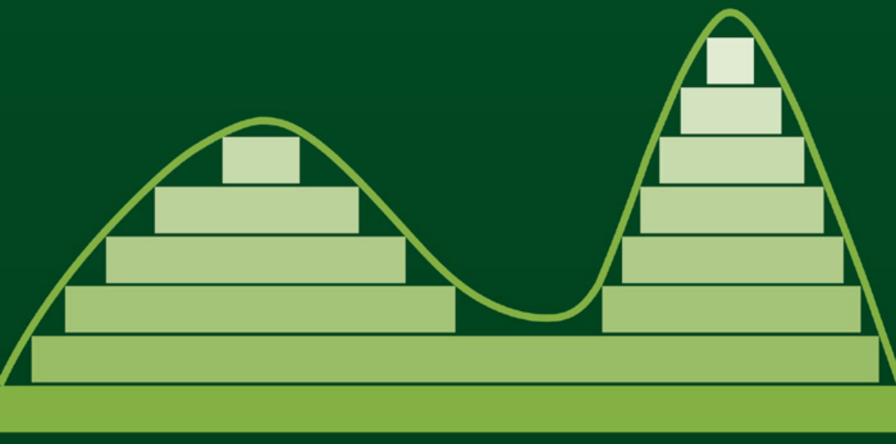
Measures, Integrals and Martingales

René L. Schilling



SECOND EDITION

MEASURES, INTEGRALS AND MARTINGALES

A concise yet elementary introduction to measure and integration theory, which are vital in many areas of mathematics, including analysis, probability, mathematical physics and finance. In this highly successful textbook the core ideas of measure and integration are explored, and martingales are used to develop the theory further. Additional topics are also covered such as: Jacobi's transformation theorem; the Radon–Nikodym theorem; differentiation of measures and Hardy–Littlewood maximal functions.

In this second edition, readers will find newly added chapters on Hausdorff measures, Fourier analysis, vague convergence, and classical proofs of the Radon–Nikodym and Riesz representation theorems. All proofs are carefully worked out with utmost clarity to ensure full understanding of the material and its background.

Requiring few prerequisites, this book is a suitable text for undergraduate lecture courses or self-study. Numerous illustrations and over 400 exercises help to consolidate and broaden the reader's knowledge. Full solutions to all exercises are available on the author's webpage at www.motapa.de.

RENÉ L. SCHILLING is a Professor of Mathematics at Technische Universität, Dresden. His main research area is stochastic analysis and stochastic processes.

MEASURES, INTEGRALS AND MARTINGALES

SECOND EDITION

RENÉ L. SCHILLING
Technische Universität, Dresden

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi – 110002, India

79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781316620243
DOI: 10.1017/9781316718339

First published 2005
Second edition 2017

Printed in the United Kingdom by Clays, St Ives plc

A catalogue record for this publication is available from the British Library.

ISBN 978-1-316-62024-3 Paperback

Contents

List of Symbols

This is intended to aid cross-referencing, so notation that is specific to a single section is generally not listed. Some symbols are used locally, without ambiguity, in senses other than those given below. Numbers following entries are page numbers, with the occasional (Pr *m.n*) referring to Problem *m.n* on the respective page.

Unless stated otherwise, binary operations between functions such as $f \pm g$, $f \cdot g$, $f \wedge g$, $f \vee g$, comparisons $f \leqslant g$, $f < g$ or limiting relations $f_n \xrightarrow[n\to\infty]{} f$, $\lim_n f_n$, $\liminf_n f_n$, $\limsup f_n$, $\sup_i f_i$ or $\inf_i f_i$ are always understood pointwise.

Alternatives are indicated by square brackets, i.e., '*if A* [*B*] … *then P* [*Q*]' should be read as '*if A … then P*' and '*if B … then Q*'.

Generalities

positive	always in the sense $\geqslant 0$
negative	always in the sense $\leqslant 0$
$\mathbb{N}$	natural numbers: $1, 2, 3, \ldots$
$\mathbb{N}_0$	positive integers: $0, 1, 2, \ldots$
$\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$	integer, rational, real, complex numbers
$\overline{\mathbb{R}}$	$[-\infty, +\infty]$
$\inf \emptyset, \sup \emptyset$	$\inf \emptyset = +\infty$, $\sup \emptyset = -\infty$
$a \vee b$	maximum of a and b
$a \wedge b$	minimum of a and b
$\liminf_n a_n$	$\sup_k \inf_{n \geqslant k} a_n$, 409
$\limsup_n a_n$	$\inf_k \sup_{n \geqslant k} a_n$, 409
$\lvert x \rvert$	Euclidean norm in $\mathbb{R}^n$, $\lvert x\rvert^2 = x_1^2 + \cdots + x_n^2$
$\langle x, y \rangle$	scalar product $\sum_{i=1}^n x_i y_i$
ω_n	volume of the unit ball in $\mathbb{R}^n$, 181

Sets and set operations

$A \cup B$	union 6
$A \uplus B$	union of disjoint sets, 6
$A \cap B$	intersection, 6
$A \setminus B$	set-theoretic difference, 6
A^c	complement of A, 6
$A \vartriangle B$	$(A \setminus B) \cup (B \setminus A)$
$A \subset B$	subset (includes '='), 6
$A \subsetneq B$	proper subset, 6
$A \times B$	Cartesian product
$\overline{A}$	closure of A
A°	open interior of A
$A_n \uparrow A$, $A_n \downarrow A$	23
$A \times B$	Cartesian product
A^n	n-fold Cartesian product
$A^{\mathbb{N}}$	infinite sequences in A
$\#A$	cardinality of A, 8
$t \cdot A$	$\{ta : a \in A\}$
$x + A$	$\{x + a : a \in A\}$
$E \cap \mathscr{A}$	$\{E \cap A : A \in \mathscr{A}\}$, 17
$\liminf_n A_n$	$\bigcup_{k\in\mathbb{N}} \bigcap_{n\geqslant k} A_n$, 413
$\limsup_n A_n$	$\bigcap_{k\in\mathbb{N}} \bigcup_{n\geqslant k} A_n$, 413
$(a, b), [a, b]$	open, closed intervals
$(a, b], [a, b)$	half-open intervals
$B_r(x)$	open ball with radius r and centre x

Families of sets

$\mathscr{A}$	generic σ-algebra, 16
$\overline{\mathscr{A}}$	completion 30 (Pr 4.15)
$(\mathscr{A}_i)_{i\in I}$	filtration, 276
$\mathscr{A}_\infty$	$\sigma(\mathscr{A}_i : i \in I) = \sigma\left(\bigcup_{i\in I} \mathscr{A}_i\right)$
$\mathscr{A}_{-\infty}$	$\bigcap_{\ell\in-\mathbb{N}} \mathscr{A}_\ell$
$\mathscr{A}_\tau$, $\mathscr{A}_\sigma$	283
$\mathscr{A} \times \mathscr{B}$	$\{A \times B : A \in \mathscr{A},\ B \in \mathscr{B}\}$ rectangles
$\mathscr{A} \otimes \mathscr{B}$	product-σ-algebra, 138
$\mathscr{B}(X)$	Borel sets in X, 18: $X = \mathbb{R}^n$ (18), $X = A \subset \mathbb{R}^n$ (22 Pr 3.13), $X = \overline{\mathbb{R}}$ (61), $X = \mathbb{C}$ (88 Pr 10.9, 423)
$\overline{\mathscr{B}(\mathbb{R}^n)}$	completion of the Borel sets, 151 (Pr 14.15), 172, 429

$\mathscr{J}, \mathscr{J}^o, \mathscr{J}_{\mathrm{rat}}$	rectangles in $\mathbb{R}^n$, 19
$\mathscr{N}_\mu$	μ-null sets, 29 (Pr 4.12), 89
$\mathscr{O}(X)$	topology, open sets in X, 18
$\mathscr{P}(X)$	all subsets of X, 13
$\delta(\mathscr{G})$	Dynkin system generated by $\mathscr{G}$, 32
$\sigma(\mathscr{G})$	σ-algebra generated by $\mathscr{G}$, 17
$\sigma(T), \sigma(T_i : i \in I)$	σ-algebra generated by the map(s) T, resp., T_i, 55
$(X, \mathscr{A})$	measurable space, 23
$(X, \mathscr{A}, \mu)$	measure space, 24
$(X, \mathscr{A}, \mathscr{A}_i, \mu)$	filtered measure space, 275, 276

Measures and integrals

μ, ν	generic measures
δ_x	Dirac measure in x, 26
λ, λ^n	Lebesgue measure, 27
$\mu \circ T^{-1}, T(\mu)$	image measure, 55
$u \cdot \mu, u\mu$	measure with density, 86
$\mu \times \nu$	product of measures, 141
$\mu \star \nu$	convolution, 157
$\mu \ll \nu$	absolute continuity, 300
$\mu \perp \nu$	singular measures, 306
$d\nu/d\mu$	Radon–Nikodým derivative, 230, 301
$\mathbb{E}^{\mathscr{G}}$	conditional expectation, 343, 346, 348
$\int u\,d\mu$	74, 82
$\int_A u\,d\mu$	$\int \mathbb{1}_A u\,d\mu$, 86
$\int u\,dx$	83
$\int u\,dT(\mu)$	$\int u \circ T\,d\mu$, 154
$\int_a^b u(x)\,dx$, $(R)\int_a^b u(x)\,dx$	Riemann integral, 102, 443

Functions and spaces

$\mathbb{1}_A$	$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$
$\operatorname{sgn}(x)$	$\mathbb{1}_{(0,\infty)}(x) - \mathbb{1}_{(-\infty,0)}(x)$
id	identity map or matrix
det	determinant (of a matrix)
$u(A)$	$\{u(x) : x \in A\}$
$u^{-1}(\mathscr{B})$	$\{u^{-1}(B) : B \in \mathscr{B}\}$, 17
u^+	$\max\{u(x), 0\}$ positive part
u^-	$-\min\{u(x), 0\}$ negative part
$\{u \in B\}$	$\{x : u(x) \in B\}$, 60
$\{u \geqslant \lambda\}$	$\{x : u(x) \geqslant \lambda\}$ etc., 60
$\operatorname{supp} u$	support $\overline{\{u \neq 0\}}$
$C(X)$	continuous functions on X
$C_b(X)$	bounded continuous functions on X
$C_\infty(X)$	continuous functions on X with $\lim\limits_{\lvert x\rvert \to \infty} u(x) = 0$
$C_c(X)$	continuous functions on X with compact support
$\mathcal{E}(\mathscr{A})$	simple functions, 63
$\mathcal{M}(\mathscr{A})$	measurable functions, 62
$\mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$	measurable functions, values in $\overline{\mathbb{R}}$, 62
$\mathcal{L}^1$	integrable functions, 82
$\mathcal{L}^1_{\overline{\mathbb{R}}}$	integrable functions, values in $\overline{\mathbb{R}}$, 82
$\mathcal{L}^p, \mathcal{L}^\infty$	116
L^p, L^∞	119
$\mathcal{L}^p_{\mathbb{C}}, L^p_{\mathbb{C}}$	423, 214
$\ell^1(\mathbb{N}), \ell^p(\mathbb{N})$	124–125
$\lVert u\rVert_p$	$\left(\int \lvert u\rvert^p\,d\mu\right)^{1/p}$, $p < \infty$, 116
$\lVert u\rVert_\infty$	$\inf\{c : \mu\{\lvert u\rvert \geqslant c\} = 0\}$, 114
$\mathfrak{M}^+_{\mathrm{r}}(X)$	regular measures on X, 437

Abbreviations

a.a., a.e.	almost all/every(where), 89
UI	uniformly integrable, 258
w.r.t.	with respect to
$\cup/\cap$-stable	stable under finite unions/intersections
$\Box$	end of proof
[✍]	indicates that a small intermediate step is required
✻	can be omitted without loss of continuity
	(in the margin) caution
(D_1)–(D_3)	Dynkin system, 32, 37
(M_0)–(M_3)	measure, 23
(OM_1)–(OM_3)	outer measure, 40, 200
$(\mathcal{O}_1)$–$(\mathcal{O}_3)$	topology, 18
(S_1)–(S_3)	semi-ring, 39
(Σ_1)–(Σ_3)	σ-algebra, 16

Prelude

The purpose of this book is to give a straightforward and yet elementary introduction to measure and integration theory that is within the grasp of second- or third- year undergraduates. Indeed, apart from interest in the subject, the only prerequisites for Chapters 1–15 are a course on rigorous ϵ–δ-analysis on the real line and basic notions of linear algebra and calculus in $\mathbb{R}^n$. The first 15 chapters form a concise introduction to Lebesgue's approach to measure and integration, which I have often taught in 10-week or 30-hour lecture courses at several universities in the UK and Germany.

Chapters 16–28 are more advanced and contain a selection of results from measure theory, probability theory and analysis. This material can be read linearly but it is also possible to select certain topics; see the dependence chart on page xvii. Although these chapters are more challenging than the first part, the prerequisites remain essentially the same and a reader who has worked through and understood Chapters 1–15 will be well prepared for all that follows. I tried to avoid topology and, when it comes in, usually an understanding of an open set and open ball (in $\mathbb{R}^n$) will suffice. From Chapter 17 onwards, I frequently use metric spaces (X, d), but you can safely think of them as $X = \mathbb{R}^n$ and $d(x, y) = |x - y|$ – or read Appendix B.

Each chapter is followed by a section of *problems*. They are not just drill exercises but contain variants of, excursions from and extensions of the material presented in the text. The proofs of the core material do not depend on any of the problems and it is as an exception that I refer to a problem in one of the proofs. Nevertheless, I do advise you to attempt as many problems as possible. The material in the *appendices* – on upper and lower limits, the Riemann integral and tools from topology – is primarily intended as back-up, for when you want to look something up.

Unlike many textbooks this is not an *introduction to integration for analysts* or a *probabilistic measure theory*. I want to reach both (future) analysts and (future) probabilists, and to provide a foundation which will be useful for both communities and for further, more specialized, studies. It goes without saying that I have had to leave out many pet choices of each discipline. On the other hand, I try to intertwine the subjects as far as possible, resulting – mostly in the latter part of the book – in the consequent use of the martingale machinery which gives 'probabilistic' proofs of 'analytic' results.

Measure and integration theory is often seen as an abstract and dry subject, which is disliked by many students. There are several reasons for this. One of them is certainly the fact that measure theory has traditionally been based on a thorough knowledge of real analysis in one and several dimensions. Many excellent textbooks are written for such an audience but today's undergraduates find it increasingly hard to follow such tracts, which are often more aptly labelled *graduate* texts. Another reason lies within the subject: measure theory has come a long way and is, in its modern purist form, stripped of its motivating roots. If, for example, one starts out with the basic definition of measures, it takes unreasonably long until one arrives at interesting examples of measures – the proof of existence and uniqueness of something as basic as Lebesgue measure already needs the full abstract machinery – and it is not easy to entertain by constantly referring to examples made up of delta functions and artificial discrete measures … I try to alleviate this by postulating the existence and properties of Lebesgue measure early on, then justifying the claims as we proceed with the abstract theory.

Technically, measure and integration theory is not more difficult than, say, complex function theory or vector calculus. Most proofs are even shorter and have a very clear structure. The one big exception, Carathéodory's extension theorem, can be safely stated without proof since an understanding of the technique is not really needed at the beginning; we will refer to the details of it only in connection with regularity questions in Chapter 16 and in Chapter 18 on Hausdorff measures. The other exceptions are the Radon–Nikodým theorem (Chapter 20) and the Riesz representation theorem (Chapter 21).

Changes in the second edition The first edition of my textbook was well received by scholars and students alike, and I would like to thank all of them for their comments and positive criticism. There are a few changes to the second edition: while the core material in Chapters 1–15 is only slightly updated, the proof of Jacobi's theorem (Chapter 16) and the material on martingales (Chapters 23 and 24), Hilbert space (Chapter 26) and conditional expectations (Chapter 27) have been re-organized and re-written. Newly added are Chapters 17–21, covering dense and measure-determining sets, Hausdorff measures and Fourier transforms as well as the classical proofs of the Radon–Nikodým and the Riesz representation theorem for L^p and C_c. I hope that these changes do not alter the character of the text and that they will make the book even more useful and accessible.

Acknowledgements I am grateful to the many people who sent me comments and corrections. My colleagues Niels Jacob, Nick Bingham, David Edmunds and Alexei Tyukov read the whole text of the first edition, and Charles Goldie and Alex Sobolev commented on large parts of the manuscript. This second edition was prepared with the help of Franziska Kühn, who did the thankless job of proofreading and suggested many improvements, and Julian Hollender, who produced most of the illustrations. Georg Berschneider and Charles Goldie read the newly added passages. Without their encouragement and help there would be more obscure passages, blunders and typos in the pages to follow. I owe a great debt to all of the students who went to my classes, challenged me to write, re-write and improve this text and who drew my attention – sometimes unbeknownst to them – to many weaknesses.

Figures 1.4, 4.1, 8.1, 8.2, 9.1, 12.2, 13.1 and 21.1 and those appearing on p. 134 and in the proofs of Theorems 3.8 and 8.8 and Lemma 13.1 are taken from my book *Maß und Integral*, De Gruyter, Berlin 2015, ISBN 978-3-11-034814-9. I am grateful for the permission of De Gruyter to use these figures here, too.

Finally, it is a pleasure to acknowledge the interest and skill of Cambridge University Press and its editors, Roger Astley and Clare Dennison, in the preparation of this book.

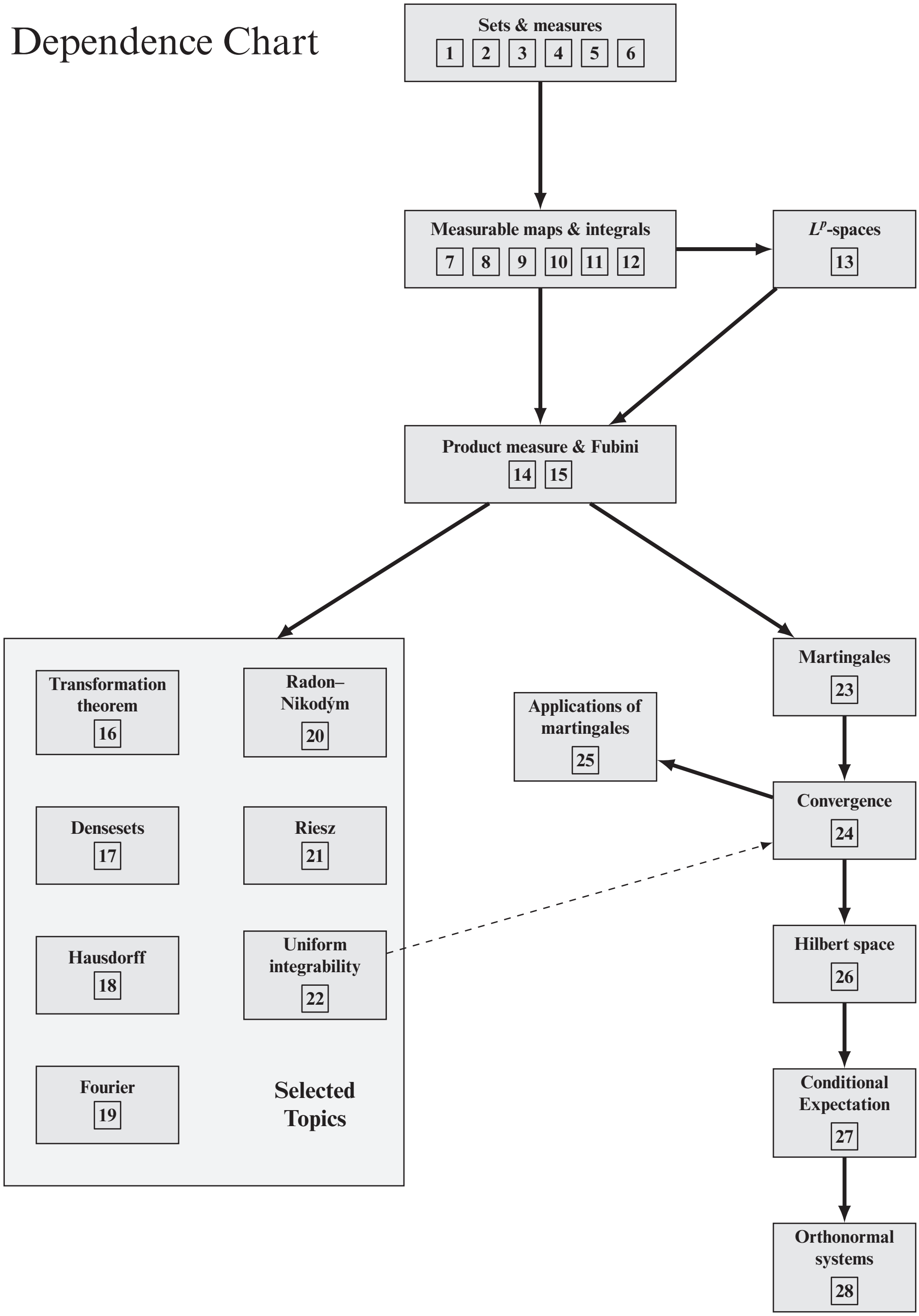
A few words on notation before getting started I try to keep unusual and special notation to a minimum. However, a few remarks are in order: $\mathbb{N}$ means the natural numbers $\{1, 2, 3, \dots\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. *Positive* or *negative* is always understood in the non-strict sense $\geqslant 0$ or $\leqslant 0$; to exclude 0, I say *strictly positive/negative*. A '$+$' as sub- or superscript refers to the positive part of a function or the positive members of a set. Finally, $a \vee b$, resp. $a \wedge b$, denote the maximum, resp. minimum, of the numbers $a, b \in \mathbb{R}$. For any other general notation there is a comprehensive list of symbols starting on page x.

In some statements I indicate alternatives using square brackets, i.e. '*if A* [*B*] ... *then P* [*Q*]' should be read as '*if A* ... *then P*' and '*if B* ... *then Q*'. The end of a proof is marked by Halmos' tombstone symbol □, and Bourbaki's dangerous bend symbol in the margin identifies a passage which requires some attention. Throughout Chapters 1–15 I have marked material which can be omitted on first reading without losing (too much) continuity by ✻.

As with every book, one cannot give all the details at every instance. On the other hand, the less experienced reader might glide over these places without even noticing that some extra effort is needed; for these readers – and, I hope, not to the annoyance of all others – I use the symbol [✍] to indicate where a short calculation on the side is appropriate.

Cross-referencing Throughout the text chapters are numbered with arabic numerals and appendices with capital letters. Theorems, definitions, examples, etc. share the same numbering sequence, e.g. Definition 4.1 is followed by Lemma 4.2 and then Corollary 4.3, and $(n.k)$ denotes formula k in Chapter n.

Dependence Chart



Dependence chart Chapters 1–15 contain core material which is needed in all later chapters. The dependence is shown by arrows, with dashed arrows indicating a minor dependence; Chapters 16–22 can be read independently and in any order.

1

Prologue

The theme of this book is the problem of how to assign a size, a content, a probability, etc. to certain sets. In everyday life this is usually pretty straightforward; we

- count: $\{a, b, c, \dots, x, y, z\}$ has 26 letters;
- take measurements: length (in one dimension), area (in two dimensions), volume (in three dimensions) or time;
- calculate: rates of radioactive decay or the odds of winning the lottery.

