathcal{C}$ to the set of probability measures on $[-1, 1]$, is a bijection. What is μ when σ is the Dirac measure at ρ?

METHOD. For (5), consider successively the cases where $\rho = 1$, $\rho = -1$, and (using Problem V-5(2) and Mehler's formula, V-1.5.8(ii)) $\rho < 1$.

REMARK. This phenomenon was observed by O. Sarmanov (1966) and generalized by Tyan, Derin, and Thomas (1976).

Index

Graduate Texts in Mathematics

continued from page ii

119 ROTMAN. An Introduction to Algebraic Topology.
120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic K-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory:
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.
179 DOUGLAS. Banach Algebra Techniques in Operator Theory. 2nd ed.
180 SRIVASTAVA. A Course on Borel Sets.
181 KRESS. Numerical Analysis.
182 WALTER. Ordinary Differential Equations.

183 MEGGINSON. An Introduction to Banach Space Theory.
184 BOLLOBÁS. Modern Graph Theory.
185 COX/LITTLE/O'SHEA. Using Algebraic Geometry.
186 RAMAKRISHNAN/VALENZA. Fourier Analysis on Number Fields.
187 HARRIS/MORRISON. Moduli of Curves.
188 GOLDBLATT. Lectures on Hyperreals.
189 LAM. Lectures on Rings and Modules.