In each case we compare (and express the outcome) with respect to some base unit; most of the measurements just mentioned are intuitively clear. Nevertheless, let's have a closer look at areas in Fig. 1.1.

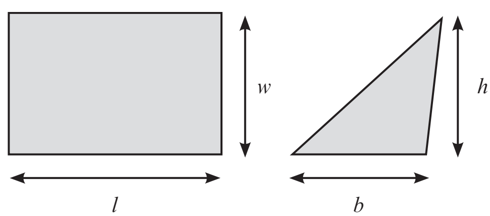


Fig. 1.1. Here area$(\square) =$ length $(\ell) \times$ width (w) and area$(\triangle) = \frac{1}{2} \times$ base $(b) \times$ height (h).

Triangles are more flexible and basic than rectangles since we can represent every rectangle, and actually any odd-shaped quadrangle, as the 'sum' of two non-overlapping triangles. In doing so we have *tacitly* assumed a few things. In Fig. 1.2 we have chosen *a particular* base line and the corresponding height arbitrarily. But the concept of *area* should not depend on such a choice and the

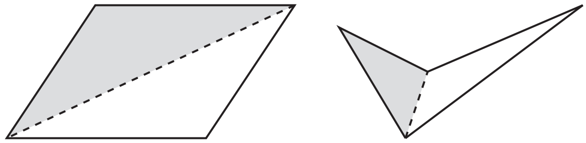


Fig. 1.2. Here area $=$ area$\big($shaded triangle$\big)$ $+$ area$\big($white triangle$\big)$.

calculation this choice entails. Independence of the area from the way we calculate it is called *well-definedness*. Plainly, we have the choices shown in Fig. 1.3. Notice that Fig. 1.3 allows us to pick the most convenient method to work out the area. In Fig. 1.2 we actually use two facts:

- the area of non-overlapping (disjoint) sets can be added, i.e.

$$\text{area}(A) = \alpha, \text{area}(B) = \beta, A \cap B = \emptyset \implies \text{area}(A \cup B) = \alpha + \beta;$$

- congruent triangles have the same area, i.e. area(◸) = area(◿).

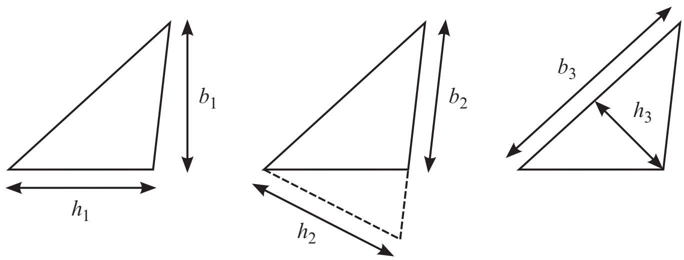


Fig. 1.3. Here area$(\triangle)$ $= \frac{1}{2} \times h_1 \times b_1 = \frac{1}{2} \times h_2 \times b_2 = \frac{1}{2} \times h_3 \times b_3$.

This shows that the least we should expect from a reasonable measure μ is that it is

$$\text{well-defined, takes values in } [0, \infty], \text{ and } \mu(\emptyset) = 0; \tag{1.1}$$

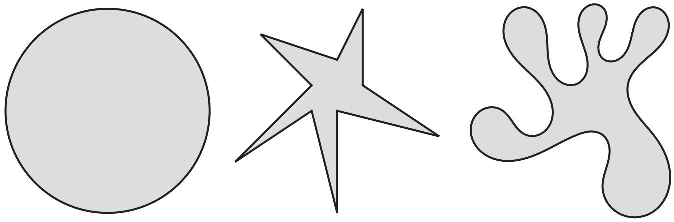
$$\text{additive, i.e. } \mu(A \cup B) = \mu(A) + \mu(B) \text{ whenever } A \cap B = \emptyset. \tag{1.2}$$

The additional property that the measure μ

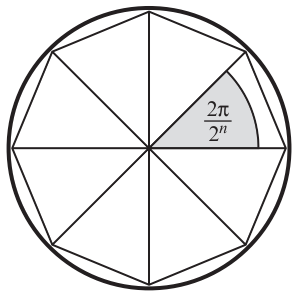
$$\text{is invariant under congruences and translations} \tag{1.3}$$

turns out to be a characteristic property of length, area and volume, i.e. of *Lebesgue measure* on $\mathbb{R}^n$.

The above rules allow us to measure arbitrarily odd-looking *polygons* using the following recipe: dissect the polygon into non-overlapping triangles and add their areas. But what about *curved* or even more complicated shapes, say,



Here is *one* possibility for the circle: inscribe a regular 2^n-gon, $n \in \mathbb{N}$, into the circle, subdivide it into congruent triangles, find the area of each of these slices and then add all 2^n pieces.



In the next step increase $n \rightsquigarrow n+1$ by doubling the number of points on the circumference and repeat the above procedure. Eventually,

$$\operatorname{area}(\bigcirc) = \lim_{n\to\infty} 2^n \times \operatorname{area}(\triangle \text{ at step } n). \tag{1.4}$$

Again, there are a few problems. Does the limit exist? Is it admissible to subdivide a set into arbitrarily many subsets? Is the procedure independent of the particular subdivision? In fact, nothing would have prevented us from paving the circle with ever smaller squares! For a reasonable notion of measure the answer to all of these questions should be *yes* and the way we pave the circle should not lead to different results, as long as our tiles are disjoint. However, finite additivity (1.2) is not enough for this and we have to use instead infinitely many pieces: σ-additivity. Thus

$$\operatorname{area}\Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big) = \sum_{n\in\mathbb{N}} \operatorname{area}(A_n).$$

The notation $\dot{\bigcup}_n A_n$ means the *disjoint union* of the sets A_n, i.e. the union where the sets A_n are pairwise disjoint: $A_n \cap A_m = \emptyset$ if $n \neq m$; a corresponding notation is used for unions of finitely many sets.

Conditions (1.1) and (1.4) lead to a notion of measure which is powerful enough to cater for all our everyday measuring needs and for much more. We will also see that a good notion of measure allows us to introduce integrals, basically starting with the naive idea that the integral of a positive function should stand for the area of the set between the graph of the function and the abscissa.

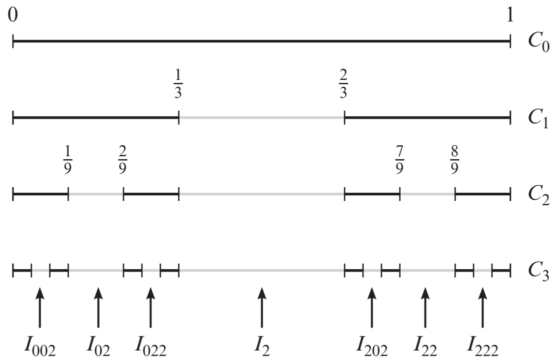


Fig. 1.4. Construction of Cantor's ternary set.

To get an idea of how far we can go with these simple principles, consider *Cantor's ternary set* on the interval $[0, 1]$, see Fig. 1.4. We obtain

1. C_1 by removing from $[0, 1]$ the middle third $I_2 = \left(\frac{1}{3}, \frac{2}{3}\right)$;
2. C_2 by removing from C_1 the middle thirds $I_{02} = \left(\frac{1}{9}, \frac{2}{9}\right)$ and $I_{22} = \left(\frac{7}{9}, \frac{8}{9}\right)$;
3. C_3 by removing from C_2 the middle thirds $I_{002}, I_{022}, I_{202}$ and I_{222};
4. ...

and $C := \bigcap_{n=1}^{\infty} C_n$ is Cantor's ternary set. By construction, C_n consists of 2^n intervals and the endpoints of these intervals will be contained in C. Thus, C is not empty. Let us calculate the 'length' ℓ of the set C. Looking at Fig. 1.4 we see that the length of C_n can be obtained by subtracting from $\ell[0, 1] = 1$ the lengths of the intervals which have been removed in the previous steps:

1. $\ell(C_1) = \ell[0, 1] - \ell(I_2) = 1 - \dfrac{1}{3}$;
2. $\ell(C_2) = \ell[0, 1] - \ell(I_2) - \ell(I_{02}) - \ell(I_{22}) = 1 - \dfrac{1}{3} - 2 \times \dfrac{1}{9}$;

....

n. $\ell(C_{n+1}) = \ell[0, 1] - 2^0 \times \dfrac{1}{3^1} - 2^1 \times \dfrac{1}{3^2} - \cdots - 2^n \times \dfrac{1}{3^{n+1}}$;

and so

$$\ell(C) = 1 - \sum_{n=0}^{\infty} \frac{2^n}{3^{n+1}} = 1 - \frac{1}{3} \sum_{n=0}^{\infty} \frac{2^n}{3^n} = 1 - \frac{1}{3} \frac{1}{1 - \frac{2}{3}} = 0.$$

Thus, the Cantor set has no length in the traditional sense, yet it is not empty.

Problems

1.1. Use (1.4) to find the area of a circle with radius r.

1.2. Where was σ-additivity used when calculating the length of the Cantor set?

1.3. Consider the following variation of Cantor's set: fix $r \in (0,1)$ and delete from $I_0 = [0,1]$ the open interval $\left(\frac{1}{2} - \frac{1}{4}r, \frac{1}{2} + \frac{1}{4}r\right)$. This defines the set I_1 consisting of two intervals, $\left[0, \frac{1}{2} - \frac{1}{4}r\right]$ and $\left[\frac{1}{2} + \frac{1}{4}r, 1\right]$. We get I_2 by removing from each of these intervals the open middles of length $r/8$ and I_3 by removing all open middles of length $r/32$. This defines recursively the sets $I_1, I_2, \dots$. Find the length of the interval I_n and of the generalized Cantor set $I := \bigcap_{n=0}^{\infty} I_n$. Is I empty?

1.4. Let $K_0 \subset \mathbb{R}^2$ be a line of length 1. We get K_1 by replacing the middle third of K_0 by the sides of an equilateral triangle. By iterating this procedure we get the curves $K_0, K_1, K_2, \dots$ (see Fig. 1.5) which defines in the limit Koch's snowflake K_∞. Find the length of K_n and K_∞.

Fig. 1.5. The first four steps in the construction of Koch's snowflake.

1.5. Let $S_0 \subset \mathbb{R}^2$ be a solid equilateral triangle. We get S_1 by removing the middle triangle whose vertices are the midpoints of the sides of S_1. By repeating this procedure with the four triangles which make up S_1 etc. we get $S_0, S_2, S_2, \dots$ (see Fig. 1.6). The Sierpiński triangle is $S_\infty := \bigcap_{n=0}^{\infty} S_n$. Find the area of S_n and S if the side-length of S_0 is $s = 1$.

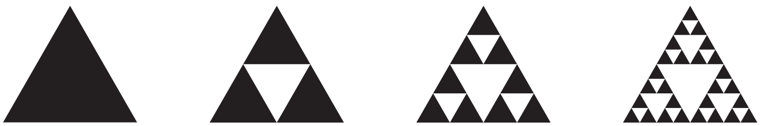


Fig. 1.6. The first four steps in the construction of Sierpiński's triangle.

2

The Pleasures of Counting

Set algebra and countability play a major rôle in measure theory. In this chapter we review briefly notation and manipulations with sets and introduce then the notion of countability. If you are not already acquainted with set algebra, you should verify all statements in this chapter and work through the exercises.

Throughout this chapter X and Y denote two arbitrary sets. For any two sets A, B (which are not necessarily subsets of a common set) we write

$$
\begin{aligned}
A \cup B &= \{x : x \in A \text{ or } x \in B \text{ or } x \in A \text{ and } B\},\\
A \cap B &= \{x : x \in A \text{ and } x \in B\},\\
A \setminus B &= \{x : x \in A \text{ and } x \notin B\};
\end{aligned}
$$

in particular, we write $A \cupdot B$ for the *disjoint union*, i.e. for $A \cup B$ if $A \cap B = \emptyset$. $A \subset B$ means that A is contained in B, including the possibility that $A = B$; to exclude the latter we write $A \subsetneq B$. If $A \subset X$, we set $A^c := X \setminus A$ for the *complement* of A (relative to X). Recall also the *distributive laws* for $A, B, C \subset X$

$$
\begin{aligned}
A \cap (B \cup C) &= (A \cap B) \cup (A \cap C),\\
A \cup (B \cap C) &= (A \cup B) \cap (A \cup C)
\end{aligned}
\tag{2.1}
$$

and *de Morgan's identities*

$$
\begin{aligned}
(A \cap B)^c &= A^c \cup B^c,\\
(A \cup B)^c &= A^c \cap B^c,
\end{aligned}
\tag{2.2}
$$

which hold also for *arbitrarily* many sets $A_i \subset X$, $i \in I$ (I stands for an arbitrary index set),

$$\begin{aligned}\Big(\bigcap_{i\in I} A_i\Big)^c &= \bigcup_{i\in I} A_i^c,\\ \Big(\bigcup_{i\in I} A_i\Big)^c &= \bigcap_{i\in I} A_i^c.\end{aligned} \tag{2.3}$$

A map $f : X \to Y$ is called

$$\begin{array}{lll} \textit{injective} \text{ (or one–one)} & \iff & f(x) = f(x') \text{ implies } x = x', \\ \textit{surjective} \text{ (or onto)} & \iff & f(X) := \{f(x) \in Y : x \in X\} = Y, \\ \textit{bijective} & \iff & f \text{ is injective and surjective.} \end{array} \tag{2.4}$$

Set operations and *direct images* under a map f are not necessarily compatible: indeed, we have, in general,

$$\begin{aligned} f(A \cup B) &= f(A) \cup f(B),\\ f(A \cap B) &\neq f(A) \cap f(B),\\ f(A \setminus B) &\neq f(A) \setminus f(B). \end{aligned} \tag{2.5}$$

Inverse images and set operations are *always* compatible. The inverse mapping f^{-1} maps subsets $C \subset Y$ into subsets of X and it is defined as

$$f^{-1}(C) := \{x \in X : f(x) \in C\} \subset X \quad \text{for all } \ C \subset Y;$$

it is, in general, multi-valued, the notation $f^{-1}(y)$ is used only if $f^{-1}(\{y\})$ has at most one element, i.e. if f is injective. For $C, C_i, D \subset Y$ one has

$$\begin{aligned} f^{-1}\Big(\bigcup_{i\in I} C_i\Big) &= \bigcup_{i\in I} f^{-1}(C_i),\\ f^{-1}\Big(\bigcap_{i\in I} C_i\Big) &= \bigcap_{i\in I} f^{-1}(C_i),\\ f^{-1}(C \setminus D) &= f^{-1}(C) \setminus f^{-1}(D). \end{aligned} \tag{2.6}$$

If we have more information about f we can, of course, say more.

Lemma 2.1 *$f : X \to Y$ is injective if, and only if, $f(A \cap B) = f(A) \cap f(B)$ for all $A, B \subset X$.*

Proof '$\Rightarrow$': Because of the inclusions $f(A \cap B) \subset f(A)$ and $f(A \cap B) \subset f(B)$, we have always $f(A \cap B) \subset f(A) \cap f(B)$. Let us check the converse inclusion '$\supset$'. If $y \in f(A) \cap f(B)$, we have $y = f(a)$ and $y = f(b)$ for some $a \in A, b \in B$.

So, $f(a)=y=f(b)$ and, by injectivity, $a=b$. This means that $a=b\in A\cap B$, hence $y\in f(A\cap B)$ and $f(A)\cap f(B)\subset f(A\cap B)$ follows.

'$\Leftarrow$': Take $x,x'\in X$ with $f(x)=f(x')$ and set $A:=\{x\}$, $B:=\{x'\}$. Then we have $\emptyset\neq f(\{x\})\cap f(\{x'\})=f(\{x\}\cap\{x'\})$ which is possible only if $\{x\}\cap\{x'\}\neq\emptyset$, i.e. if $x=x'$. This shows that f is injective. □

Lemma 2.2 *$f: X\to Y$ is injective if, and only if, $f(X\setminus A)=f(X)\setminus f(A)$ for all $A\subset X$.*

Proof [1] '$\Rightarrow$': Assume that f is injective. If $f(x)\notin f(A)$, we have $x\notin A$, and so $f(X)\setminus f(A)\subset f(X\setminus A)$ (for this inclusion we do not require injectivity). For the reverse inclusion we pick $x\in X\setminus A$ and observe that by injectivity $f(x)\neq f(x')$ for all $x'\in A$. Thus $f(x)\notin f(A)$, and we have shown that $f(x)\in f(X)\setminus f(A)$. This proves $f(X\setminus A)\subset f(X)\setminus f(A)$ as desired.

'$\Leftarrow$': Specialise $f(X\setminus A)=f(X)\setminus f(A)$ to the case $A=\{x\}$ for some $x\in X$. For any $x'\neq x$ we infer that $x'\in X\setminus\{x\}$, and so $f(x')\in f(X)\setminus f(\{x\})$, hence $f(x')\neq f(x)$. □

We can now start with the main topic of this chapter: counting.

Definition 2.3 Two sets X, Y have the same *cardinality* if there exists a bijection $f: X\to Y$. In this case we write $\#X=\#Y$.

If there is an injection $g: X\to Y$, we say that the cardinality of X is less than or equal to the cardinality of Y and write $\#X\leqslant\#Y$. If $\#X\leqslant\#Y$ but $\#X\neq\#Y$, we say that X is of strictly smaller cardinality than Y and write $\#X<\#Y$ (in this case, no injection $g: X\to Y$ can be surjective).

That Definition 2.3 is indeed *counting* becomes clear if we choose $Y=\mathbb{N}$ since in this case $\#X=\#\mathbb{N}$ or $\#X\leqslant\#\mathbb{N}$ just means that we can label each $x\in X$ with a unique tag from the set $\{1,2,3,\dots\}$, i.e. we are numbering X. This particular example is, in fact, of central importance.

Definition 2.4 A set X is *countable* if $\#X\leqslant\#\mathbb{N}$. If $\#\mathbb{N}<\#X$, the set X is said to be *uncountable*. The cardinality of $\mathbb{N}$ is called $\aleph_0$, *aleph null*.

Plainly, Definition 2.4 requires that we can find for every countable set some *enumeration* $X=(x_1,x_2,x_3,\dots)$ which may or may not be finite (and which may contain any x_i more than once).

Caution Sometimes *countable* is used only to indicate $\#X=\#\mathbb{N}$, while sets where $\#X\leqslant\#\mathbb{N}$ are called *at most countable* or *finite or countable*. This has the effect that a countable set is always infinite. We do not adopt this convention.

[1] I owe this short argument to Charles Goldie.

The following examples show that (countable) sets with infinitely many elements can behave strangely.

Example 2.5 (i) Finite sets are countable: $\{a, b, \dots, z\} \to \{1, 2, \dots, 26\}$, where $a \leftrightarrow 1, \dots, z \leftrightarrow 26$, is bijective and $\{1, 2, 3, \dots, 26\} \to \mathbb{N}$ is clearly an injection. Thus

$$\#\{a, b, c, \dots, z\} = \#\{1, 2, 3, \dots, 26\} \leqslant \#\mathbb{N}.$$

(ii) The even numbers are countable. This follows from the fact that the map

$$f \colon \{2, 4, 6, \dots, 2i, \dots\} \to \mathbb{N}, \quad i \mapsto \frac{i}{2},$$

is an injection and even a bijection. [✍] This means that there are 'as many' even numbers as there are natural numbers.

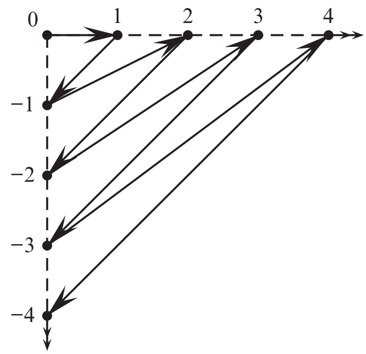
(iii) The set of integers

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$$

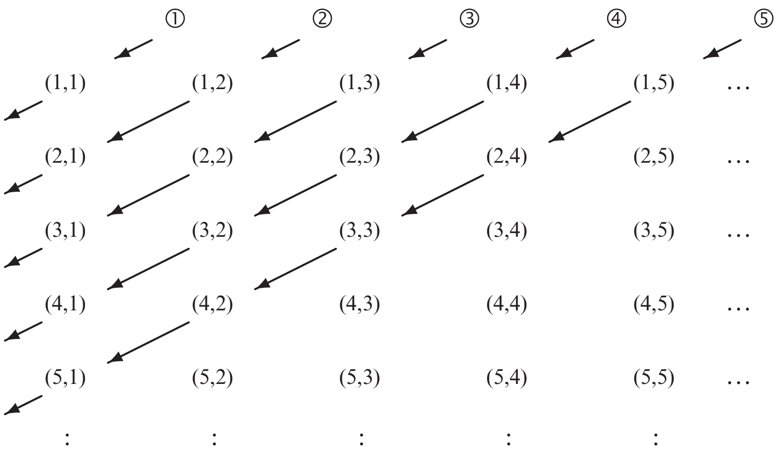
is countable. A possible counting scheme is shown on the right or, more formally,

$$g : i \in \mathbb{Z} \mapsto \begin{cases} 2i & \text{if } i > 0, \\ 2|i| + 1 & \text{if } i \leqslant 0, \end{cases}$$

hence $\#\mathbb{Z} \leqslant \#\mathbb{N}$. [✍]



(iv) The Cartesian product $\mathbb{N} \times \mathbb{N} := \{(i, k) : i, k \in \mathbb{N}\}$ is countable. To see this, arrange the pairs (i, k) in an array and count along the diagonals:



Notice that each line contains only finitely many elements, so that each diagonal can be dealt with in finitely many steps. The map for the above counting scheme is given by

$$h:(i,k)\mapsto\frac{(i+k)(i+k-1)}{2}-k+1\in\mathbb{N},\qquad(i,k)\in\mathbb{N}\times\mathbb{N}.\tag{2.7}$$

(v) The rational numbers $\mathbb{Q}$ are countable: set $Q_\pm:=\{q\in\mathbb{Q}:\pm q>0\}$. Every element $\frac{m}{n}\in Q_+$ can be identified with at least one pair $(m,n)\in\mathbb{N}\times\mathbb{N}$, so that

$$Q_+\subset\Big\{\underbrace{\tfrac11,}_{①}\ \underbrace{\tfrac12,\tfrac21,}_{②}\ \underbrace{\tfrac13,\tfrac22,\tfrac31,}_{③}\ \underbrace{\tfrac14,\tfrac23,\tfrac32,\tfrac41,}_{④}\ldots\Big\};$$

in the set $\{\ldots\}$ on the right we distinguish between cancelled and uncancelled forms of a rational, i.e. $\frac{6}{18},\frac13,\frac26,\frac39,\ldots$, etc. are counted whenever they appear. The numbers ⓘ refer to the corresponding diagonals in the counting scheme in part (iv). This shows that we can find injections $Q_+\xrightarrow{\imath}\{\ldots\}\xrightarrow{\imath}\mathbb{N}\times\mathbb{N}$; the set $\mathbb{N}\times\mathbb{N}$ is countable, thus Q_+ is countable [✍] and so is Q_-. Finally,

$$\mathbb{Q}=Q_-\uplus\{0\}\uplus Q_+=\{r_1,r_2,r_3,\ldots\}\uplus\{0\}\uplus\{q_1,q_2,q_3,\ldots\}$$

and $p_1:=0,\ p_{2i}:=q_i,\ p_{2i+1}:=r_i$ gives an enumeration $(p_1,p_2,p_3,\ \ldots)$ of $\mathbb{Q}$.

Theorem 2.6 *Let $A_1,A_2,A_3,\ldots$ be countably many countable sets. Then their union $A=\bigcup_{i\in\mathbb{N}}A_i$ is countable, i.e. countable unions of countable sets are countable.*

Proof Since each A_i is countable we can find an enumeration

$$A_i=\big(a_{i,1},a_{i,2},\ldots,a_{i,k},\ldots\big)$$

(if A_i is a finite set, we repeat the last element of the list infinitely often), so that

$$A=\bigcup_{i\in\mathbb{N}}A_i=\big(a_{i,k}:(i,k)\in\mathbb{N}\times\mathbb{N}\big).$$

Using Example 2.5(iv) we can relabel $\mathbb{N}\times\mathbb{N}$ by $\mathbb{N}$ and (after deleting all duplicates) we have found an enumeration. □

It is not hard to see that for cardinalities '$\leqslant$' is *reflexive* ($\#A\leqslant\#A$) and *transitive* ($\#A\leqslant\#B,\ \#B\leqslant\#C\implies\#A\leqslant\#C$). *Antisymmetry*, which makes '$\leqslant$' into a partial order relation, is less obvious. The proof of the following important result is somewhat technical and can be left out at first reading.

***Theorem 2.7** (Cantor–Bernstein) *Let X, Y be two sets. If both $\#X\leqslant\#Y$ and $\#Y\leqslant\#X$, then $\#X=\#Y$.*

Proof By assumption,

$$\#X \leqslant \# Y \iff \text{there exists an injection} \quad f\colon X \to Y;$$
$$\# Y \leqslant \#X \iff \text{there exists an injection} \quad g\colon Y \to X.$$

In order to prove $\#X = \# Y$ we have to construct a bijection $h\colon X \to Y$.

Step 1. Without loss of generality we may assume that $Y \subset X$. Indeed, since $g\colon Y \to g(Y)$ is a bijection, we know that $\# Y = \#g(Y)$ and it suffices to show $\#g(Y) = \#X$. As $g(Y) \subset X$ we can simplify things and identify $g(Y)$ with Y, i.e. assume that $g = \mathrm{id}$ or, equivalently, $Y \subset X$.

Step 2. Let $Y \subset X$ and $g = \mathrm{id}$. Recursively we define

$$X_0 := X, \dots, X_{i+1} := f(X_i); \qquad Y_0 := Y, \dots, Y_{i+1} := f(Y_i).$$

As usual, we write $f^i := \underbrace{f \circ f \circ \cdots \circ f}_{i \text{ times}}$ and $f^0 := \mathrm{id}$. Then

$$\begin{array}{ccccccc} f^{i+1}(X) & = & f^i(f(X)) & \overset{f(X)\subset Y}{\subset} & f^i(Y) & \overset{Y\subset X}{\subset} & f^i(X) \\ \| & & & & \| & & \| \\ X_{i+1} & & \subset & & Y_i & \subset & X_i \end{array}$$

and we can define a map $h\colon X \to Y$ by

$$h(x) := \begin{cases} f(x) & \text{if } x \in X_i \setminus Y_i \text{ for some } i \in \mathbb{N}_0, \\ x & \text{if } x \notin \bigcup_{i\in\mathbb{N}_0} X_i \setminus Y_i. \end{cases}$$

Step 3. The map h is surjective: $h(X) = Y$. Indeed, we have by definition

$$\begin{aligned} h(X) &= \bigcup_{i\in\mathbb{N}_0} f(X_i \setminus Y_i) \cup \Big(\bigcup_{i\in\mathbb{N}_0} X_i \setminus Y_i\Big)^c \\ &\overset{2.2}{=} \bigcup_{i\in\mathbb{N}_0} \big(f(X_i) \setminus f(Y_i)\big) \cup \Big(X \setminus Y \cup \bigcup_{i\in\mathbb{N}} X_i \setminus Y_i\Big)^c \\ &= \underbrace{\bigcup_{i\in\mathbb{N}_0} \big(X_{i+1} \setminus Y_{i+1}\big)}_{=:A} \cup \Big[\big(X \setminus Y\big)^c \cap \Big(\bigcup_{i\in\mathbb{N}_0} X_{i+1} \setminus Y_{i+1}\Big)^c\Big] \\ &= A \cup [(X^c \cup Y) \cap A^c] = A \cup [Y \cap A^c] = Y \cap X = Y, \end{aligned}$$

where we use that $A = \bigcup_{i\in\mathbb{N}_0} X_{i+1} \setminus Y_{i+1} \subset \bigcup_{i\in\mathbb{N}_0} X_{i+1} \subset X_1 = f(X) \subset Y$.

Step 4. The map h is injective. To see this, let $x, x' \in X$ and $h(x) = h(x')$. We have four possibilities.

(a) $x, x' \in X_i \setminus Y_i$ for some $i \in \mathbb{N}_0$. Then we have $f(x) = h(x) = h(x') = f(x')$ so that $x = x'$ since f is injective.
(b) $x, x' \notin \bigcup_{i\in\mathbb{N}_0} X_i \setminus Y_i$. Then $x = h(x) = h(x') = x'$.
(c) $x \in X_i \setminus Y_i$ for some $i \in \mathbb{N}_0$ and $x' \notin X_k \setminus Y_k$ for all $k \in \mathbb{N}_0$. As $f(x) = h(x) = h(x') = x'$ we see

$$x' = f(x) \in f(X_i \setminus Y_i) \overset{2.2}{=} f(X_i) \setminus f(Y_i) = X_{i+1} \setminus Y_{i+1},$$

which is impossible, i.e. (c) cannot occur.
(d) $x' \in X_i \setminus Y_i$ for some $i \in \mathbb{N}_0$ and $x \notin X_k \setminus Y_k$ for all $k \in \mathbb{N}_0$. This is analogous to (c). □

Theorem 2.7 says that $\#X < \#Y$ and $\#Y < \#X$ cannot occur at the same time; it does not claim that we can compare the cardinality of any two sets X and Y, i.e. that '$\leqslant$' is a linear ordering. This is indeed true but its proof requires the *axiom of choice*, see Hewitt–Stromberg [22, p. 20].

Not all sets are countable. The following proof goes back to Georg Cantor and is called *Cantor's diagonal method*.

Theorem 2.8 *The interval* $(0, 1)$ *is uncountable; its cardinality* $\mathfrak{c} := \#(0, 1)$ *is called the continuum.*

Proof We can write each $x \in (0, 1)$ as a decimal fraction, i.e. $x = 0.y_1y_2y_3\ldots$ with $y_i \in \{0, 1, \ldots, 9\}$. If x has a finite decimal representation, say $x = 0.y_1y_2y_3\ldots y_n$, $y_n \neq 0$, we replace the last digit y_n by $y_n - 1$ and fill it up with trailing 9s. For example, $0.24 = 0.23\overline{99}\ldots$. This yields a *unique* representation of x by an *infinite* decimal expansion.

Assume that $(0, 1)$ were countable and let $\{x_1, x_2, \ldots\}$ be an enumeration (containing no element more than once!). Then we can write

$$\begin{aligned} x_1 &= 0.\boldsymbol{a_{1,1}}a_{1,2}a_{1,3}a_{1,4}\ldots, \\ x_2 &= 0.a_{2,1}\boldsymbol{a_{2,2}}a_{2,3}a_{2,4}\ldots, \\ x_3 &= 0.a_{3,1}a_{3,2}\boldsymbol{a_{3,3}}a_{3,4}\ldots, \\ x_4 &= 0.a_{4,1}a_{4,2}a_{4,3}\boldsymbol{a_{4,4}}\ldots, \\ &\vdots \quad \vdots \end{aligned} \tag{2.8}$$

and construct a new number $x := 0.y_1y_2y_3\ldots \in (0, 1)$ with digits

$$y_i := \begin{cases} 1 & \text{if } a_{i,i} = 5, \\ 5 & \text{if } a_{i,i} \neq 5. \end{cases} \tag{2.9}$$

By construction, $x \neq x_i$ for any x_i from the list (2.8): x and x_i differ at the ith decimal. But then we have found a number $x \in (0,1)$ which is not contained in our supposedly complete enumeration of $(0,1)$ and we get a contradiction. □

By $(0,1)^{\mathbb{N}}$ we denote the set of all sequences $(x_i)_{i\in\mathbb{N}}$ where $x_i \in (0,1)$.

Theorem 2.9 *We have* $\#(0,1)^{\mathbb{N}} = \mathfrak{c}$.

Proof We have to assign to every sequence $(x_i)_{i\in\mathbb{N}} \subset (0,1)$ a unique number $x \in (0,1)$ – and vice versa. For this we write, as in the proof of Theorem 2.8, each x_i as a unique infinite decimal fraction

$$x_i = 0.a_{i,1}a_{i,2}a_{i,3}a_{i,4}\ldots, \quad i \in \mathbb{N},$$

and we organize the array $(a_{i,k})_{i,k\in\mathbb{N}}$ into one sequence with the help of the counting scheme of Example 2.5(iv):

$$x := 0.\,\underbrace{a_{1,1}}_{\textcircled{1}}\,\underbrace{a_{1,2}\,a_{2,1}}_{\textcircled{2}}\,\underbrace{a_{1,3}\,a_{2,2}\,a_{3,1}}_{\textcircled{3}}\,\underbrace{a_{1,4}\,a_{2,3}\,a_{3,2}\,a_{4,1}}_{\textcircled{4}}\ldots$$

(the numbers $\textcircled{i}$ refer to the corresponding diagonals in the counting scheme of Example 2.5(iv)). Since the counting scheme was bijective, this procedure is reversible, i.e. we can start with the decimal expansion of $x \in (0,1)$ and get a unique sequence of x_is. We have thus found an injection from $(0,1)^{\mathbb{N}}$ to $(0,1)$, hence $\#(0,1)^{\mathbb{N}} \leqslant \#(0,1)$. On the other hand, $\#(0,1) \leqslant \#(0,1)^{\mathbb{N}}$ is obvious, [✍] and an application of the Cantor–Bernstein theorem (Theorem 2.7) finishes the proof. □

We write $\mathscr{P}(X)$ for the *power set* $\{A : A \subset X\}$ which is the family of all subsets of a given set X. For finite sets it is clear that the power set is of strictly larger cardinality than X. This is still true for infinite sets.

Theorem 2.10 *For any set X we have* $\#X < \#\mathscr{P}(X)$.

Proof We have to show that no injection $\Phi : X \to \mathscr{P}(X)$ can be surjective. Fix such an injection and define

$$B := \{x \in X : x \notin \Phi(x)\}$$

(mind: $\Phi(x)$ is a set!). Clearly $B \in \mathscr{P}(X)$. If Φ were surjective, we would have $B = \Phi(z)$ for some element $z \in X$. Then, however,

$$z \in B \quad \overset{\text{def}}{\iff} \quad z \notin \Phi(z) \iff z \notin B \quad (\text{since } \Phi(z) = B),$$

which is impossible. Thus Φ cannot be surjective. □

Problems

2.1. Let $A, B, C \subset X$ be sets. Show that

(i) $A \setminus B = A \cap B^c$;
(ii) $(A \setminus B) \setminus C = A \setminus (B \cup C)$;
(iii) $A \setminus (B \setminus C) = (A \setminus B) \cup (A \cap C)$;
(iv) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$;
(v) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$;
(vi) $(A \cup B) \setminus C = (A \setminus C) \cup (B \setminus C)$.

2.2. Let $A, B, C \subset X$. The *symmetric difference* of A and B is $A \vartriangle B := (A \setminus B) \cup (B \setminus A)$. Verify that

$$(A \cup B \cup C) \setminus (A \cap B \cap C) = (A \vartriangle B) \cup (B \vartriangle C).$$

2.3. Prove de Morgan's identities (2.2) and (2.3).

2.4. (i) Find examples showing that $f(A \cap B) \neq f(A) \cap f(B)$ and $f(A \setminus B) \neq f(A) \setminus f(B)$. In both relations one inclusion '$\subset$' or '$\supset$' is always true. Which one?
(ii) Prove (2.6).

2.5. The *indicator function* of a set $A \subset X$ is defined by

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

Check that for $A, B, A_i \subset X$, $i \in I$ (arbitrary index set) the following equalities hold:

(i) $\mathbb{1}_{A \cap B} = \mathbb{1}_A \mathbb{1}_B$; (ii) $\mathbb{1}_{A \cup B} = \min\{\mathbb{1}_A + \mathbb{1}_B, 1\}$;
(iii) $\mathbb{1}_{A \setminus B} = \mathbb{1}_A - \mathbb{1}_{A \cap B}$; (iv) $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A \cap B}$;
(v) $\mathbb{1}_{A \cup B} = \max\{\mathbb{1}_A, \mathbb{1}_B\}$; (vi) $\mathbb{1}_{A \cap B} = \min\{\mathbb{1}_A, \mathbb{1}_B\}$;
(vii) $\mathbb{1}_{\bigcup_{i \in I} A_i} = \sup_{i \in I} \mathbb{1}_{A_i}$; (viii) $\mathbb{1}_{\bigcap_{i \in I} A_i} = \inf_{i \in I} \mathbb{1}_{A_i}$.

2.6. Let $A, B, C \subset X$ and denote by $A \vartriangle B$ the symmetric difference as in Problem 2.2. Show that

(i) $\mathbb{1}_{A \vartriangle B} = \mathbb{1}_A + \mathbb{1}_B - 2\mathbb{1}_A \mathbb{1}_B = |\mathbb{1}_A - \mathbb{1}_B| = \mathbb{1}_A + \mathbb{1}_B \mod 2$;
(ii) $A \vartriangle (B \vartriangle C) = (A \vartriangle B) \vartriangle C$;
(iii) $\mathscr{P}(X)$ is a commutative ring (in the usual algebraists' sense) with 'addition' $\vartriangle$ and 'multiplication' $\cap$.

[Hint: use indicator functions for (ii) and (iii).]

2.7. Let $f: X \to Y$ be a map, $A \subset X$ and $B \subset Y$. Show that, in general,

$$f \circ f^{-1}(B) \subsetneq B \quad \text{and} \quad f^{-1} \circ f(A) \supsetneq A.$$

When does '$=$' hold in these relations? Provide an example showing that the above inclusions are strict.

2.8. Let f and g be two injective maps. Show that $f \circ g$, if it exists, is injective.

2.9. Show that the following sets have the same cardinality as $\mathbb{N}$: $\{m \in \mathbb{N} : m \text{ is odd}\}$, $\mathbb{N} \times \mathbb{Z}$, $\mathbb{Q}^m$ $(m \in \mathbb{N})$, $\bigcup_{m \in \mathbb{N}} \mathbb{Q}^m$.

2.10. Use Theorem 2.7 to show that $\#\mathbb{N} \times \mathbb{N} = \#\mathbb{N}$.
[Hint: $\#\mathbb{N} = \#\mathbb{N} \times \{1\}$ and $\mathbb{N} \times \{1\} \subset \mathbb{N} \times \mathbb{N}$.]

2.11. Show that if $E \subset F$ we have $\#E \leqslant \#F$. In particular, subsets of countable sets are again countable.

2.12. Show that $\{0,1\}^{\mathbb{N}} = \{\text{all infinite sequences consisting of 0 and 1}\}$ is uncountable.
[Hint: use the diagonal method.]

2.13. Show that the set $\mathbb{R}$ is uncountable and that $\#(0,1) = \#\mathbb{R}$.
[Hint: find a bijection $f\colon (0,1) \to \mathbb{R}$.]

2.14. Let $(A_i)_{i\in\mathbb{N}}$ be a sequence of sets of cardinality $\mathfrak{c}$. Show that $\#\bigcup_{i\in\mathbb{N}} A_i = \mathfrak{c}$.
[Hint: map A_i bijectively onto $(i-1,i)$ and use that $(0,1) \subset \bigcup_{i=1}^{\infty}(i-1,i) \subset \mathbb{R}$.]

2.15. Adapt the proof of Theorem 2.8 to show that $\#\{1,2\}^{\mathbb{N}} \leqslant \#(0,1) \leqslant \#\{0,1\}^{\mathbb{N}}$ and conclude that $\#(0,1) = \#\{0,1\}^{\mathbb{N}}$.
Remark. This is the reason for writing $\mathfrak{c} = 2^{\aleph_0}$.
[Hint: interpret $\{0,1\}^{\mathbb{N}}$ as base-2 expansions of all numbers in $(0,1)$ while $\{1,2\}^{\mathbb{N}}$ are all infinite base-3 expansions lacking the digit 0.]

2.16. Extend Problem 2.15 to deduce $\#\{0,1,2,\dots,n\}^{\mathbb{N}} = \#(0,1)$ for all $n \in \mathbb{N}$.

2.17. Mimic the proof of Theorem 2.9 to show that $\#(0,1)^2 = \mathfrak{c}$. Use the fact that $\#\mathbb{R} = \#(0,1)$ to conclude that $\#\mathbb{R}^2 = \mathfrak{c}$.

2.18. Show that the set of all infinite sequences of natural numbers $\mathbb{N}^{\mathbb{N}}$ has cardinality $\mathfrak{c}$.
[Hint: use that $\#\{0,1\}^{\mathbb{N}} = \#\{1,2\}^{\mathbb{N}}$, $\{1,2\}^{\mathbb{N}} \subset \mathbb{N}^{\mathbb{N}} \subset \mathbb{R}^{\mathbb{N}}$ and $\#\mathbb{R}^{\mathbb{N}} = \#(0,1)^{\mathbb{N}}$.]

2.19. Let $\mathscr{F} := \{F \subset \mathbb{N} : \#F < \infty\}$. Show that $\#\mathscr{F} = \#\mathbb{N}$.
[Hint: embed $\mathscr{F}$ into $\bigcup_{k\in\mathbb{N}} \mathbb{N}^k$ *or* show that $F \mapsto \sum_{j\in F} 2^j$ is a bijection between $\mathscr{F}$ and $\mathbb{N}$.]

2.20. Show – not using Theorem 2.10 – that $\#\mathscr{P}(\mathbb{N}) > \#\mathbb{N}$. Conclude that there are more than countably many maps $f\colon \mathbb{N} \to \mathbb{N}$.
[Hint: use the diagonal method.]

2.21. If $A \subset \mathbb{N}$ we can identify the indicator function $\mathbb{1}_A : \mathbb{N} \to \{0,1\}$ with the 0–1-sequence $\left(\mathbb{1}_A(i)\right)_{i\in\mathbb{N}}$, i.e. $\mathbb{1}_A \in \{0,1\}^{\mathbb{N}}$. Show that the map $\mathscr{P}(\mathbb{N}) \ni A \mapsto \mathbb{1}_A \in \{0,1\}^{\mathbb{N}}$ is a bijection and conclude that $\#\mathscr{P}(\mathbb{N}) = \mathfrak{c}$.

2.22. Show that for $A_n^0, A_n^1 \subset X$, $n \in \mathbb{N}$, we have

$$\bigcup_{n\in\mathbb{N}} (A_n^0 \cap A_n^1) = \bigcap_{i=(i(k))_{k\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}} \ \bigcup_{k\in\mathbb{N}} A_k^{i(k)}.$$

3

σ-Algebras

We have seen in the prologue that a reasonable measure should be able to deal with disjoint *countable* partitions of sets. Therefore, a measure function must be defined on a system of sets which is stable whenever we repeat any of the basic set operations – $\cup, \cap, {}^c$ – countably many times.

Definition 3.1 A *σ-algebra* $\mathscr{A}$ on a set X is a family of subsets of X with the following properties:

$$X \in \mathscr{A}, \tag{Σ_1}$$

$$A \in \mathscr{A} \implies A^c \in \mathscr{A}, \tag{Σ_2}$$

$$(A_n)_{n\in\mathbb{N}} \subset \mathscr{A} \implies \bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}. \tag{Σ_3}$$

A set $A \in \mathscr{A}$ is said to be *measurable* or *$\mathscr{A}$-measurable*.

Properties 3.2 (of a σ-algebra) (i) $\emptyset \in \mathscr{A}$.

Indeed: $\emptyset = X^c \in \mathscr{A}$ by (Σ_1), (Σ_2).

(ii) $A, B \in \mathscr{A} \implies A \cup B \in \mathscr{A}$.

Indeed: if $A_1 = A,\ A_2 = B,\ A_3 = A_4 = \cdots = \emptyset$, then $A \cup B = \bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}$ by (Σ_3).

(iii) $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A} \implies \bigcap_{n\in\mathbb{N}} A_n \in \mathscr{A}$.

Indeed: if $A_n \in \mathscr{A}$, then $A_n^c \in \mathscr{A}$ by (Σ_2), hence $\bigcup_{n\in\mathbb{N}} A_n^c \in \mathscr{A}$ by (Σ_3) and, again by (Σ_2), $\bigcap_{n\in\mathbb{N}} A_n = \big(\bigcup_{n\in\mathbb{N}} A_n^c\big)^c \in \mathscr{A}$.

Example 3.3 (i) $\mathscr{P}(X)$ is a σ-algebra (the maximal σ-algebra in X).

(ii) $\{\emptyset, X\}$ is a σ-algebra (the minimal σ-algebra in X).

(iii) $\{\emptyset, B, B^c, X\}$, $B \subset X$, is a σ-algebra.

(iv) $\{\emptyset, B, X\}$ is **no** σ-algebra (unless $B = \emptyset$ or $B = X$).

(v) $\mathscr{A} := \{A \subset X : \#A \leqslant \#\mathbb{N} \text{ or } \#A^c \leqslant \#\mathbb{N}\}$ is a σ-algebra.
Proof: Let us verify (Σ_1)–(Σ_3).
(Σ_1): $X^c = \emptyset$, which is certainly countable.
(Σ_2): if $A \in \mathscr{A}$, either A or A^c is by definition countable, so $A^c \in \mathscr{A}$.
(Σ_3): if $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$, then the following two cases can occur.

- All A_n are countable. Then $A = \bigcup_{n\in\mathbb{N}} A_n$ is a countable union of countable sets, which is, by Theorem 2.6, itself countable.
- At least one A_{n_0} is uncountable. Then $A_{n_0}^c$ must be countable, so that

$$\Big(\bigcup_{n\in\mathbb{N}} A_n\Big)^c = \bigcap_{n\in\mathbb{N}} A_n^c \subset A_{n_0}^c.$$

Hence $\left(\bigcup_{n\in\mathbb{N}} A_n\right)^c$ is countable (Problem 2.11) and so $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}$.

(vi) **(Trace σ-algebra)** Let $E \subset X$ be any set and let $\mathscr{A}$ be some σ-algebra in X. Then

$$\mathscr{A}_E := E \cap \mathscr{A} := \{E \cap A : A \in \mathscr{A}\} \tag{3.1}$$

is a σ-algebra in E.

(vii) **(Pre-image σ-algebra)** Let $f: X \to X'$ be a map and let $\mathscr{A}'$ be a σ-algebra in X'. Then

$$\mathscr{A} := f^{-1}(\mathscr{A}') := \{f^{-1}(A') : A' \in \mathscr{A}'\} \tag{3.2}$$

is a σ-algebra in X.

Theorem 3.4 (and Definition) (i) *The intersection $\bigcap_{i\in I} \mathscr{A}_i$ of arbitrarily many σ-algebras $\mathscr{A}_i$ in X is again a σ-algebra in X.*

(ii) *For every system of sets $\mathscr{G} \subset \mathscr{P}(X)$ there exists a smallest (also: minimal, coarsest) σ-algebra containing $\mathscr{G}$. This is the σ-algebra generated by $\mathscr{G}$, denoted by $\sigma(\mathscr{G})$, and $\mathscr{G}$ is called its* generator.

Proof (i) We check (Σ_1)–(Σ_3). (Σ_1): since $X \in \mathscr{A}_i$ for all $i \in I$, $X \in \bigcap_i \mathscr{A}_i$.

(Σ_2): if $A \in \bigcap_i \mathscr{A}_i$, then $A^c \in \mathscr{A}_i$ for all $i \in I$, so $A^c \in \bigcap_i \mathscr{A}_i$.

(Σ_3): let $(A_n)_{n\in\mathbb{N}} \subset \bigcap_i \mathscr{A}_i$. Then $A_n \in \mathscr{A}_i$ for all $n \in \mathbb{N}$ and all $i \in I$, hence $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}_i$ for each $i \in I$ and so $\bigcup_{n\in\mathbb{N}} A_n \in \bigcap_{i\in I} \mathscr{A}_i$.

(ii) Consider the family

$$\mathscr{A} := \bigcap_{\substack{\mathscr{F}\ \sigma\text{-alg.} \\ \mathscr{F} \supset \mathscr{G}}} \mathscr{F}.$$

Since $\mathscr{G} \subset \mathscr{P}(X)$ and since $\mathscr{P}(X)$ is a σ-algebra, the above intersection is non-void. This means that the definition of $\mathscr{A}$ makes sense and yields, by part (i), a σ-algebra containing $\mathscr{G}$. If $\mathscr{A}'$ is a further σ-algebra with $\mathscr{A}' \supset \mathscr{G}$, then $\mathscr{A}'$ would be included in the intersection used for the definition of $\mathscr{A}$, hence we have $\mathscr{A} \subset \mathscr{A}'$. In this sense, $\mathscr{A}$ is the smallest σ-algebra containing $\mathscr{G}$. □

Remark 3.5 (i) If $\mathscr{G}$ is a σ-algebra, then $\mathscr{G} = \sigma(\mathscr{G})$.
(ii) For $A \subset X$ we have $\sigma(\{A\}) = \{\emptyset, A, A^c, X\}$.
(iii) If $\mathscr{F} \subset \mathscr{G} \subset \mathscr{A}$, then $\sigma(\mathscr{F}) \subset \sigma(\mathscr{G}) \subset \sigma(\mathscr{A}) \overset{3.5(i)}{=} \mathscr{A}$.

On the Euclidean space $\mathbb{R}^n$ there is a canonical σ-algebra, which is generated by the open sets. Recall that

$$U \subset \mathbb{R}^n \quad \text{is } \textit{open} \iff \forall x \in U \; \exists \epsilon > 0 : B_\epsilon(x) \subset U,$$

where $B_\epsilon(x) := \{y \in \mathbb{R}^n : |x - y| < \epsilon\}$ is the open ball with centre x and radiusϵ. A set is *closed* if its complement is open. The system of open sets in $X = \mathbb{R}^n$, $\mathscr{O} = \mathscr{O}_X$, has the following properties:

$$\emptyset, X \in \mathscr{O}, \tag{$\mathscr{O}_1$}$$

$$U, V \in \mathscr{O} \implies U \cap V \in \mathscr{O}, \tag{$\mathscr{O}_2$}$$

$$U_i \in \mathscr{O}, \; i \in I \,(\text{arbitrary}) \implies \bigcup_{i \in I} U_i \in \mathscr{O}. \tag{$\mathscr{O}_3$}$$

Note, however, that countable or arbitrary intersections of open sets need not be open. [✍] A family of subsets $\mathscr{O}$ of a general space X satisfying the conditions ($\mathscr{O}_1$)–($\mathscr{O}_3$) is called a *topology*, and the pair $(X, \mathscr{O})$ is called a *topological space*; in analogy to $\mathbb{R}^n$, $U \in \mathscr{O}$ is said to be *open*, while *closed* sets are exactly the complements of open sets.

Definition 3.6 The σ-algebra $\sigma(\mathscr{O})$ generated by the open sets $\mathscr{O} = \mathscr{O}_{\mathbb{R}^n}$ of $\mathbb{R}^n$ is called *Borel σ-algebra*, and its members are called *Borel sets* or *Borel measurable sets*. We write $\mathscr{B}(\mathbb{R}^n)$ for the Borel sets in $\mathbb{R}^n$.

The Borel sets are fundamental for the study of measures on $\mathbb{R}^n$. Since the Borel σ-algebra depends on the topology of $\mathbb{R}^n$, $\mathscr{B}(\mathbb{R}^n)$ is often also called the *topological σ-algebra*.

Theorem 3.7 *Denote by $\mathscr{O}, \mathscr{C}$ and $\mathscr{K}$ the families of open, closed and compact*[1] *sets in $\mathbb{R}^n$. Then*

$$\mathscr{B}(\mathbb{R}^n) = \sigma(\mathscr{O}) = \sigma(\mathscr{C}) = \sigma(\mathscr{K}).$$

[1] That is, closed and bounded.

Proof Since compact sets are closed, we have $\mathscr{K} \subset \mathscr{C}$ and, by Remark 3.5(iii), $\sigma(\mathscr{K}) \subset \sigma(\mathscr{C})$. On the other hand, if $C \in \mathscr{C}$, then $C_k := C \cap \overline{B_k(0)}$ [2] is closed and bounded, hence $C_k \in \mathscr{K}$. By construction $C = \bigcup_{k\in\mathbb{N}} C_k$, thus $\mathscr{C} \subset \sigma(\mathscr{K})$ and also $\sigma(\mathscr{C}) \subset \sigma(\mathscr{K})$.

Since $(\mathscr{O})^c := \{U^c : U \in \mathscr{O}\} = \mathscr{C}$ (and $(\mathscr{C})^c = \mathscr{O}$) we have $\mathscr{C} = (\mathscr{O})^c \subset \sigma(\mathscr{O})$, hence $\sigma(\mathscr{C}) \subset \sigma(\mathscr{O})$ and the converse inclusion is similar. $\square$

The Borel σ-algebra $\mathscr{B}(\mathbb{R}^n)$ is generated by many different systems of sets. For our purposes the most interesting generators are the families of open rectangles

$$\mathscr{J}^o = \mathscr{J}^{o,n} = \mathscr{J}^o(\mathbb{R}^n) = \{(a_1, b_1) \times \cdots \times (a_n, b_n) : a_i, b_i \in \mathbb{R}\}$$

and (from the right) half-open rectangles

$$\mathscr{J} = \mathscr{J}^n = \mathscr{J}(\mathbb{R}^n) = \{[a_1, b_1) \times \cdots \times [a_n, b_n) : a_i, b_i \in \mathbb{R}\}.$$

We use the conventions that $[a_i, b_i) = (a_i, b_i) = \emptyset$ if $b_i \leqslant a_i$ and, of course, that $[a_1, b_1) \times \cdots \times \emptyset \times \cdots \times [a_n, b_n) = \emptyset$. Sometimes we use the shorthand $[\![a, b)\!) = [a_1, b_1) \times \cdots \times [a_n, b_n)$ for vectors $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n)$ from $\mathbb{R}^n$. Finally, we write $\mathscr{J}_{\mathrm{rat}}, \mathscr{J}^o_{\mathrm{rat}}$ for the (half-)open rectangles with rational endpoints. Notice that the half-open rectangles are intervals in $\mathbb{R}$, rectangles in $\mathbb{R}^2$, cuboids in $\mathbb{R}^3$ and hypercubes in all dimensions $n \geqslant 4$ (see Fig. 3.1).

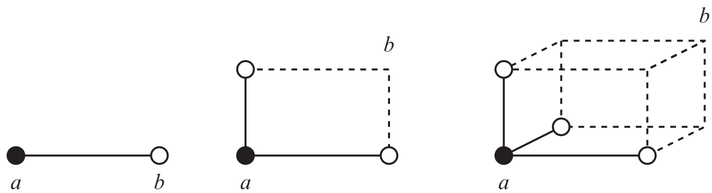


Fig. 3.1. ‘Rectangles’ in dimension $n = 1, 2, 3$.

Theorem 3.8 *We have* $\mathscr{B}(\mathbb{R}^n) = \sigma(\mathscr{J}^n_{rat}) = \sigma(\mathscr{J}^{o,n}_{rat}) = \sigma(\mathscr{J}^n) = \sigma(\mathscr{J}^{o,n})$.

Proof We begin with open rectangles having rational endpoints. Since the open rectangle $(\!(a, b)\!) = \bigtimes_{i=1}^n (a_i, b_i)$ is an open set [✍], we get the following inclusions: $\sigma(\mathscr{O}) \supset \sigma(\mathscr{J}^o) \supset \sigma(\mathscr{J}^o_{\mathrm{rat}})$.

[2] $B_k(0)$, $\overline{B_k(0)}$ denote the open, resp., closed balls with centre 0 and radius k.

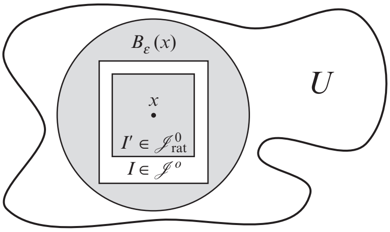


Fig. 3.2. The ball $B_\epsilon(x) \subset U$.

Conversely, if $U \in \mathscr{O}$, we have

$$U = \bigcup_{I \in \mathscr{J}^o_{\mathrm{rat}}, I \subset U} I. \tag{3.3}$$

Here '$\supset$' is clear from the definition and for the other direction '$\subset$' we fix some $x \in U$. Since U is open, there is some ball $B_\epsilon(x) \subset U$ – see Fig. 3.2 – and we can inscribe a square into $B_\epsilon(x)$ and then shrink this square to get a rectangle $I' = I'(x) \in \mathscr{J}^o_{\mathrm{rat}}$ containing x. Since every rectangle is uniquely determined by its main diagonal, there are at most $\#(\mathbb{Q}^n \times \mathbb{Q}^n) = \#\mathbb{N}$ many I in the union (3.3). Thus

$$U \in \mathscr{O} \subset \sigma(\mathscr{J}^o_{\mathrm{rat}}),$$

proving the other inclusion $\sigma(\mathscr{O}) \subset \sigma(\mathscr{J}^o_{\mathrm{rat}})$, and so $\sigma(\mathscr{O}) = \sigma(\mathscr{J}^o) = \sigma(\mathscr{J}^o_{\mathrm{rat}})$.

Every half-open rectangle (with rational endpoints) can be written as

$$[a_1, b_1) \times \cdots \times [a_n, b_n) = \bigcap_{i \in \mathbb{N}} (a_1 - \tfrac{1}{i}, b_1) \times \cdots \times (a_n - \tfrac{1}{i}, b_n),$$

while every open rectangle (with rational endpoints) can be represented as

$$(c_1, d_1) \times \cdots \times (c_n, d_n) = \bigcup_{i \in \mathbb{N}} [c_1 + \tfrac{1}{i}, d_1) \times \cdots \times [c_n + \tfrac{1}{i}, d_n).$$

These formulae imply that $\mathscr{J} \subset \sigma(\mathscr{J}^o)$ and $\mathscr{J}^o \subset \sigma(\mathscr{J})$ $\big[$resp. $\mathscr{J}_{\mathrm{rat}} \subset \sigma(\mathscr{J}^o_{\mathrm{rat}})$ and $\mathscr{J}^o_{\mathrm{rat}} \subset \sigma(\mathscr{J}_{\mathrm{rat}})\big]$, hence, by Remark 3.5(iii), $\sigma(\mathscr{J}^o) = \sigma(\mathscr{J})$ $\big[$resp. $\sigma(\mathscr{J}^o_{\mathrm{rat}}) = \sigma(\mathscr{J}_{\mathrm{rat}})\big]$ and the proof follows as we already know that $\sigma(\mathscr{J}^o_{\mathrm{rat}}) = \sigma(\mathscr{J}^o) = \sigma(\mathscr{O})$. $\square$

Remark 3.9 Let D be a dense subset of $\mathbb{R}$, e.g. $D = \mathbb{Q}$ or $D = \mathbb{R}$. The Borel sets of the real line $\mathbb{R}$ are also generated by any of the following systems:

$$\{(-\infty, a) : a \in D\}, \ \{(-\infty, a] : a \in D\}, \ \{(a, \infty) : a \in D\}, \ \{[a, \infty) : a \in D\}.$$

Remark 3.10 One might think that $\sigma(\mathscr{G})$ can be explicitly constructed for any given $\mathscr{G}$ by adding to the family $\mathscr{G}$ all possible countable unions of its members and their complements:

$$\mathscr{G}_{\sigma c} := \left\{ \bigcup_{n \in \mathbb{N}} G_n, \ \Big(\bigcup_{n \in \mathbb{N}} G_n \Big)^c \ : \ G_n \in \mathscr{G} \right\}.$$

But $\mathscr{G}_{\sigma c}$ is not necessarily a σ-algebra. [✍] Even if we repeat this procedure *countably* often, i.e.

$$\mathscr{G}_n := \underbrace{(\ldots(\mathscr{G}_{\sigma c})_{\sigma c}\cdots)_{\sigma c}}_{n \text{ times}}, \qquad \widehat{\mathscr{G}} := \bigcup_{n\in\mathbb{N}} \mathscr{G}_n,$$

we end up, in general, with a set that is too small: $\widehat{\mathscr{G}} \subsetneq \sigma(\mathscr{G})$.[3]

This shows that the σ-operation produces a pretty big family; so big, in fact, that no approach using countably many countable set operations will give the whole of $\sigma(\mathscr{G})$. On the other hand, it is rather typical that a σ-algebra is given through its generator. In order to deal with these cases, we need the notion of *Dynkin systems* which will be introduced in Chapter 5.

Problems

3.1. Let $\mathscr{A}$ be a σ-algebra. Show that

(i) if $A_1, A_2, \ldots, A_N \in \mathscr{A}$, then $A_1 \cap A_2 \cap \cdots \cap A_N \in \mathscr{A}$;
(ii) $A \in \mathscr{A}$ if, and only if, $A^c \in \mathscr{A}$;
(iii) if $A, B \in \mathscr{A}$, then $A \setminus B \in \mathscr{A}$ and $A \vartriangle B \in \mathscr{A}$.

3.2. Prove the assertions made in Example 3.3(iv), (vi) and (vii).
[Hint: use (2.6) for (vii).]

3.3. Let $X = \mathbb{R}$. Find the σ-algebra generated by the singletons $\{\{x\} : x \in \mathbb{R}\}$.

3.4. Verify the assertions made in Remark 3.5.

3.5. Let $X = [0, 1]$. Find the σ-algebra generated by the sets

(i) $\left(0, \frac{1}{2}\right)$;
(ii) $\left[0, \frac{1}{4}\right), \left(\frac{3}{4}, 1\right]$;
(iii) $\left[0, \frac{3}{4}\right], \left[\frac{1}{4}, 1\right]$.

3.6. Let $A_1, A_2, \ldots, A_N$ be non-empty subsets of X.

(i) If the A_n are disjoint and $\dot\bigcup A_n = X$, then $\#\sigma(A_1, A_2, \ldots, A_N) = 2^N$.
Remark. A set A in a σ-algebra $\mathscr{A}$ is called an *atom*, if there is no proper subset $\emptyset \neq B \subsetneq A$ such that $B \in \mathscr{A}$. In this sense all A_n are atoms.
(ii) Show that $\sigma(A_1, A_2, \ldots, A_N)$ consists of finitely many sets.
[Hint: show that $\sigma(A_1, A_2, \ldots, A_N)$ has only finitely many atoms.]

3.7. Let $(X, \mathscr{A})$ be a measurable space. Show that there cannot be a σ-algebra $\mathscr{A}$ which contains countably infinitely many sets.
[Hint: recall that $A \in \mathscr{A}$ is an *atom* if A contains no proper subset $\emptyset \neq B \in \mathscr{A}$. Show that $\#\mathscr{A} = \#\mathbb{N}$ implies that $\mathscr{A}$ has countably infinitely many atoms. This will lead to a contradiction.]

3.8. Let $(X, \mathscr{A})$ be a measurable space and $(\mathscr{A}_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence of σ-algebras, i.e. $\mathscr{A}_n \subsetneq \mathscr{A}_{n+1}$. Show that $\mathscr{A}_\infty := \bigcup_{n\in\mathbb{N}} \mathscr{A}_n$ is *never* a σ-algebra.

3.9. Verify the properties $(\mathscr{O}_1)$–$(\mathscr{O}_3)$ for open sets in $\mathbb{R}^n$. Is $\mathscr{O}$ a σ-algebra?

[3] A 'constructive' approach along these lines is nevertheless possible if we use transfinite induction, see Hewitt and Stromberg [22, Theorem 10.23] or Appendix G.

3.10. Find an example (e.g. in $\mathbb{R}$) showing that $\bigcap_{n\in\mathbb{N}} U_n$ need not be open even if all U_n are open sets.

3.11. Prove any one of the assertions made in Remark 3.9.

3.12. Denote by $B_r(x)$ an open ball in $\mathbb{R}^n$ with centre x and radius r. Show that the Borel sets $\mathscr{B}(\mathbb{R}^n)$ are generated by all open balls $\mathbb{B} := \{B_r(x) : x \in \mathbb{R}^n,\ r > 0\}$. Is this still true for the family $\mathbb{B}' := \{B_r(x) : x \in \mathbb{Q}^n,\ r \in \mathbb{Q}^+\}$?
[Hint: mimic the proof of Theorem 3.8.]

3.13. Let $\mathscr{O}$ be the collection of open sets (topology) in $\mathbb{R}^n$ and let $A \subset \mathbb{R}^n$ be an arbitrary subset. We can introduce a topology $\mathscr{O}_A$ on A as follows: a set $V \subset A$ is called open (relative to A) if $V = U \cap A$ for some $U \in \mathscr{O}$. We write $\mathscr{O}_A$ for the open sets relative to A.

(i) Show that $\mathscr{O}_A$ is a topology on A, i.e. a family satisfying $(\mathscr{O}_1)$–$(\mathscr{O}_3)$.

(ii) If $A \in \mathscr{B}(\mathbb{R}^n)$, show that the trace σ-algebra $A \cap \mathscr{B}(\mathbb{R}^n)$ coincides with $\sigma(\mathscr{O}_A)$ (the latter is usually denoted by $\mathscr{B}(A)$: the Borel sets relative to A).

3.14. **Monotone classes (1).** A family $\mathscr{M} \subset \mathscr{P}(X)$ which contains X and is stable under countable unions of increasing sets and countable intersections of decreasing sets

$$(A_n)_{n\in\mathbb{N}} \subset \mathscr{M}, \quad A_1 \subset \cdots \subset A_n \subset A_{n+1} \uparrow A = \bigcup_{n\in\mathbb{N}} A_n \implies A \in \mathscr{M} \qquad (\mathrm{MC}_1)$$

$$(B_n)_{n\in\mathbb{N}} \subset \mathscr{M}, \quad B_1 \supset \cdots \supset B_n \supset B_{n+1} \downarrow B = \bigcap_{n\in\mathbb{N}} B_n \implies B \in \mathscr{M} \qquad (\mathrm{MC}_2)$$

is called a *monotone class*. Assume that $\mathscr{M}$ is a monotone class and $\mathscr{F} \subset \mathscr{P}(X)$ any family of sets.

(i) Mimic the proof of Theorem 3.4 to show that there is a minimal monotone class $\mathfrak{m}(\mathscr{F})$ such that $\mathscr{F} \subset \mathfrak{m}(\mathscr{F})$.

(ii) If $\mathscr{F}$ is stable w.r.t. complements, i.e. $F \in \mathscr{F} \implies F^c \in \mathscr{F}$, then so is $\mathfrak{m}(\mathscr{F})$.

(iii) If $\mathscr{F}$ is $\cap$-stable, i.e. $F, G \in \mathscr{F} \implies F \cap G \in \mathscr{F}$, then so is $\mathfrak{m}(\mathscr{F})$.
[Hint: show that the families

$$\Sigma := \{M \in \mathfrak{m}(\mathscr{F}) : M \cap F \in \mathfrak{m}(\mathscr{F})\ \forall F \in \mathscr{F}\}$$
$$\Sigma' := \{M \in \mathfrak{m}(\mathscr{F}) : M \cap N \in \mathfrak{m}(\mathscr{F})\ \forall N \in \mathfrak{m}(\mathscr{F})\}$$

are again monotone classes satisfying $\mathscr{F} \subset \Sigma, \Sigma'$.]

(iv) Use (i)–(iii) to prove the following.

Monotone class theorem. *Let $\mathscr{F} \subset \mathscr{P}(X)$ be a family of sets which is stable under the formation of complements and intersections. If $\mathscr{M} \supset \mathscr{F}$ is a monotone class, then $\mathscr{M} \supset \sigma(\mathscr{F})$.*

3.15. **Alternative characterization of $\mathscr{B}(\mathbb{R}^n)$.** In older books the Borel sets are often introduced as the smallest family $\mathscr{M}$ of sets which is stable under countable intersections of decreasing and countable unions of increasing sequences of sets, and which contains all open sets $\mathscr{O}$. Use Problem 3.14 to show that $\mathscr{M} = \mathscr{B}(\mathbb{R}^n)$.
Can we omit 'decreasing' and 'increasing' in the above characterization, i.e. is $\mathscr{B}(\mathbb{R}^n)$ also the smallest family containing countable intersections and countable unions of its members, and all open sets?

3.16. Let X be an arbitrary set and $\mathscr{F} \subset \mathscr{P}(X)$. Show that

$$\sigma(\mathscr{F}) = \bigcup \{\sigma(\mathscr{C}) : \mathscr{C} \subset \mathscr{F} \text{ countable sub-family}\}.$$

4

Measures

We are now ready to introduce one of the central concepts of measure and integration theory: measures. As before, X is some set and $\mathscr{A}$ is a σ-algebra on X.

Definition 4.1 A (positive) *measure* μ on X is a map $\mu : \mathscr{A} \to [0, \infty]$ satisfying

$$\mathscr{A} \text{ is a } \sigma\text{-algebra in } X, \tag{M_0}$$

$$\mu(\emptyset) = 0, \tag{M_1}$$

$$(A_n)_{n\in\mathbb{N}} \subset \mathscr{A} \text{ pairwise disjoint} \implies \mu\Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n). \tag{M_2}$$

If μ satisfies (M_1), (M_2), but $\mathscr{A}$ is not a σ-algebra, then μ is said to be a *pre-measure*.

Caution (M_2) requires implicitly that $\dot{\bigcup}_n A_n$ is again in $\mathscr{A}$ – this is clearly the case for σ-algebras, but needs special attention if one deals with pre-measures.

We will frequently use the following notation for *increasing* and *decreasing sequences of sets*:

$$A_n \uparrow A \iff A_1 \subset A_2 \subset A_3 \subset \ldots \quad \text{and} \quad A = \bigcup\nolimits_{n\in\mathbb{N}} A_n,$$

$$B_n \downarrow B \iff B_1 \supset B_2 \supset B_3 \supset \ldots \quad \text{and} \quad B = \bigcap\nolimits_{n\in\mathbb{N}} B_n.$$

Definition 4.2 Let X be a set and $\mathscr{A}$ a σ-algebra on X. The pair $(X, \mathscr{A})$ is called *measurable space*. If μ is a measure on X, $(X, \mathscr{A}, \mu)$ is called *measure space*.

A *finite measure* is a measure with $\mu(X) < \infty$, and a *probability measure* is a measure with $\mu(X) = 1$. The corresponding measure spaces are called a *finite measure space*, resp. *probability space*.

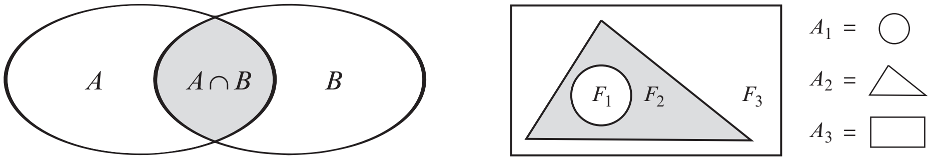


Fig. 4.1. *Left:* strong additivity of measures (Proposition 4.3(iv)). *Right:* Continuity from below vs. σ-additivity (Proposition 4.3(vi)).

A measure μ is said to be *σ-finite* and $(X, \mathscr{A}, \mu)$ is called a *σ-finite measure space*, if $\mathscr{A}$ contains a sequence $(A_n)_{n\in\mathbb{N}}$ such that $A_n \uparrow X$ and $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$.

Let us derive some immediate properties of measures and pre-measures.

Proposition 4.3 *Let $(X, \mathscr{A}, \mu)$ be a measure space and $A, B, A_n, B_n \in \mathscr{A}$, $n \in \mathbb{N}$. Then*

(i) $A \cap B = \emptyset \implies \mu(A \mathbin{\dot\cup} B) = \mu(A) + \mu(B)$ (*additive*),

(ii) $A \subset B \implies \mu(A) \leqslant \mu(B)$ (*monotone*),

(iii) $A \subset B,\ \mu(A) < \infty \implies \mu(B \setminus A) = \mu(B) - \mu(A)$,

(iv) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$ (*strongly additive*),

(v) $\mu(A \cup B) \leqslant \mu(A) + \mu(B)$ (*subadditive*),

(vi) $A_n \uparrow A \implies \mu(A) = \sup_{n\in\mathbb{N}} \mu(A_n) = \lim_{n\to\infty} \mu(A_n)$ (*continuous from below*),

(vii) $B_n \downarrow B,\ \mu(B_1) < \infty \implies \mu(B) = \inf_{n\in\mathbb{N}} \mu(B_n) = \lim_{n\to\infty} \mu(B_n)$ (*continuous from above*),

(viii) $\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \leqslant \sum_{n\in\mathbb{N}} \mu(A_n)$ (*σ-subadditive*).

Proof (i) Set $A_1 := A, A_2 := B, A_3 = A_4 = \cdots = \emptyset$. Then $(A_n)_{n\in\mathbb{N}}$ is a family of pairwise disjoint sets from $\mathscr{A}$. Moreover, $A \mathbin{\dot\cup} B = \dot\bigcup_n A_n$ and by (M_1), (M_2)

$$\begin{aligned}\mu(A \mathbin{\dot\cup} B) = \mu\Big(\dot\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n) &= \mu(A) + \mu(B) + \mu(\emptyset) + \ldots \\ &= \mu(A) + \mu(B).\end{aligned}$$

(ii) If $A \subset B$, we have $B = A \mathbin{\dot\cup} (B \setminus A)$, and by (i)

$$\mu(B) = \mu(A \mathbin{\dot\cup} (B \setminus A)) = \mu(A) + \mu(B \setminus A) \tag{4.1}$$

$$\geqslant \mu(A). \tag{4.2}$$

(iii) If $A \subset B$, we can subtract the finite number $\mu(A)$ from both sides of (4.1) to get $\mu(B) - \mu(A) = \mu(B \setminus A)$.

(iv) For all $A, B \in \mathscr{A}$ we have $A \cup B = \big(A \mathbin{\dot\cup} (B \setminus (A \cap B)\big)$ (see Figure 4.1) and using (i) and (4.1) we get

$$\begin{aligned}
\mu(A \cup B) + \mu(A \cap B) &= \mu\big(A \mathbin{\dot\cup} (B \setminus (A \cap B)\big) + \mu(A \cap B) \\
&\overset{\text{(i)}}{=} \mu(A) + \mu(B \setminus (A \cap B)) + \mu(A \cap B) \\
&\overset{(4.1)}{=} \mu(A) + \mu(B).
\end{aligned}$$

(v) From (iv) we get $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B) \geqslant \mu(A \cup B)$ for all $A, B \in \mathscr{A}$.

(vi) Set $F_1 := A_1$, $F_2 := A_2 \setminus A_1$, ..., $F_{n+1} := A_{n+1} \setminus A_n$. Since the sets F_n are pairwise disjoint, we get

$$A_n = \dot{\bigcup_{i=1}^{n}} F_i \implies A = \dot{\bigcup_{i=1}^{\infty}} F_i = \bigcup_{n=1}^{\infty} A_n$$

and

$$\mu(A) = \mu\left(\dot{\bigcup_{i=1}^{\infty}} F_i\right) \overset{(\mathrm{M}_2)}{=} \sum_{i=1}^{\infty} \mu(F_i) = \lim_{n\to\infty} \underbrace{\sum_{i=1}^{n} \mu(F_i)}_{\overset{\text{(i)}}{=} \lim_{n\to\infty} \mu\left(\dot\bigcup_{i=1}^{n} F_i\right)} = \lim_{n\to\infty} \mu(A_n).$$

(vii) Since $B_n \downarrow B$, we get $B_1 \setminus B_n \uparrow B_1 \setminus B$. By assumption $\mu(B_1) < \infty$, which means that we can use (vi) and (iii) to get

$$\begin{aligned}
\mu(B_1) - \mu(B) = \mu(B_1 \setminus B) &= \lim_{n\to\infty} \mu(B_1 \setminus B_n) \\
&= \lim_{n\to\infty} (\mu(B_1) - \mu(B_n)) \\
&= \mu(B_1) - \lim_{n\to\infty} \mu(B_n).
\end{aligned}$$

A simple rearrangement proves the claim.

(viii)
$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} A_n\right) &\overset{\text{(vi)}}{=} \lim_{n\to\infty} \mu(A_1 \cup \dots \cup A_n) \\
&\overset{\text{(v)}}{\leqslant} \lim_{n\to\infty} (\mu(A_1) + \dots + \mu(A_n)) = \sum_{n=1}^{\infty} \mu(A_n).
\end{aligned}$$
□

Remark 4.4 With some obvious re-wordings, Proposition 4.3 remains valid for pre-measures, i.e. for families $\mathscr{A}$ which are not σ-algebras. Of course, one has to make sure that $\emptyset \in \mathscr{A}$ and that $\mathscr{A}$ is stable under

- finite unions, intersections and differences of sets (for 4.3(i)–(v));[1]
- finite differences and countable unions of sets (for 4.3(vi), (viii));
- finite differences and countable intersections of sets (for 4.3(vii)).

The proofs are literally the same.

It is about time to give some examples of measures. At this stage this is, unfortunately, a somewhat difficult task! The main problem is that we have to say for *every* set of the σ-algebra $\mathscr{A}$ what its measure $\mu(A)$ shall be. Since $\mathscr{A}$ can be very large – see Remark 3.10 – this is, in general, possible only (explicitly!) if either μ or $\mathscr{A}$ is very simple. Nevertheless …

Example 4.5 (i) **(Dirac measure, unit mass)** Let $(X, \mathscr{A})$ be a measurable space and let $x \in X$ be some point. Then $\delta_x : \mathscr{A} \to \{0, 1\}$, defined for $A \in \mathscr{A}$ by

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

is a measure. It is called *Dirac's delta measure* or *unit mass* at the point x.

(ii) Consider $(\mathbb{R}, \mathscr{A})$ with $\mathscr{A}$ from Example 3.3(v) (i.e. $A \in \mathscr{A}$ if A or A^c is countable). Then $\gamma : \mathscr{A} \to \{0, 1\}$, defined for $A \in \mathscr{A}$ by

$$\gamma(A) := \begin{cases} 0 & \text{if } A \text{ is countable}, \\ 1 & \text{otherwise}, \end{cases}$$

is a measure.

(iii) **(Counting measure)** Let $(X, \mathscr{A})$ be a measurable space and $A \in \mathscr{A}$. Then

$$|A| := \begin{cases} \#A & \text{if } A \text{ is finite}, \\ +\infty & \text{if } A \text{ is infinite}, \end{cases}$$

defines a measure. It is called *counting measure*.

(iv) **(Discrete probability measure)** Let $\Omega = \{\omega_1, \omega_2, \dots\}$ be a countable set and $(p_n)_{n\in\mathbb{N}}$ be a sequence of real numbers $p_n \in [0, 1]$ such that $\sum_{n\in\mathbb{N}} p_n = 1$. On $(\Omega, \mathscr{P}(\Omega))$ the set function

$$\mathbb{P}(A) = \sum_{n:\omega_n\in A} p_n = \sum_{n\in\mathbb{N}} p_n \delta_{\omega_n}(A), \qquad A \subset \Omega,$$

defines a probability measure. The triplet $(\Omega, \mathscr{P}(\Omega), \mathbb{P})$ is called *discrete probability space*.[2]

[1] Such a family is called a *ring* of sets.

[2] Since $\mathbb{P}(A) = \mathbb{P}\big(\bigcup_{\omega\in A}\{\omega\}\big) = \sum_{\omega\in A} \mathbb{P}(\{\omega\})$, the measure $\mathbb{P}$ is uniquely determined by the values $\mathbb{P}(\{\omega_n\}) = p_n$, i.e. (iv) describes the most general *discrete* probability measure.

(v) **(Trivial measures)** Let $(X, \mathscr{A})$ be a measurable space and $A \in \mathscr{A}$. Then

$$\mu(A) := \begin{cases} 0 & \text{if } A = \emptyset, \\ +\infty & \text{if } A \neq \emptyset \end{cases} \quad \text{and} \quad \nu(A) := 0$$

are measures.

Our list of examples does not include the most familiar of all measures: length, area and volume.

Definition 4.6 The set function λ^n on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ that assigns every half-open rectangle $[\![a, b)\!) = [a_1, b_1) \times \cdots \times [a_n, b_n) \in \mathscr{J}$ the value

$$\lambda^n [\![a, b)\!) := \prod_{i=1}^{n} (b_i - a_i)$$

is called *n-dimensional Lebesgue measure*.

The problem here is that we do not know whether λ^n is a measure in the sense of Definition 4.1: λ^n is explicitly given only on the half-open rectangles $\mathscr{J}$ and it is not obvious at all that λ^n is a pre-measure on $\mathscr{J}$; much less clear is the question of whether and how we can extend this pre-measure from $\mathscr{J}$ to a proper measure on $\sigma(\mathscr{J})$. Over the next few chapters we will see that such an extension is indeed possible. But this requires some extra work and a more abstract approach. One of the main obstacles is, of course, that $\sigma(\mathscr{J})$ cannot be obtained by a bare-hands construction from $\mathscr{J}$.

Let us, meanwhile, note the upshot of what will be proved in the next chapters.

Theorem 4.7 *Lebesgue measure λ^n exists, is a measure on the Borel sets $\mathscr{B}(\mathbb{R}^n)$ and is uniquely determined by its values on the rectangles $\mathscr{J}$. Moreover, λ^n enjoys the following properties for $B \in \mathscr{B}(\mathbb{R}^n)$:*

(i) *λ^n is invariant under translations: $\lambda^n(x + B) = \lambda^n(B)$, $x \in \mathbb{R}^n$;*
(ii) *λ^n is invariant under motions: $\lambda^n(R^{-1}(B)) = \lambda^n(B)$, where R is a motion, i.e. a combination of translations, rotations and reflections;*
(iii) *$\lambda^n(M^{-1}(B)) = |\det M|^{-1} \lambda^n(B)$ for any invertible matrix $M \in \mathbb{R}^{n \times n}$.*

Clearly, the sets $x + B := \{x + y : y \in B\}$, $R^{-1}(B) := \{R^{-1}(y) : y \in B\}$ and $M^{-1}(B)$ must again be Borel sets, otherwise the statement of Theorem 4.7 would be senseless, see Theorem 5.8 and Chapter 7.

Let us finally show that the continuity of measures, see Proposition 4.3(vi) and (vii), is essentially equivalent to σ-additivity (M_2).

✻Lemma 4.8 *Let $(X,\mathscr{A})$ be a measure space and $\mu:\mathscr{A}\to[0,\infty]$ be an additive set function such that $\mu(\emptyset)=0$. Then μ is a measure if, and only if, μ is continuous from below, i.e. if $A_n\in\mathscr{A}$ and $A_n\uparrow A$, then $\mu(A)=\sup_{n\in\mathbb{N}}\mu(A_n)=\lim_{n\to\infty}\mu(A_n)$.*

Proof Any measure μ is continuous from below, see Proposition 4.3(vi). Conversely, assume that μ is an additive set function which is continuous from below. If $(B_n)_{n\in\mathbb{N}}\subset\mathscr{A}$ are disjoint sets, then $A_n:=B_1\cup\cdots\cup B_n\in\mathscr{A}$ is an increasing sequence satisfying

$$A=\bigcup_{n\in\mathbb{N}}A_n=\dot{\bigcup_{n\in\mathbb{N}}}B_n.$$

Therefore,

$$\mu\left(\dot{\bigcup_{i=1}^{\infty}}B_i\right)\overset{4.3(\text{vi})}{=}\lim_{n\to\infty}\mu\left(\dot{\bigcup_{i=1}^{n}}B_i\right)\overset{4.3(\text{i})}{=}\lim_{n\to\infty}\sum_{i=1}^{n}\mu(B_i)=\sum_{i=1}^{\infty}\mu(B_i).\qquad\Box$$

✻Lemma 4.9 *Let $(X,\mathscr{A})$ be a measure space and $\mu:\mathscr{A}\to[0,\infty)$ be an additive set function such that $\mu(\emptyset)=0$ and $\mu(A)<\infty$ for all $A\in\mathscr{A}$. Then μ is a measure if, and only if, one of the of the following continuity properties is satisfied:*

(i) *μ is continuous from below, i.e. Proposition 4.3(vi) holds;*
(ii) *μ is continuous from above, i.e. Proposition 4.3(vii) holds;*
(iii) *μ is continuous at $\emptyset$, i.e. Proposition 4.3(vii) holds for $B=\emptyset$.*

Proof By Proposition 4.3, every measure enjoys the properties (i)–(iii). On the other hand, we know from the proof of Proposition 4.3(vii) that (i) $\Rightarrow$ (ii) while (ii) $\Rightarrow$ (iii) is trivial. It then suffices, therefore, to show that (iii) entails σ-additivity. Let $(A_n)_{n\in\mathbb{N}}\subset\mathscr{A}$ be pairwise disjoint sets and $A=\dot{\bigcup}_{n=1}^{\infty}A_n$. Since $B_n:=A\setminus(A_1\,\dot{\cup}\,\ldots\,\dot{\cup}\,A_n)\downarrow\emptyset$, we find by additivity

$$\begin{aligned}\mu(A)&\overset{\text{additive}}{=}\mu(A\setminus(A_1\,\dot{\cup}\,\ldots\,\dot{\cup}\,A_n))+\mu(A_1\,\dot{\cup}\,\ldots\,\dot{\cup}\,A_n)\\&\overset{\text{additive}}{=}\mu(B_n)+\sum_{i=1}^{n}\mu(A_i)\xrightarrow[n\to\infty]{(\text{iii})}\sum_{i=1}^{\infty}\mu(A_i).\end{aligned}\qquad\Box$$

Problems

4.1. Extend Proposition 4.3(i), (iv) and (v) to finitely many sets $A_1,A_2,\ldots,A_N\in\mathscr{A}$.

4.2. Check that the set functions defined in Example 4.5 are measures in the sense of Definition 4.1.

4.3. Is the set function γ of Example 4.5(ii) still a measure on the measurable space $(\mathbb{R},\mathscr{B}(\mathbb{R}))$? Is it still a measure on the measurable space $(\mathbb{Q},\mathbb{Q}\cap\mathscr{B}(\mathbb{R}))$?

4.4. Let $X = \mathbb{R}$. For which σ-algebras are the following set functions measures:

$$\text{(i)}\quad \mu(A) = \begin{cases} 0, & \text{if } A = \emptyset, \\ 1, & \text{if } A \neq \emptyset; \end{cases} \qquad \text{(ii)}\quad \nu(A) = \begin{cases} 0, & \text{if } A \text{ is finite}, \\ 1, & \text{if } A^c \text{ is finite?} \end{cases}$$

4.5. Find an example showing that the finiteness condition in Proposition 4.3(vii) or Lemma 4.9 is essential.
[Hint: use Lebesgue measure or the counting measure on infinite tails $[k, \infty) \downarrow \emptyset$.]

4.6. Find a measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ which is σ-finite but assigns to every interval $[a, b)$ with $b - a > 2$ finite mass.

4.7. Let $(X, \mathscr{A})$ be a measurable space.

(i) Let μ, ν be two measures on $(X, \mathscr{A})$. Show that the set function $\rho(A) := a\mu(A) + b\nu(A)$, $A \in \mathscr{A}$, for all $a, b \geqslant 0$ is again a measure.

(ii) Let $\mu_1, \mu_2, \ldots$ be countably many measures on $(X, \mathscr{A})$ and let $(\alpha_i)_{i\in\mathbb{N}}$ be a sequence of positive numbers. Show that $\mu(A) := \sum_{i=1}^{\infty} \alpha_i \mu_i(A)$, $A \in \mathscr{A}$, is again a measure.
[Hint: to show σ-additivity use (and prove) the following helpful lemma. *For any double sequence β_{ik}, $i, k \in \mathbb{N}$, of real numbers we have*

$$\sup_{i\in\mathbb{N}} \sup_{k\in\mathbb{N}} \beta_{ik} = \sup_{k\in\mathbb{N}} \sup_{i\in\mathbb{N}} \beta_{ik}.$$

Thus $\lim_{i\to\infty} \lim_{k\to\infty} \beta_{ik} = \lim_{k\to\infty} \lim_{i\to\infty} \beta_{ik}$ if $i \mapsto \beta_{ik}$, and $k \mapsto \beta_{ik}$ increases when the other index is fixed.]

4.8. Let $(X, \mathscr{A})$ be a measurable space and assume that $\mu : \mathscr{A} \to [0, \infty]$ finitely additive and σ-subadditive. Show that μ is σ-additive.

4.9. Let $(X, \mathscr{A}, \mu)$ be a measure space and $F \in \mathscr{A}$. Show that $\mathscr{A} \ni A \mapsto \mu(A \cap F)$ defines a measure.

4.10. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ a sequence of sets with $\mathbb{P}(A_n) = 1$ for all $n \in \mathbb{N}$. Show that $\mathbb{P}\big(\bigcap_{n\in\mathbb{N}} A_n\big) = 1$.

4.11. Let $(X, \mathscr{A}, \mu)$ be a finite measure space and $(A_n)_{n\in\mathbb{N}}, (B_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ such that $A_n \supset B_n$ for all $n \in \mathbb{N}$. Show that

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) - \mu\Big(\bigcup_{n\in\mathbb{N}} B_n\Big) \leqslant \sum_{n\in\mathbb{N}} \big(\mu(A_n) - \mu(B_n)\big).$$

[Hint: show first that $\bigcup_n A_n \setminus \bigcup_k B_k \subset \bigcup_n (A_n \setminus B_n)$ then use Proposition 4.3(viii).]

4.12. **Null sets.** Let $(X, \mathscr{A}, \mu)$ be a measure space. A set $N \in \mathscr{A}$ is called a *null set* or *μ-null set* if $\mu(N) = 0$. We write $\mathscr{N}_\mu$ for the family of all μ-null sets. Check that $\mathscr{N}_\mu$ has the following properties:

(i) $\emptyset \in \mathscr{N}_\mu$;

(ii) if $N \in \mathscr{N}_\mu$, $M \in \mathscr{A}$ and $M \subset N$, then$M \in \mathscr{N}_\mu$;

(iii) if $(N_n)_{n\in\mathbb{N}} \subset \mathscr{N}_\mu$, then $\bigcup_{n\in\mathbb{N}} N_n \in \mathscr{N}_\mu$.

4.13. Let λ be one-dimensional Lebesgue measure.

(i) Show that for all $x \in \mathbb{R}$ the set $\{x\}$ is a Borel set with $\lambda\{x\} = 0$.
[Hint: consider the intervals $[x - 1/k, x + 1/k)$, $k \in \mathbb{N}$ and use Proposition 4.3(vii).]

(ii) Prove in two ways that $\mathbb{Q}$ is a Borel set and $\lambda(\mathbb{Q}) = 0$:

(a) by using the first part of the problem;

(b) by considering the set $C(\epsilon) := \bigcup_{n\in\mathbb{N}} [q_n - \epsilon 2^{-n}, q_n + \epsilon 2^{-n})$, where $(q_n)_{n\in\mathbb{N}}$ is an enumeration of $\mathbb{Q}$, and letting $\epsilon \to 0$.

(iii) Use the trivial fact that $[0,1]=\bigcup_{0\leqslant x\leqslant 1}\{x\}$ to show that a non-countable union of null sets (here $\{x\}$) is not necessarily a null set.

4.14. Determine all null sets of the measure $\delta_a+\delta_b$, $a,b\in\mathbb{R}$, on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$.

4.15. Completion (1). We have seen in Problem 4.12 that *measurable* subsets of null sets are again null sets: $M\in\mathscr{A}$, $M\subset N\in\mathscr{A}$, $\mu(N)=0$ then $\mu(M)=0$; but there might be subsets of N which are not in $\mathscr{A}$. This motivates the following definition: *a measure space* $(X,\overline{\mathscr{A}},\mu)$ (*or a measure* μ) *is* ***complete*** *if all subsets of* μ*-null sets are again in* $\overline{\mathscr{A}}$. In other words, it holds if all subsets of a null set are null sets.
The following exercise shows that a measure space $(X,\mathscr{A},\mu)$ which is not yet complete can be completed.

(i) $\overline{\mathscr{A}}:=\{A\cup N: A\in\mathscr{A},\quad N \text{ is a subset of some } \mathscr{A}\text{-measurable null set}\}$ is a σ-algebra satisfying $\mathscr{A}\subset\overline{\mathscr{A}}$.

(ii) $\bar{\mu}(A^*):=\mu(A)$ for $A^*=A\cup N\in\overline{\mathscr{A}}$ is well-defined, i.e. it is independent of the way we can write A^*, say as $A^*=A\cup N=B\cup M$, where $A,B\in\mathscr{A}$ and M,N are subsets of null sets.

(iii) $\bar{\mu}$ is a measure on $\overline{\mathscr{A}}$ and $\bar{\mu}(A)=\mu(A)$ for all $A\in\mathscr{A}$.

(iv) $(X,\overline{\mathscr{A}},\bar{\mu})$ is complete.

(v) we have $\overline{\mathscr{A}}=\{A^*\subset X: \exists A,B\in\mathscr{A},\ A\subset A^*\subset B,\ \mu(B\setminus A)=0\}$.

4.16. Let $(X,\mathscr{A},\mu)$ be a measure space, let $\mathscr{F}\subset\mathscr{A}$ be a sub-σ-algebra and denote the collection of all μ-null sets by $\mathscr{N}=\{N\in\mathscr{A}:\mu(N)=0\}$. Then

$$\sigma(\mathscr{F},\mathscr{N})=\big\{F\,\triangle\,N: F\in\mathscr{F},\ N\in\mathscr{N}\big\},$$

where $F\,\triangle\,N:=(F\setminus N)\cup(N\setminus F)$ denotes the symmetric difference, see Problem 2.2.

4.17. Let $\overline{\mathscr{A}}$ denote the completion of $\mathscr{A}$ as in Problem 4.15 and write

$$\mathscr{N}:=\{N\subset X: \exists M\in\mathscr{A},\ N\subset M,\ \mu(M)=0\}$$

for the family of all subsets of $\mathscr{A}$-measurable null sets. Show that

$$\overline{\mathscr{A}}=\sigma(\mathscr{A},\mathscr{N})=\big\{A\,\triangle\,N: A\in\mathscr{A},\ N\in\mathscr{N}\big\}.$$

Conclude that for every set $A^*\in\overline{\mathscr{A}}$ there is some $A\in\mathscr{A}$ such that $A\,\triangle\,A^*\in\mathscr{N}$.

4.18. Consider on $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$ the Dirac measure δ_x for some fixed $x\in\mathbb{R}^n$. Find the completion of $\mathcal{B}(\mathbb{R}^n)$ with respect to δ_x.

4.19. Restriction. Let $(X,\mathscr{A},\mu)$ be a measure space and let $\mathscr{B}\subset\mathscr{A}$ be a sub-σ-algebra. Denote by $\nu:=\mu|_{\mathscr{B}}$ the *restriction* of μ to $\mathscr{B}$.

(i) Show that ν is again a measure.

(ii) Assume that μ is a finite measure [a probability measure]. Is ν still a finite measure [a probability measure]?

(iii) Does ν inherit σ-finiteness from μ?

4.20. Show that a measure space $(X,\mathscr{A},\mu)$ is σ-finite if, and only if, there exists a sequence of measurable sets $(E_n)_{n\in\mathbb{N}}\subset\mathscr{A}$ such that $\bigcup_{n\in\mathbb{N}}E_n=X$ and $\mu(E_n)<\infty$ for all $n\in\mathbb{N}$.

4.21. Regularity. Let X be a metric space and μ be a finite measure on the Borel sets $\mathscr{B}=\mathscr{B}(X)$ and denote the open sets by $\mathscr{O}$ and the closed sets by $\mathscr{F}$. Define a family of sets

$$\Sigma:=\{A\subset X: \forall\epsilon>0\ \exists U\in\mathscr{O},\ F\in\mathscr{F}\ \text{ s.t. }\ F\subset A\subset U,\ \mu(U\setminus F)<\epsilon\}.$$

(i) Show that $A\in\Sigma\implies A^c\in\Sigma$ and that $\mathscr{F}\subset\Sigma$.

(ii) Show that Σ is stable under finite intersections.

(iii) Show that Σ is a σ-algebra containing the Borel sets $\mathscr{B}$.

(iv) Conclude that μ is *regular*, i.e. for all Borel sets $B \in \mathscr{B}$

$$\mu(B) = \sup_{F \subset B,\, F \in \mathscr{F}} \mu(F) = \inf_{U \supset B,\, U \in \mathscr{O}} \mu(U).$$

(v) Assume that there exists an increasing sequence of compact sets K_j such that $K_j \uparrow X$. Show that μ satisfies

$$\mu(B) = \sup_{K \subset B,\, K \text{ compact}} \mu(K).$$

(vi) Extend the equality $\mu(B) = \sup_{F \subset B,\, F \in \mathscr{F}} \mu(F)$ to a σ-finite measure μ.

4.22. Regularity on Polish spaces. A Polish space X is a complete metric space which has a countable dense subset $D \subset X$.

Let μ be a finite measure on $(X, \mathscr{B}(X))$. Then μ is regular in the sense that

$$\mu(B) = \sup_{K \subset B,\, K \text{ compact}} \mu(K) = \inf_{U \supset B,\, U \text{ open}} \mu(U).$$

[Hint: use Problem 4.21 and show that the set $K := \bigcap_n \bigcup_{j=1}^{k(n)} K_{1/n}(d_{k(n)})$ is compact; $K_\epsilon(x)$ denotes a closed ball of radius ϵ and centre x and $\{d_k\}_k$ is an enumeration of D. Choosing $k(n)$ sufficiently large, we can achieve that $\mu(X \setminus K) < \epsilon$.]

5
Uniqueness of Measures

Before we embark on the proof of the existence of measures in the following chapter, let us first check whether it suffices to consider measures on some *generator* $\mathscr{G}$ of a σ-algebra – otherwise our construction of Lebesgue measure would be flawed from the start.

As mentioned in Remark 3.10 a major problem is that, apart from trivial cases, $\sigma(\mathscr{G})$ cannot be constructively obtained from $\mathscr{G}$. To overcome this obstacle we need a new concept.

Definition 5.1 A family $\mathscr{D} \subset \mathscr{P}(X)$ is a *Dynkin system* if

$$X \in \mathscr{D}, \tag{D$_1$}$$

$$D \in \mathscr{D} \implies D^c \in \mathscr{D}, \tag{D$_2$}$$

$$(D_n)_{n\in\mathbb{N}} \subset \mathscr{D} \text{ pairwise disjoint} \implies \dot{\bigcup_{n\in\mathbb{N}}} D_n \in \mathscr{D}. \tag{D$_3$}$$

Remark 5.2 As for σ-algebras, see Properties 3.2, one sees that $\emptyset \in \mathscr{D}$ and that finite disjoint unions are again in $\mathscr{D}$: $D, E \in \mathscr{D}$, $D \cap E = \emptyset \implies D \,\dot{\cup}\, E \in \mathscr{D}$. Of course, every σ-algebra is a Dynkin system, but the converse is, in general, wrong, [✍] Problem 5.2.

Proposition 5.3 *Let $\mathscr{G} \subset \mathscr{P}(X)$. Then there is a smallest (also minimal, coarsest) Dynkin system $\delta(\mathscr{G})$ containing $\mathscr{G}$. $\delta(\mathscr{G})$ is called the Dynkin system generated by $\mathscr{G}$. Moreover, $\mathscr{G} \subset \delta(\mathscr{G}) \subset \sigma(\mathscr{G})$.*

Proof The proof that $\delta(\mathscr{G})$ exists parallels the proof of Theorem 3.4(ii). As in the case of σ-algebras, $\delta(\mathscr{D}) = \mathscr{D}$ if $\mathscr{D}$ is a Dynkin system (by minimality) and so $\delta(\sigma(\mathscr{G})) = \sigma(\mathscr{G})$. Hence, $\mathscr{G} \subset \sigma(\mathscr{G})$ implies that $\delta(\mathscr{G}) \subset \delta(\sigma(\mathscr{G})) = \sigma(\mathscr{G})$. □

It is important to know when a Dynkin system is already a σ-algebra.

Lemma 5.4 *A Dynkin system $\mathscr{D}$ is a σ-algebra if, and only if, it is stable under finite intersections:*[1] $D, E \in \mathscr{D} \implies D \cap E \in \mathscr{D}$.

Proof Since a σ-algebra is $\cap$-stable (see Properties 3.2 and Problem 3.1) as well as a Dynkin system (Remark 5.2) it remains only to show that a $\cap$-stable Dynkin system $\mathscr{D}$ is a σ-algebra.

Let $(D_n)_{n\in\mathbb{N}}$ be a sequence of sets in $\mathscr{D}$. Let us show that $D := \bigcup_{n\in\mathbb{N}} D_n \in \mathscr{D}$. Set $E_1 := D_1 \in \mathscr{D}$ and

$$\begin{aligned} E_{n+1} &:= (\cdots((D_{n+1} \setminus D_n) \setminus D_{n-1}) \setminus \cdots) \setminus D_1 \\ &= D_{n+1} \cap D_n^c \cap D_{n-1}^c \cap \cdots \cap D_1^c \in \mathscr{D} \end{aligned}$$

where we use (D_2) and the assumed $\cap$-stability of $\mathscr{D}$. The E_n are obviously mutually disjoint and $D = \biguplus_{n\in\mathbb{N}} E_n \in \mathscr{D}$ by (D_3). ☐

Lemma 5.4 is not applicable if $\mathscr{D}$ is given in terms of a generator $\mathscr{G}$, which is often the case. The next theorem is very important for applications as it extends Lemma 5.4 to the much more convenient setting of generators.

Theorem 5.5 *If $\mathscr{G} \subset \mathscr{P}(X)$ is stable under finite intersections, then $\delta(\mathscr{G}) = \sigma(\mathscr{G})$.*

Proof We have already established $\delta(\mathscr{G}) \subset \sigma(\mathscr{G})$ in Proposition 5.3. If we knew that $\delta(\mathscr{G})$ was a σ-algebra, the minimality of $\sigma(\mathscr{G})$ and $\mathscr{G} \subset \delta(\mathscr{G})$ would immediately imply $\sigma(\mathscr{G}) \subset \delta(\mathscr{G})$, and hence equality.

In view of Lemma 5.4 it will suffice to show that $\delta(\mathscr{G})$ is $\cap$-stable. For this we fix some $D \in \delta(\mathscr{G})$ and introduce the family

$$\mathscr{D}_D := \{Q \subset X : Q \cap D \in \delta(\mathscr{G})\}.$$

Let us check that $\mathscr{D}_D$ is a Dynkin system. (D_1) is obviously true. (D_2): take $Q \in \mathscr{D}_D$. Then

$$Q^c \cap D = (Q^c \cup D^c) \cap D = (Q \cap D)^c \cap D = (\underbrace{(Q \cap D)}_{\in\delta(\mathscr{G})} \uplus \underbrace{D^c}_{\in\delta(\mathscr{G})})^c \tag{5.1}$$

and disjoint unions of sets from $\delta(\mathscr{G})$ are still in $\delta(\mathscr{G})$. Thus $Q^c \in \mathscr{D}_D$. (D_3): let $(Q_n)_{n\in\mathbb{N}}$ be a sequence of pairwise disjoint sets from $\mathscr{D}_D$. Then $(Q_n \cap D)_{n\in\mathbb{N}}$ is a disjoint sequence in $\delta(\mathscr{G})$ and (D_3) for the Dynkin system $\delta(\mathscr{G})$ shows

$$\left(\biguplus_{n\in\mathbb{N}} Q_n\right) \cap D = \biguplus_{n\in\mathbb{N}} (Q_n \cap D) \in \delta(\mathscr{G}),$$

[1] $\cap$-stable, for short.

which means that $\dot{\bigcup}_{n\in\mathbb{N}} Q_n \in \mathscr{D}_D$.

Since $\mathscr{G} \subset \delta(\mathscr{G})$ and since $\mathscr{G}$ is $\cap$-stable, we have

$$\begin{aligned}
& \mathscr{G} \subset \mathscr{D}_G && \forall G \in \mathscr{G} \\
\Longrightarrow\quad & \delta(\mathscr{G}) \subset \mathscr{D}_G && \forall G \in \mathscr{G} \ \text{ (since } \mathscr{D}_G \text{ is a Dynkin system)} \\
\Longrightarrow\quad & G \cap D \in \delta(\mathscr{G}) && \forall G \in \mathscr{G},\ \forall D \in \delta(\mathscr{G}) \ \text{ (by the definition of } \mathscr{D}_G) \\
\Longrightarrow\quad & G \in \mathscr{D}_D && \forall G \in \mathscr{G},\ \forall D \in \delta(\mathscr{G}) \\
\Longrightarrow\quad & \mathscr{G} \subset \mathscr{D}_D && \forall D \in \delta(\mathscr{G}) \\
\Longrightarrow\quad & \delta(\mathscr{G}) \subset \mathscr{D}_D && \forall D \in \delta(\mathscr{G}) \text{ (since } \mathscr{D}_D \text{ is a Dynkin system).}
\end{aligned}$$

The latter just says that $\delta(\mathscr{G})$ is stable under intersections with $D \in \delta(\mathscr{G})$. By Lemma 5.4 $\delta(\mathscr{G})$ is a σ-algebra, and the theorem is proved. □

Remark 5.6 The technique used in the proof of Theorem 5.5 is an extremely important and powerful tool. It is the analogue of 'transfinite induction' as it allows us to reduce assertions to generators rather than checking them on the (usually much larger) σ-algebra itself. We will use it almost exclusively in this chapter to prove the uniqueness of measures theorem and some properties of Lebesgue measure λ^n.

Theorem 5.7 (uniqueness of measures) *Let $(X, \mathscr{A})$ be a measurable space and assume that $\mathscr{A} = \sigma(\mathscr{G})$ is generated by a family $\mathscr{G}$ such that*

(i) *$\mathscr{G}$ is stable under finite intersections: $G, H \in \mathscr{G} \Longrightarrow G \cap H \in \mathscr{G}$;*
(ii) *there exists an exhausting sequence $(G_n)_{n\in\mathbb{N}} \subset \mathscr{G}$ with $G_n \uparrow X$.*

Any two measures μ, ν that coincide on $\mathscr{G}$ and are finite for all members of the exhausting sequence $\mu(G_n) = \nu(G_n) < \infty$, are equal on $\mathscr{A}$, i.e. $\mu(A) = \nu(A)$ for all $A \in \mathscr{A}$.

If μ and ν are probability measures, assumption (ii) in Theorem 5.7 may be omitted. To see this, adjoin X to $\mathscr{G}$ and pick the trivial exhausting sequence $G_n = X$.

Proof of Theorem 5.7 For $n \in \mathbb{N}$ we define

$$\mathscr{D}_n := \left\{ A \in \mathscr{A} : \mu(G_n \cap A) = \nu(G_n \cap A) \quad (< \infty) \right\}$$

and we claim that every $\mathscr{D}_n$ is a Dynkin system. (D_1) is clear. (D_2): if $A \in \mathscr{D}_n$ we have

$$\begin{aligned}
\mu(G_n \cap A^c) = \mu(G_n \setminus A) &= \mu(G_n) - \mu(G_n \cap A) \\
&= \nu(G_n) - \nu(G_n \cap A) \\
&= \nu(G_n \setminus A) = \nu(G_n \cap A^c),
\end{aligned}$$

so that $A^c \in \mathscr{D}_n$. (D_3): if $(A_k)_{k\in\mathbb{N}} \subset \mathscr{D}_n$ are mutually disjoint sets, we get

$$\begin{aligned}\mu\Big(G_n \cap \dot{\bigcup_{k\in\mathbb{N}}} A_k\Big) &= \mu\Big(\dot{\bigcup_{k\in\mathbb{N}}}(G_n \cap A_k)\Big) = \sum_{k\in\mathbb{N}} \mu(G_n \cap A_k)\\ &= \sum_{k\in\mathbb{N}} \nu(G_n \cap A_k) = \nu\Big(\dot{\bigcup_{k\in\mathbb{N}}}(G_n \cap A_k)\Big)\\ &= \nu\Big(G_n \cap \dot{\bigcup_{k\in\mathbb{N}}} A_k\Big),\end{aligned}$$

and $\dot{\bigcup}_{k\in\mathbb{N}} A_k \in \mathscr{D}_n$ follows.

Since $\mathscr{G}$ is $\cap$-stable, we know from Theorem 5.5 that $\delta(\mathscr{G}) = \sigma(\mathscr{G})$; therefore,

$$\mathscr{D}_n \supset \mathscr{G} \implies \mathscr{D}_n \supset \delta(\mathscr{G}) = \sigma(\mathscr{G}) \qquad \forall n \in \mathbb{N}.$$

On the other hand, $\mathscr{A} = \sigma(\mathscr{G}) \subset \mathscr{D}_n \subset \mathscr{A}$, which means that $\mathscr{A} = \mathscr{D}_n$ for every $n \in \mathbb{N}$, and so

$$\mu(G_n \cap A) = \nu(G_n \cap A) \quad \forall n \in \mathbb{N}, \quad \forall A \in \mathscr{A}. \tag{5.2}$$

Using Proposition 4.3(vi) we can let $n \to \infty$ in (5.2) and we get

$$\mu(A) = \lim_{n\to\infty} \mu(G_n \cap A) = \lim_{n\to\infty} \nu(G_n \cap A) = \nu(A) \quad \forall A \in \mathscr{A}. \qquad \square$$

The following two theorems show why Lebesgue measure (if it exists ...) plays a very special rôle indeed.

Theorem 5.8 (i) *The n-dimensional Lebesgue measure λ^n is invariant under translations, i.e.*

$$\lambda^n(x + B) = \lambda^n(B) \quad \forall x \in \mathbb{R}^n,\ \forall B \in \mathscr{B}(\mathbb{R}^n). \tag{5.3}$$

(ii) *Every measure μ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ which is invariant under translations and satisfies $\kappa = \mu([0,1)^n) < \infty$ is a multiple of Lebesgue measure: $\mu = \kappa\lambda^n$.*

Proof First of all we should convince ourselves that

$$B \in \mathscr{B}(\mathbb{R}^n) \implies x + B \in \mathscr{B}(\mathbb{R}^n) \quad \forall x \in \mathbb{R}^n \tag{5.4}$$

– otherwise the statement of Theorem 5.8 would be senseless. For this set

$$\mathscr{A}_x := \big\{B \in \mathscr{B}(\mathbb{R}^n) : x + B \in \mathscr{B}(\mathbb{R}^n)\big\} \subset \mathscr{B}(\mathbb{R}^n).$$

It is clear that $\mathscr{A}_x$ is a σ-algebra and that the rectangles $\mathscr{J} \subset \mathscr{A}_x$. [✍] Hence, $\mathscr{B}(\mathbb{R}^n) = \sigma(\mathscr{J}) \subset \mathscr{A}_x \subset \mathscr{B}(\mathbb{R}^n)$ and (5.4) follows. We can now start the proof proper.

(i) Set $\nu(B) := \lambda^n(x+B)$ for some fixed $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. It is easy to check that ν is a measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. [✍]

Take $I = [a_1, b_1) \times \cdots \times [a_n, b_n) \in \mathscr{J}$ and observe that

$$x + I = [a_1 + x_1, b_1 + x_1) \times \cdots \times [a_n + x_n, b_n + x_n) \in \mathscr{J},$$

so that

$$\nu(I) = \lambda^n(x+I) = \prod_{i=1}^n \big((b_i + x_i) - (a_i + x_i)\big) = \prod_{i=1}^n \big(b_i - a_i\big) = \lambda^n(I).$$

This means that $\nu|_{\mathscr{J}} = \lambda^n|_{\mathscr{J}}$.[2] But $\mathscr{J}$ is $\cap$-stable,[3] generates $\mathscr{B}(\mathbb{R}^n)$ and admits the exhausting sequence

$$[-k,k)^n \uparrow \mathbb{R}^n, \quad \lambda^n\big([-k,k)^n\big) = (2k)^n < \infty.$$

We can now invoke Theorem 5.7 to see that $\lambda^n = \nu$ on the whole of $\mathscr{B}(\mathbb{R}^n)$.

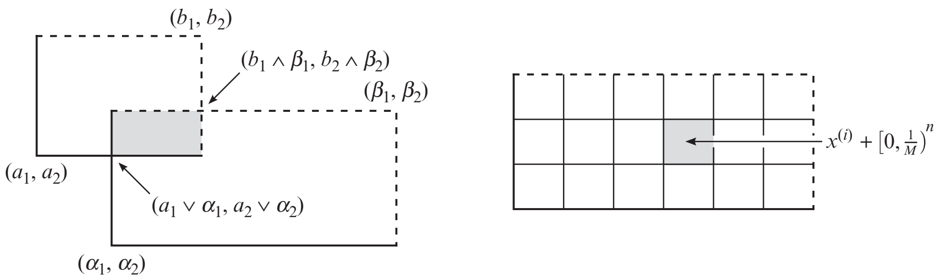


Fig. 5.1. *Left:* We have $\times_{i=1}^n [a_i, b_i) \cap \times_{i=1}^n [\alpha_i, \beta_i) = \times_{i=1}^n [\max\{a_i, \alpha_i\}, \min\{b_i, \beta_i\})$. *Right:* We can pave I with squares of the form $x^{(i)} + \big[0, \frac{1}{M}\big)^n$ having side-length $1/M$ and lower left corner $x^{(i)}$.

(ii) Take $I \in \mathscr{J}$ as in part (i) but with rational endpoints $a_i, b_i \in \mathbb{Q}$. Thus there is some $M \in \mathbb{N}$ and $k(I) \in \mathbb{N}$ and points $x^{(i)} \in \mathbb{R}^n$, such that

$$I = \bigcupdot_{i=1}^{k(I)} \big(x^{(i)} + \big[0, \tfrac{1}{M}\big)^n\big)$$

– i.e. we pave the rectangle I by little squares $\big[0, \frac{1}{M}\big)^n$ of side-length $1/M$ (Fig. 5.1), where M is, say, the common denominator of all a_i and b_i. Using

[2] This is short for $\nu(I) = \lambda^n(I) \quad \forall I \in \mathscr{J}$.

[3] Use $\times_{i=1}^n [a_i, b_i) \cap \times_{i=1}^n [a_i', b_i') = \times_{i=1}^n [a_i \vee a_i', b_i \wedge b_i')$. [✍]

the translation invariance of μ and λ^n, we see that

$$\mu(I) = k(I)\mu\big(\big[0, \tfrac{1}{M}\big)^n\big), \qquad \mu\big([0,1)^n\big) = M^n\mu\big(\big[0, \tfrac{1}{M}\big)^n\big),$$
$$\lambda^n(I) = k(I)\lambda^n\big(\big[0, \tfrac{1}{M}\big)^n\big), \qquad \underbrace{\lambda^n\big([0,1)^n\big)}_{=1} = M^n\lambda^n\big(\big[0, \tfrac{1}{M}\big)^n\big),$$

and dividing the top two and bottom two equalities gives

$$\mu(I) = \frac{k(I)}{M^n}\mu\big([0,1)^n\big), \qquad \lambda^n(I) = \frac{k(I)}{M^n}\lambda^n\big([0,1)^n\big) = \frac{k(I)}{M^n}.$$

Thus $\mu(I) = \mu\big([0,1)^n\big)\lambda^n(I) = \kappa\lambda^n(I)$ for all $I \in \mathscr{J}$ and, as in part (i), an application of Theorem 5.5 finishes the proof. □

Incidentally, Theorem 5.8 proves Theorem 4.7(i). Further properties of Lebesgue measure will be studied in the following chapters, but first we concentrate on its existence.

Problems

5.1. Verify the claims made in Remark 5.2.

5.2. The following exercise shows that Dynkin systems and σ-algebras are, in general, different. Let $X = \{1, 2, 3, \dots, 4k-1, 4k\}$ for some $k \in \mathbb{N}$. Then $\mathscr{D} = \{A \subset X : \#A \text{ is even}\}$ is a Dynkin system, but not a σ-algebra.

5.3. Let $\mathscr{D}$ be a Dynkin system. Show that for all $A, B \in \mathscr{D}$ with $A \subset B$ the difference $B \setminus A \in \mathscr{D}$. [Hint: use $R \setminus Q = ((R \cap Q) \cup R^c)^c$, where $R, Q \subset X$.]

5.4. Let $\mathscr{A}$ be a σ-algebra, $\mathscr{D}$ be a Dynkin system and $\mathscr{G} \subset \mathscr{H} \subset \mathscr{P}(X)$ two collections of subsets of X. Show that

(i) $\delta(\mathscr{A}) = \mathscr{A}$ and $\delta(\mathscr{D}) = \mathscr{D}$;
(ii) $\delta(\mathscr{G}) \subset \delta(\mathscr{H})$;
(iii) $\delta(\mathscr{G}) \subset \sigma(\mathscr{G})$.

5.5. Let $A, B \subset X$. Compare $\delta(\{A, B\})$ and $\sigma(\{A, B\})$. When are they equal?

5.6. **An alternative definition of Dynkin systems**. A family $\mathscr{F} \subset \mathscr{P}(X)$ is a Dynkin system if, and only if,

$$X \in \mathscr{F}, \tag{D$_1$}$$
$$F, G \in \mathscr{F},\ F \subset G \implies G \setminus F \in \mathscr{F}, \tag{D$_2'$}$$
$$(F_n)_{n\in\mathbb{N}} \subset \mathscr{F},\ F_n \uparrow F \implies F = \bigcup_{n\in\mathbb{N}} F_n \in \mathscr{F}. \tag{D$_3'$}$$

Conclude that any Dynkin system is a monotone class in the sense of Problem 3.14.

5.7. Show that Theorem 5.7 is still valid, if $(G_n)_{n\in\mathbb{N}} \subset \mathscr{G}$ is not an increasing sequence but *any* countable family of sets such that

$$(1) \quad \bigcup_{n\in\mathbb{N}} G_n = X \quad \text{and} \quad (2) \quad \nu(G_n) = \mu(G_n) < \infty.$$

[Hint: set $F_N := G_1 \cup \dots \cup G_N = F_{N-1} \cup G_N$ and check by induction that $\mu(F_N) = \nu(F_N)$; use then Theorem 5.7.]

5.8. Show that the half-open intervals $\mathcal{J}$ in $\mathbb{R}^n$ are stable under finite intersections.
[Hint: check that $\times_{i=1}^n [a_i, b_i) \cap \times_{i=1}^n [a_i', b_i') = \times_{i=1}^n [a_i \vee a_i', b_i \wedge b_i')$.]

5.9. Dilations. Mimic the proof of Theorem 5.8(i) and show that $t \cdot B := \{tb : b \in B\}$ is a Borel set for all $B \in \mathscr{B}(\mathbb{R}^n)$ and $t > 0$. Moreover,

$$\lambda^n(t \cdot B) = t^n \lambda^n(B) \qquad \forall B \in \mathscr{B}(\mathbb{R}^n),\ \forall t > 0. \tag{5.5}$$

5.10. Invariant measures. Let $(X, \mathscr{A}, \mu)$ be a finite measure space where $\mathscr{A} = \sigma(\mathscr{G})$ for some $\cap$-stable generator $\mathscr{G}$. Assume that $\theta : X \to X$ is a map such that $\theta^{-1}(A) \in \mathscr{A}$ for all $A \in \mathscr{A}$. Prove that

$$\mu(G) = \mu(\theta^{-1}(G)) \quad \forall G \in \mathscr{G} \implies \mu(A) = \mu(\theta^{-1}(A)) \quad \forall A \in \mathscr{A}.$$

(A measure μ with this property is called *invariant* w.r.t. the map θ.)

5.11. Independence (1). Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and let $\mathscr{B}, \mathscr{C} \subset \mathscr{A}$ be two sub-σ-algebras of $\mathscr{A}$. We call $\mathscr{B}$ and $\mathscr{C}$ *independent*, if

$$\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) \quad \forall B \in \mathscr{B},\ C \in \mathscr{C}.$$

Assume now that $\mathscr{B} = \sigma(\mathscr{G})$ and $\mathscr{C} = \sigma(\mathcal{H})$, where $\mathscr{G}$, $\mathcal{H}$ are $\cap$-stable collections of sets. Prove that $\mathscr{B}$ and $\mathscr{C}$ are independent if, and only if,

$$\mathbb{P}(G \cap H) = \mathbb{P}(G)\mathbb{P}(H) \quad \forall G \in \mathscr{G},\ H \in \mathcal{H}.$$

5.12. Approximation of σ-algebras. Let $\mathscr{G}$ be a *Boolean algebra* in X, i.e. a family of sets such that $X \in \mathscr{G}$ and $\mathscr{G}$ is stable under the formation of finite unions, intersections and complements. Let $\mathscr{A} = \sigma(\mathscr{G})$ and μ be a finite measure on $(X, \mathscr{A}, \mu)$. Let $A \triangle B := (A \setminus B) \cup (B \setminus A)$ denote the symmetric difference of $A, B \subset X$.

(i) For every $\epsilon > 0$ and $A \in \mathscr{A}$ there is some $G \in \mathscr{G}$ such that $\mu(A \triangle G) \leqslant \epsilon$.
[Hint: $\{A \in \mathscr{A} : \forall \epsilon > 0\, \exists G \in \mathscr{G} : \mu(A \triangle G) \leqslant \epsilon\}$ is a Dynkin system.]

(ii) Let μ, ν be finite measures on $(X, \mathscr{A})$. For every $\epsilon > 0$ and $A \in \mathscr{A}$ there is some $G \in \mathscr{G}$ such that $\mu(A \triangle G) \leqslant \epsilon$ and $\nu(A \triangle G) \leqslant \epsilon$.

(iii) Assume that $X = \mathbb{R}^n$, $\mathscr{A} = \mathscr{B}(\mathbb{R}^n)$ and $\mu = \lambda^n$. A set $A \in \mathscr{A}$ satisfies $\mu(A) = 0$ if, and only if, for every $\epsilon > 0$ there is a sequence $(I_n)_{n \in \mathbb{N}} \subset \mathcal{J}$, such that $A \subset \bigcup_n I_n$ and $\mu\left(\bigcup_n I_n\right) \leqslant \epsilon$.

Remark. Compare the result of this exercise with Theorem 27.25.

5.13. Monotone classes (2). Recall from Problem 3.14 that a monotone class $\mathscr{M} \subset \mathscr{P}(X)$ is a family which contains X and is stable under countable unions of increasing sets and countable intersections of decreasing sets; denote by $\mathfrak{m}(\mathscr{G})$ the smallest monotone class containing the family $\mathscr{G} \subset \mathscr{P}(X)$. Show the following analogues of Lemma 5.4 and Theorem 5.5.

(i) $\mathscr{M}$ is a σ-algebra if, and only if, it is stable under the formation of complements.

(ii) If $\mathscr{G}$ is stable under the formation of complements and finite intersections, then $\mathscr{M} = \mathfrak{m}(\mathscr{G})$ is a σ-algebra (namely $\sigma(\mathscr{G})$).
[Hint: use the monotone class theorem from Problem 3.14.]

6

Existence of Measures

In Chapter 4 we saw that it is not a trivial task to assign *explicitly* a μ-value to *every* set A from a σ-algebra $\mathscr{A}$. Rather than doing this it is often more natural to assign μ-values to, say, rectangles (in the case of the Borel σ-algebra) or, in general, to sets from some generator $\mathscr{G}$ of $\mathscr{A}$. Because of Proposition 4.3 (and Remark 4.4) $\mu|_{\mathscr{G}}$ should be a pre-measure. If $\mathscr{G}$ and μ satisfy the conditions of the uniqueness theorem (Theorem 5.7), this approach will lead to a unique measure on $\mathscr{A}$, provided that we can extend μ from $\mathscr{G}$ onto $\sigma(\mathscr{G}) = \mathscr{A}$.

To get such an automatic extension the following (technically motivated) class of generators is useful. A *semi-ring* is a family $\mathscr{S} \subset \mathscr{P}(X)$ with the following properties:

$$\emptyset \in \mathscr{S}, \tag{S$_1$}$$

$$S, T \in \mathscr{S} \implies S \cap T \in \mathscr{S}, \tag{S$_2$}$$

$$\begin{array}{c}\text{for } S, T \in \mathscr{S} \text{ there exist finitely many disjoint} \\ S_1, S_2, \ldots, S_M \in \mathscr{S} \text{ such that } S \setminus T = \displaystyle\bigcup_{i=1}^{M}{}^{\bullet}\, S_i.\end{array} \tag{S$_3$}$$

The solution to our problems is the following deep extension theorem for measures which goes back to Carathéodory [8].

Theorem 6.1 (Carathéodory) *Let $\mathscr{S} \subset \mathscr{P}(X)$ be a semi-ring and $\mu : \mathscr{S} \to [0, \infty]$ a pre-measure, i.e. a set function with*

(i) $\mu(\emptyset) = 0$;

(ii) $(S_n)_{n\in\mathbb{N}} \subset \mathscr{S}$, *disjoint and* $S = \dot{\bigcup}_{n\in\mathbb{N}} S_n \in \mathscr{S} \implies \mu(S) = \sum_{n\in\mathbb{N}} \mu(S_n)$.

Then μ has an extension to a measure μ on $\sigma(\mathcal{S})$. If, moreover, $\mathcal{S}$ contains an exhausting sequence $(S_n)_{n\in\mathbb{N}}$, $S_n \uparrow X$ such that $\mu(S_n) < \infty$ for all $n \in \mathbb{N}$, then the extension is unique.

Remark 6.2 From the definition of a measure (Definition 4.1) it is clear that the conditions 6.1(i) and (ii) are necessary for μ to become a measure. Theorem 6.1 says that they are even sufficient. Remarkable is the fact that (ii) is needed only relative to $\mathcal{S}$ – its extension to $\sigma(\mathcal{S})$ is then automatic.

Idea of the proof of Theorem 6.1 The fundamental problem is how to extend the pre-measure μ. The following auxiliary set-function $\mu^* : \mathscr{P}(X) \to [0, \infty]$ will play a central role. Define for each $A \subset X$ the family of countable $\mathcal{S}$-coverings

$$\mathscr{C}(A) := \Big\{ (S_n)_{n\in\mathbb{N}} \subset \mathcal{S} : \underbrace{\bigcup_{n\in\mathbb{N}} S_n}_{\text{need not be disjoint or in } \mathcal{S}} \supset A \Big\}$$

and the set function

$$\mu^*(A) := \inf \Big\{ \sum_{n\in\mathbb{N}} \mu(S_n) : (S_n)_{n\in\mathbb{N}} \in \mathscr{C}(A) \Big\}. \tag{6.1}$$

If A cannot be covered by sets from $\mathcal{S}$, we have $\mathscr{C}(A) = \emptyset$ and $\mu^*(A) = \inf \emptyset = \infty$.

Step 1. Verify that μ^* is an *outer measure*, i.e. a map $\mu^* : \mathscr{P}(X) \to [0, \infty]$ with the following properties:

$$\mu^*(\emptyset) = 0, \tag{OM$_1$}$$

$$A \subset B \implies \mu^*(A) \leqslant \mu^*(B), \tag{OM$_2$}$$

$$\mu^*\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \leqslant \sum_{n\in\mathbb{N}} \mu^*(A_n). \tag{OM$_3$}$$

Step 2. Show that μ^* extends μ, i.e. $\mu^*(S) = \mu(S)$ for all $S \in \mathcal{S}$.

Step 3. Define the μ^*-measurable sets

$$\mathscr{A}^* := \{ A \subset X : \mu^*(Q) = \mu^*(Q \cap A) + \mu^*(Q \setminus A) \quad \forall Q \subset X \} \tag{6.2}$$

and show that $\mathscr{A}^*$ is a σ-algebra with $\mathscr{A}^* \supset \mathcal{S}$ and $\mathscr{A}^* \supset \sigma(\mathcal{S})$.

Step 4. Show that $\mu^*|_{\mathscr{A}^*}$ is a measure. In particular, $\mu^*|_{\sigma(\mathcal{S})}$ is a measure which extends μ. $\square$

The proof of Carathéodory's theorem is a bit involved and not particularly rewarding when read superficially. Therefore we recommend skipping the proof on first reading and resuming on p. 46.

Proof of Theorem 6.1 We follow the steps outlined above but, for clarity, we make further subdivisions.

Step 1. Claim: μ^* is an outer measure. (OM_1) is obvious since we can take in (6.1) the constant sequence $S_1 = S_2 = \cdots = \emptyset$ which is clearly in $\mathscr{C}(\emptyset)$.

(OM_2): if $B \supset A$, then each $\mathcal{S}$-cover of B also covers A, i.e. $\mathscr{C}(B) \subset \mathscr{C}(A)$. Therefore,

$$\begin{aligned}\mu^*(A) &= \inf\left\{\sum_{n\in\mathbb{N}} \mu(S_n) : (S_n)_{n\in\mathbb{N}} \in \mathscr{C}(A)\right\}\\ &\leqslant \inf\left\{\sum_{k\in\mathbb{N}} \mu(T_k) : (T_k)_{k\in\mathbb{N}} \in \mathscr{C}(B)\right\} = \mu^*(B).\end{aligned}$$

(OM_3): without loss of generality we can assume that $\mu^*(A_n) < \infty$ for all $n \in \mathbb{N}$ and so $\mathscr{C}(A_n) \neq \emptyset$. Fix $\epsilon > 0$ and observe that by the very nature of the infimum we find for each A_n a cover $(S_k{}^n)_{k\in\mathbb{N}} \in \mathscr{C}(A_n)$ with

$$\sum_{k\in\mathbb{N}} \mu(S_k^n) \leqslant \mu^*(A_n) + \frac{\epsilon}{2^n}, \quad n \in \mathbb{N}. \tag{6.3}$$

The double sequence $(S_k{}^n)_{k,n\in\mathbb{N}}$ is an $\mathcal{S}$-cover of $A := \bigcup_{n\in\mathbb{N}} A_n$, and so

$$\begin{aligned}\mu^*(A) \leqslant \sum_{(k,n)\in\mathbb{N}\times\mathbb{N}} \mu(S_k{}^n) &= \sum_{n\in\mathbb{N}}\sum_{k\in\mathbb{N}} \mu(S_k^n)\\ &\overset{(6.3)}{\leqslant} \sum_{n\in\mathbb{N}} \left(\mu^*(A_n) + \frac{\epsilon}{2^n}\right)\\ &= \sum_{n\in\mathbb{N}} \mu^*(A_n) + \epsilon,\end{aligned}$$

where the second '$\leqslant$' follows from (6.3). Letting $\epsilon \to 0$ proves (OM_3).

Step 2a. Extension of μ to the family $\mathcal{S}_\cup := \{S_1 \uplus \cdots \uplus S_M : M \in \mathbb{N},\ S_n \in \mathcal{S}\}$. Define

$$\bar{\mu}(S_1 \uplus \cdots \uplus S_M) := \sum_{n=1}^{M} \mu(S_n). \tag{6.4}$$

Since (6.4) is *necessary* for an additive set function on $\mathcal{S}_\cup$, (6.4) implies the uniqueness of the extension [✍] once we know that $\bar{\mu}$ is well-defined – that is, independent of the particular representation of sets in $\mathcal{S}_\cup$. To see this assume that

$$S_1 \uplus \cdots \uplus S_M = T_1 \uplus \cdots \uplus T_N, \qquad M, N \in \mathbb{N}, \quad S_i, T_k \in \mathcal{S}.$$

Then

$$S_i = S_i \cap (T_1 \uplus \cdots \uplus T_N) = \biguplus_{k=1}^{N} (S_i \cap T_k),$$

and the additivity of μ on $\mathcal{S}$ shows that

$$\mu(S_i) = \sum_{k=1}^{N} \mu(S_i \cap T_k).$$

Summing over $i = 1, 2, \ldots, M$ and swapping the rôles of S_i and T_k gives

$$\sum_{i=1}^{M} \mu(S_i) = \sum_{i=1}^{M} \sum_{k=1}^{N} \mu(S_i \cap T_k) = \sum_{k=1}^{N} \mu(T_k),$$

which proves that (6.4) does not depend on the representation of $\mathcal{S}_\cup$-sets.

Step 2b. Extension of μ to finite unions of not necessarily disjoint $\mathcal{S}$-sets. The family $\mathcal{S}_\cup$ is clearly stable under finite *disjoint* unions. If $S, T \in \mathcal{S}_\cup$ we find (notation as before)

$$S \cap T = (S_1 \uplus \cdots \uplus S_M) \cap (T_1 \uplus \cdots \uplus T_N) = \biguplus_{i=1}^{M} \biguplus_{k=1}^{N} \underbrace{(S_i \cap T_k)}_{\in \mathcal{S}} \in \mathcal{S}_\cup.$$

Because of (S_3), we know that $S_i \setminus T_k \in \mathcal{S}_\cup$, and so

$$\begin{aligned} S \setminus T &= (S_1 \uplus \cdots \uplus S_M) \setminus (T_1 \uplus \cdots \uplus T_N) \\ &= \biguplus_{i=1}^{M} \bigcap_{k=1}^{N} \left(S_i \cap T_k^c\right) = \biguplus_{i=1}^{M} \underbrace{\bigcap_{k=1}^{N} \underbrace{S_i \setminus T_k}_{\in \mathcal{S}_\cup}}_{\in \mathcal{S}_\cup} \in \mathcal{S}_\cup, \end{aligned}$$

where we use the $\cap$- and $\uplus$-stability of $\mathcal{S}_\cup$. Finally,[1]

$$S \cup T = \big(S \setminus T\big) \uplus \big(S \cap T\big) \uplus \big(T \setminus S\big) \in \mathcal{S}_\cup,$$

and the prescription (6.4) can be used to extend μ to finite unions of $\mathcal{S}$-sets.

[1] this shows that $\mathcal{S}_\cup$ is the ring generated by $\mathcal{S}$, i.e. the smallest ring containing $\mathcal{S}$.

Step 2c. Claim: $\bar{\mu}$ is a pre-measure on $\mathscr{S}_{\cup}$. We have to show that $\bar{\mu}$ is σ-additive on $\mathscr{S}_{\cup}$. For this take $(T_k)_{k\in\mathbb{N}} \subset \mathscr{S}_{\cup}$ such that $T := \dot\bigcup_{k\in\mathbb{N}} T_k \in \mathscr{S}_{\cup}$. By the definition of the family $\mathscr{S}_{\cup}$ we find a sequence of disjoint sets $(S_n)_{n\in\mathbb{N}} \subset \mathscr{S}$ and a sequence of integers $0 = i(0) \leqslant i(1) \leqslant i(2) \leqslant \cdots$ such that

$$T_k = S_{i(k-1)+1} \dot\cup \cdots \dot\cup S_{i(k)}, \quad k \in \mathbb{N},$$

and $T = U_1 \dot\cup \cdots \dot\cup U_L$, where $U_\ell = \dot\bigcup_{i\in J_\ell} S_i \in \mathscr{S}$ [✍] with disjoint index sets $J_1 \dot\cup J_2 \dot\cup \cdots \dot\cup J_L = \mathbb{N}$ partitioning $\mathbb{N}$. Since μ is σ-additive on $\mathscr{S}$,

$$\bar{\mu}(T) \overset{\text{def}}{=} \sum_{\ell=1}^{L} \mu(U_\ell) \overset{6.1(\text{ii})}{=} \sum_{\ell=1}^{L} \sum_{i\in J_\ell} \mu(S_i) = \sum_{k\in\mathbb{N}} \sum_{n=i(k-1)+1}^{i(k)} \mu(S_n) \overset{\text{def}}{=} \sum_{k\in\mathbb{N}} \bar{\mu}(T_k),$$

which proves σ-additivity of $\bar{\mu}$.

Step 2d. μ^* extends μ. Using the pre-measure $\bar{\mu}$ we get from Proposition 4.3(viii) for any cover $(S_n)_{n\in\mathbb{N}} \in \mathscr{C}(S)$, $S \in \mathscr{S}$, that

$$\begin{aligned}\mu(S) = \bar{\mu}(S) = \bar{\mu}\Bigg(\bigcup_{n\in\mathbb{N}} S_n \cap S\Bigg) &\leqslant \sum_{n\in\mathbb{N}} \bar{\mu}(S_n \cap S) \\ &= \sum_{n\in\mathbb{N}} \mu(S_n \cap S) \leqslant \sum_{n\in\mathbb{N}} \mu(S_n),\end{aligned}$$

and passing to the infimum over $\mathscr{C}(S)$ shows that $\mu(S) \leqslant \mu^*(S)$. The special cover $(S, \emptyset, \emptyset, \ldots) \in \mathscr{C}(S)$, on the other hand, yields $\mu^*(S) \leqslant \mu(S)$ and this shows that $\mu|_{\mathscr{S}} = \mu^*|_{\mathscr{S}}$.

Step 3a. Claim: $\mathscr{S} \subset \mathscr{A}^*$. Let $S, T \in \mathscr{S}$. From (S_3) we get

$$T = (S \cap T) \dot\cup (T \setminus S) = (S \cap T) \dot\cup \dot\bigcup_{i=1}^{M} S_i$$

for some mutually disjoint sets $S_i \in \mathscr{S}$, $i = 1, 2, \ldots, M$. Since μ is additive on $\mathscr{S}$ and μ^* is (σ-)subadditive by (OM_3), we find

$$\mu^*(S \cap T) + \mu^*(T \setminus S) \leqslant \mu(S \cap T) + \sum_{i=1}^{M} \mu(S_n) = \mu(T). \tag{6.5}$$

Take any $B \subset X$ and some $\mathscr{S}$-cover $(T_n)_{n\in\mathbb{N}} \in \mathscr{C}(B)$. Using $\mu^*(T_n) = \mu(T_n)$ and summing the inequality (6.5) for $T = T_n$ over $n \in \mathbb{N}$ yields

$$\sum_{n\in\mathbb{N}} \mu^*(T_n \setminus S) + \sum_{n\in\mathbb{N}} \mu^*(T_n \cap S) \leqslant \sum_{n\in\mathbb{N}} \mu^*(T_n),$$

and the σ-subadditivity (OM_3) and monotonicity (OM_2) of μ^* give (recall that $B \subset \bigcup_{n\in\mathbb{N}} T_n$)

$$\begin{aligned}\mu^*(B\setminus S)+\mu^*(B\cap S) &\leqslant \mu^*\Big(\bigcup_{n\in\mathbb{N}} T_n\setminus S\Big)+\mu^*\Big(\bigcup_{n\in\mathbb{N}} T_n\cap S\Big)\\ &\leqslant \sum_{n\in\mathbb{N}}\mu^*(T_n) \;=\; \sum_{n\in\mathbb{N}}\mu(T_n).\end{aligned}$$

We can now pass to the inf over $\mathscr{C}(B)$ and find $\mu^*(B\setminus S)+\mu^*(B\cap S)\leqslant\mu^*(B)$. Since the reverse inequality follows easily from the (σ-)subadditivity (OM_3) of μ^*, $S\in\mathscr{A}^*$ holds for all $S\in\mathcal{S}$.

Step 3b. Claim: $\mathscr{A}^*$ is a σ-algebra and μ^* is a measure on $\mathscr{A}^*$.

(Σ_1): clearly, $\emptyset\in\mathscr{A}^*$.

(Σ_2): the symmetry (w.r.t. A and A^c) of the definition of $\mathscr{A}^*$ shows that $A\in\mathscr{A}^*$ if, and only if, $A^c\in\mathscr{A}^*$, see (6.2).

(Σ_3): let us first show that $\mathscr{A}^*$ is $\cup$-stable. Using the (σ-)subadditivity (OM_3) of μ^* we find for $A, A'\in\mathscr{A}^*$ and any $P\subset X$

$$\begin{aligned}&\mu^*(P\cap(A\cup A'))+\mu^*(P\setminus(A\cup A'))\\ &\qquad= \mu^*(P\cap(A\cup[A'\setminus A]))+\mu^*(P\setminus(A\cup A'))\\ &\qquad\leqslant \mu^*(P\cap A)+\mu^*(P\cap(A'\setminus A))+\mu^*(P\setminus(A\cup A'))\\ &\qquad= \mu^*(P\cap A)+\mu^*((P\setminus A)\cap A')+\mu^*((P\setminus A)\setminus A')\\ &\qquad\overset{(6.2)}{=} \mu^*(P\cap A)+\mu^*(P\setminus A) \qquad (6.6)\\ &\qquad\overset{(6.2)}{=} \mu^*(P), \qquad (6.6')\end{aligned}$$

where we use in the last two steps the definition (6.2) of $\mathscr{A}^*$ with $Q\mathrel{\widehat{=}}P\setminus A$ and $Q\mathrel{\widehat{=}}P$, respectively. The reverse inequality follows from (OM_3), and we conclude that $A\cup A'\in\mathscr{A}^*$.

If $A\cap A'=\emptyset$, the equality '(6.6) = (6.6′)' becomes, for $P:=(A\uplus A')\cap Q$ and any $Q\subset X$,

$$\mu^*(Q\cap(A\uplus A'))=\mu^*(Q\cap A)+\mu^*(Q\cap A')\quad\forall Q\subset X.$$

Iterating this argument yields

$$\mu^*(Q\cap(A_1\uplus\cdots\uplus A_M))=\sum_{i=1}^{M}\mu^*(Q\cap A_i)\quad\forall Q\subset X \tag{6.7}$$

for all mutually disjoint $A_1, A_2, \dots, A_M \in \mathscr{A}^*$.

Let $A := \bigcupdot_{n\in\mathbb{N}} A_n$ for a sequence $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}^*$ of mutually disjoint sets. Since $A_1 \cup \dots \cup A_M \in \mathscr{A}^*$, we can use (OM$_2$) and (6.7) to deduce

$$\begin{aligned}\mu^*(Q) &= \mu^*(Q \cap (A_1 \cup \dots \cup A_M)) + \mu^*(Q \setminus (A_1 \cup \dots \cup A_M)) \\ &\geqslant \mu^*(Q \cap (A_1 \cup \dots \cup A_M)) + \mu^*(Q \setminus A) \\ &= \sum_{n=1}^{M} \mu^*(Q \cap A_n) + \mu^*(Q \setminus A).\end{aligned}$$

The left-hand side is independent of M; therefore, we can let $M \to \infty$ and get

$$\mu^*(Q) \geqslant \sum_{n=1}^{\infty} \mu^*(Q \cap A_n) + \mu^*(Q \setminus A) \overset{(\mathrm{OM}_3)}{\geqslant} \mu^*(Q \cap A) + \mu^*(Q \setminus A). \tag{6.8}$$

The reverse inequality $\mu^*(Q) \leqslant \mu^*(Q \cap A) + \mu^*(Q \setminus A)$ follows at once from the subadditivity of μ^*. This means that equality holds throughout (6.8), and we get $A \in \mathscr{A}^*$. If we take $Q := A$ in (6.8) we even see the σ-additivity of μ^* on $\mathscr{A}^*$.

Up to now we have seen that $\mathscr{A}^*$ is a $\cup$-stable Dynkin system. Because we have that $A \cap B = (A^c \cup B^c)^c$ we see that $\mathscr{A}^*$ is also $\cap$-stable and, by Lemma 5.4, a σ-algebra.

Step 4. Claim: μ^* is a measure on $\mathscr{A}^*$ and $\sigma(\mathscr{S})$ which extends μ. We have seen in Step 3b that μ^* is σ-additive, and hence is a measure on $\mathscr{A}^*$. By Step 3a, $\mathscr{S} \subset \mathscr{A}^*$ and thus $\sigma(\mathscr{S}) \subset \sigma(\mathscr{A}^*) = \mathscr{A}^*$ since $\mathscr{A}^*$ is itself a σ-algebra (by Step 3b). Again by Step 3b, $\mu^*|_{\sigma(\mathscr{S})}$ is a measure which, by Step 2d, extends μ.

Step 5. Uniqueness of the extension $\mu^*|_{\sigma(\mathscr{S})}$. If there is an exhausting sequence $(S_n)_{n\in\mathbb{N}} \subset \mathscr{S}$, $S_n \uparrow X$ such that $\mu(S_n) < \infty$ for all $n \in \mathbb{N}$, it follows from Theorem 5.7 that any two extensions of μ to $\sigma(\mathscr{S})$ coincide. $\square$

Caution The core of Carathéodory's theorem, Theorem 6.1, is the definition (6.2) of μ^*-measurable sets, i.e. of the σ-algebra $\mathscr{A}^*$. The proof shows that, in general, we cannot expect μ^* to be (σ-)additive outside $\mathscr{A}^*$. In many situations the σ-algebra $\mathscr{P}(X)$ is simply too big to support a non-trivial measure. Notable exceptions are countable sets X or Dirac measures. [✍] For n-dimensional Lebesgue measure, this was first remarked by Hausdorff [21, pp. 401–402]. The general case depends on the cardinality of X and the behaviour of μ on one-point sets; see the discussion in Oxtoby [35, Chapter 5].

Put in other words, this says that even a household measure like Lebesgue measure cannot assign a content to *every* set! In $\mathbb{R}^3$ (and higher dimensions) we even have the *Banach–Tarski paradox*: the open balls $B_1(0)$ and $B_2(0)$ with centre 0 and radii 1, resp. 2, have finite disjoint decompositions $B_1(0) = \bigcupdot_{n=1}^{M} E_n$

and $B_2(0) = \dot{\bigcup}_{n=1}^{M} F_n$ such that for every $n = 1, 2, \dots, M$ the sets E_n and F_n are geometrically congruent (hence, they should have the same Lebesgue measure); see Stromberg [53] or Wagon [57]. Of course, not all of the sets E_n and F_n can be Borel sets.

This brings us to the question of whether and how we can construct a *non-Borel measurable set*, i.e. a set $A \in \mathscr{P}(\mathbb{R}^n) \setminus \mathscr{B}(\mathbb{R}^n)$. Such constructions are possible but they are based on the axiom of choice, see for example Hewitt and Stromberg [22, pp. 135–6], Oxtoby [35, pp. 22–3] or Appendix G. One can show that the axiom of choice is really needed for such constructions, see Ciesielski [9, p. 55].

Existence of Lebesgue measure in $\mathbb{R}$

Let us now apply Theorem 6.1 to prove the existence of one-dimensional Lebesgue measure λ^1 which was defined for half-open intervals $\mathscr{J}^1$ in Definition 4.6:

$$\lambda^1[a, b) = b - a, \quad \forall [a, b) \in \mathscr{J}^1.$$

It is obvious [✍] that $\mathscr{J}^1$ is a semi-ring. The tricky part is the σ-additivity relative to $\mathscr{J}^1$.

Proposition 6.3 *λ^1 is a pre-measure on $\mathscr{J}^1$.*

Proof Write $\lambda := \lambda^1$ and $\mathscr{J} := \mathscr{J}^1$. We have for $a < b' \leqslant a' < b$ that

$$\lambda\big([a, a') \cup [b', b)\big) = \lambda[a, b) = b - a \leqslant (b - b') + (a' - a) = \lambda[b', b) + \lambda[a, a').$$

This shows that λ is subadditive on $\mathscr{J}$. If $a' = b'$, then the above calculation holds with '$=$', i.e. λ is finitely additive.

Let $I_n = [a_n, b_n)$ be mutually disjoint intervals such that $\dot{\bigcup}_{n\in\mathbb{N}} I_n = [a, b)$. For $0 < \epsilon < b - a$ we then define $I_{n,\epsilon} := [a_n - 2^{-n}\epsilon, b_n)$. The open intervals $I_{n,\epsilon}^\circ = (a_n - 2^{-n}\epsilon, b_n)$ cover the compact interval $[a, b - \epsilon]$, and because of compactness, we find a finite sub-cover, i.e. there is some $N = N_\epsilon \in \mathbb{N}$ such that

$$[a, b - \epsilon) \subset [a, b - \epsilon] \subset \bigcup_{n=1}^{N} I_{n,\epsilon}^\circ \subset \bigcup_{n=1}^{N} I_{n,\epsilon}.$$

Since $\lambda(I_n) = \lambda(I_{n,\epsilon}) - \epsilon/2^n$, we see

$$\lambda[a, b) - \sum_{n=1}^{N} \lambda(I_n) = \epsilon + \underbrace{\lambda[a, b - \epsilon) - \sum_{n=1}^{N} \lambda(I_{n,\epsilon})}_{\leqslant 0 \ \text{subadditivity}} + \sum_{n=1}^{N} \frac{\epsilon}{2^n} \leqslant 2\epsilon. \tag{6.9}$$

On the other hand, monotonicity and finite additivity of λ entail

$$\bigcupdot_{n=1}^{N} I_n \subset [a,b) \implies \sum_{n=1}^{N} \lambda(I_n) = \lambda\left(\bigcupdot_{n=1}^{N} I_n\right) \leqslant \lambda[a,b),$$

which shows that left-hand side of (6.9) is positive.

Letting, in the expression (6.9), first $N \to \infty$ (for fixed $\epsilon > 0$) and then $\epsilon \to 0$, we get $\sum_{n\in\mathbb{N}} \lambda(I_n) = \lambda[a,b)$. This proves that λ is σ-additive on $\mathscr{J}$. $\square$

The existence and uniqueness of Lebesgue measure on $\sigma(\mathscr{J}) = \mathscr{B}(\mathbb{R})$ now follows directly from Theorem 6.2.

Corollary 6.4 (Lebesgue measure on $\mathbb{R}$) *λ^1 is a measure on $\mathscr{B}(\mathbb{R})$. It is the only measure such that $\lambda^1[a,b) = b - a$ for $a < b$.*

Proof We know from Theorem 3.8 that $\mathscr{B}(\mathbb{R}) = \sigma(\mathscr{J})$. Since $[-k,k) \uparrow \mathbb{R}$ is an exhausting sequence of half-open intervals and $\lambda^1[-k,k) = 2k < \infty$, all conditions of Carathéodory's theorem, Theorem 6.1, are fulfilled, and λ^1 has a unique extension to a measure on $\mathscr{B}(\mathbb{R})$. $\square$

A very elegant way to introduce n-dimensional Lebesgue measure is to use product measures and Fubini's theorem, see Chapter 14 and Corollary 14.7. Go directly to the next chapter if you opt for this approach. If you do not want to wait any longer with the construction of Lebesgue measure, then continue reading …

✻Existence of Lebesgue measure in $\mathbb{R}^n$

We will again use Theorem 6.1 to prove the existence of n-dimensional Lebesgue measure. Recall from Definition 4.6 that on the half-open rectangles $\mathscr{J}^n = \mathscr{J}(\mathbb{R}^n)$

$$\lambda^n [\![a,b)\!) = \prod_{i=1}^{n} (b_i - a_i), \qquad [\![a,b)\!) = \bigtimes_{i=1}^{n} [a_i, b_i) \in \mathscr{J}^n.$$

Proposition 6.5 *The family of n-dimensional rectangles $\mathscr{J}^n$ is a semi-ring.*

Proof (by induction). It is obvious that $\mathscr{J}^1$ satisfies the properties (S_1)–(S_3) from page 39. Assume that $\mathscr{J}^n$ is a semi-ring for some $n \geqslant 1$. From the definition of rectangles it is clear that

$$\mathscr{J}^{n+1} = \mathscr{J}^n \times \mathscr{J}^1 := \{I_n \times I_1 : I_n \in \mathscr{J}^n,\ I_1 \in \mathscr{J}^1\}.$$

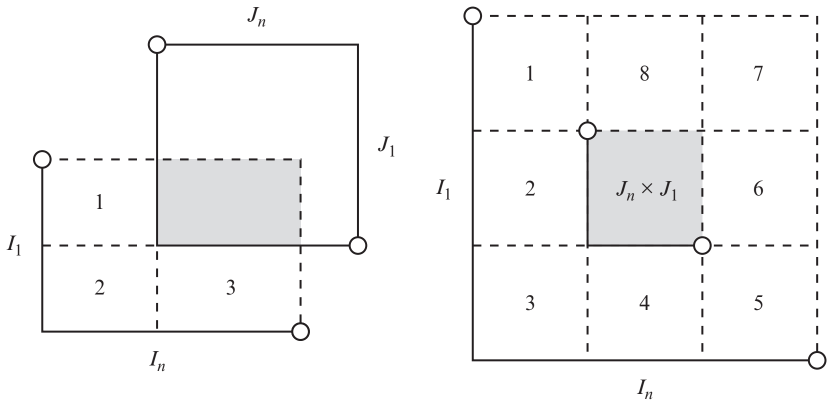


Fig. 6.1. In $\mathbb{R}^2$ it is easy to depict the two typical situations that occur in the proof of (S_3) in Proposition 6.5.

(S_1) is obviously true, and (S_2) follows from the identity

$$(I_n \times I_1) \cap (J_n \times J_1) = (I_n \cap J_n) \times (I_1 \cap J_1), \tag{6.10}$$

where $I_n, J_n \in \mathscr{J}^n$ and $I_1, J_1 \in \mathscr{J}^1$. Since

$$\begin{aligned}
&(J_n \times J_1)^c \\
&\quad = \big\{(x,y) : x \notin J_n,\ y \notin J_1 \text{ or } x \in J_n,\ y \notin J_1 \text{ or } x \notin J_n,\ y \in J_1\big\} \\
&\quad = (J_n^c \times J_1^c) \uplus (J_n \times J_1^c) \uplus (J_n^c \times J_1)
\end{aligned}$$

we see, using (6.10),

$$\begin{aligned}
&(I_n \times I_1) \setminus (J_n \times J_1) \\
&\quad = (I_n \times I_1) \cap (J_n \times J_1)^c \\
&\quad = \big[(I_n \setminus J_n) \times (I_1 \setminus J_1)\big] \uplus \big[(I_n \cap J_n) \times (I_1 \setminus J_1)\big] \uplus \big[(I_n \setminus J_n) \times (I_1 \cap J_1)\big].
\end{aligned}$$

Both $I_n \setminus J_n$ and $I_1 \setminus J_1$ are made up of finitely many disjoint rectangles from $\mathscr{J}^n$ and $\mathscr{J}^1$, and therefore $(I_n \times I_1) \setminus (J_n \times J_1)$ is a finite union of disjoint rectangles from $\mathscr{J}^n \times \mathscr{J}^1$; thus ($S_3$) holds (Fig. 6.1). □

The proof of Proposition 6.5 reveals a bit more: the Cartesian product of any two semi-rings is again a semi-ring. [✍]

Proposition 6.6 *λ^n is a pre-measure on $\mathscr{J}^n$.*

Proof We have to show that $\lambda^n(\emptyset) = 0$ and that λ^n is σ-additive on $\mathscr{J}^n$. It is clear that $\lambda^n(\emptyset) = 0$.

To see σ-additivity, we use induction with respect to the dimension n. Proposition 6.3 covers the case $n = 1$.

Now assume that $\nu = \lambda^n$ is σ-additive the rectangles $\mathscr{J}^n$ for some $n \geqslant 1$. Let $I_i = I_i{}^1 \times I_i{}^n \in \mathscr{J}^1 \times \mathscr{J}^n = \mathscr{J}^{n+1}$ be mutually disjoint rectangles such that $\dot\bigcup_{i\in\mathbb{N}} I_i = I \in \mathscr{J}^{n+1}$. Since $I \in \mathscr{J}^{n+1}$ we know that

$$I = \left(\bigcup_{i\in\mathbb{N}} I_i{}^1\right) \times \left(\bigcup_{i\in\mathbb{N}} I_i{}^n\right), \quad \bigcup_{i\in\mathbb{N}} I_i{}^d \in \mathscr{J}^d, \quad d \in \{1, n\}.$$

Define $\hat{I}_1{}^d := I_1{}^d$ and $\hat{I}_{i+1}{}^d := I_{i+1}{}^d \setminus (I_1{}^d \cup \cdots \cup I_i^d) = (I_{i+1}{}^d \setminus I_i{}^d) \setminus \cdots \setminus I_1{}^d$. Clearly, the sets $\hat{I}_i{}^d$ are disjoint and

$$\bigcup_{i=1}^N I_i{}^d = \dot\bigcup_{i=1}^N \hat{I}_i{}^d \quad \text{for all} \quad N \in \mathbb{N} \quad \text{and} \quad d \in \{1, n\}.$$

Since $\mathscr{J}^d$ is a semi-ring, the property (S_3) shows that each $\hat{I}_i{}^d$ is a finite union of disjoint rectangles. Thus, there exist disjoint sets $\tilde{I}_k^1 \in \mathscr{J}^1$ and $\tilde{I}_\ell^n \in \mathscr{J}^n$ such that

$$I = \dot\bigcup_{i\in\mathbb{N}} (I_i^1 \times I_i^n) = \dot\bigcup_{k\in\mathbb{N}} \dot\bigcup_{\ell\in\mathbb{N}} (\tilde{I}_k^1 \times \tilde{I}_\ell^n) = \dot\bigcup_{k\in\mathbb{N}} \tilde{I}_k^1 \times \dot\bigcup_{\ell\in\mathbb{N}} \tilde{I}_\ell^n \in \mathscr{J}^1 \times \mathscr{J}^n.$$

Now we can use the σ-additivity of λ^1 and λ^n and the very definition of λ^{n+1} as a product on $\mathscr{J}^{n+1}$ to see that

$$\begin{aligned}\lambda^{n+1}\left(\dot\bigcup_{i\in\mathbb{N}} I_i\right) = \lambda^{n+1}\left(\dot\bigcup_{k\in\mathbb{N}} \tilde{I}_k^1 \times \dot\bigcup_{\ell\in\mathbb{N}} \tilde{I}_\ell^n\right) &= \lambda^1\left(\dot\bigcup_{k\in\mathbb{N}} \tilde{I}_k^1\right) \cdot \lambda^n\left(\dot\bigcup_{\ell\in\mathbb{N}} \tilde{I}_\ell^n\right)\\ &= \sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} \lambda^1(\tilde{I}_k^1) \cdot \lambda^n(\tilde{I}_\ell^n)\\ &= \sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} \lambda^{n+1}(\tilde{I}_k^1 \times \tilde{I}_\ell^n).\end{aligned}$$

A similar equality holds for each rectangle $I_i = I_i^1 \times I_i{}^n$, i.e.

$$\lambda^{n+1}(I_i) = \sum\sum_{(k,\ell)\,:\,\tilde{I}_k^1 \times \tilde{I}_\ell^n \subset I_i^1 \times I_i^n} \lambda^{n+1}(\tilde{I}_k^1 \times \tilde{I}_\ell{}^n),$$

and summation over all $i \in \mathbb{N}$ proves that

$$\sum_{i\in\mathbb{N}} \lambda^{n+1}(I_i) = \sum_{i\in\mathbb{N}} \sum\sum_{(k,\ell):\tilde{I}_k^1 \times \tilde{I}_\ell^n \subset I_i^1 \times I_i^n} \lambda^{n+1}(\tilde{I}_k^1 \times \tilde{I}_\ell^n) = \sum_{k\in\mathbb{N}} \sum_{\ell\in\mathbb{N}} \lambda^{n+1}(\tilde{I}_k^1 \times \tilde{I}_\ell^n).$$

This shows that $\lambda^{n+1}\left(\dot\bigcup_{i\in\mathbb{N}} I_i\right) = \sum_{i\in\mathbb{N}} \lambda^{n+1}(I_i)$ which means that λ^{n+1} is σ-additive on $\mathscr{J}^{n+1}$. $\square$

After these preparations, we can prove the existence of n-dimensional Lebesgue measure just as in the one-dimensional case, see Corollary 6.4.

Corollary 6.7 (Lebesgue measure on $\mathbb{R}^n$) *λ^n is a measure on $\mathscr{B}(\mathbb{R}^n)$. It is the only measure such that*

$$\lambda^n\left(\bigtimes_{i=1}^n [a_i, b_i)\right) = \prod_{i=1}^n (b_i - a_i), \quad \forall a_i < b_i,\ i = 1, \ldots, n.$$

Remark 6.8 The uniqueness of Lebesgue measure and its properties (see Theorem 4.7) show that it is necessarily the familiar volume (length, area …) function $\mathrm{vol}^{(n)}(\cdot)$ from elementary geometry in the sense that $\mathrm{vol}^{(n)}$ can be extended to a measure on the Borel σ-algebra in only one way.

Problems

6.1. Stieltjes measure (1).

(i) Let μ be a measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\mu[-n, n) < \infty$ for all $n \in \mathbb{N}$. Show that

$$F_\mu(x) := \begin{cases} \mu[0, x) & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -\mu[x, 0) & \text{if } x < 0 \end{cases}$$

is a monotonically increasing and left-continuous function $F_\mu : \mathbb{R} \to \mathbb{R}$.
Remark. Increasing and left-continuous functions are called *Stieltjes functions*.

(ii) Let $F : \mathbb{R} \to \mathbb{R}$ be a Stieltjes function (see part (i)). Show that

$$\nu_F([a, b)) := F(b) - F(a), \quad \forall a, b \in \mathbb{R},\ a < b,$$

has a unique extension to a measure on $\mathscr{B}(\mathbb{R})$.
[Hint: check the assumptions of Theorem 6.1 with $\mathcal{S} = \{[a, b) : a \leqslant b\}$.]

(iii) Conclude that for every measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ satisfying $\mu[-r, r) < \infty$, $r > 0$, there is some Stieltjes function $F = F_\mu$ such that $\mu = \nu_F$ (as in part (ii)).

(iv) Which Stieltjes function F corresponds to λ (one-dimensional Lebesgue measure)?

(v) Which Stieltjes function F corresponds to δ_0 (Dirac measure on $\mathbb{R}$)?

(vi) Show that F_μ as in part (i) is continuous at $x \in \mathbb{R}$ if, and only if, $\mu\{x\} = 0$.

6.2. Let μ^* be an outer measure on X, and let $A_1, A_2, \ldots$ be a sequence of mutually disjoint μ^*-measurable sets, i.e. $A_i \in \mathcal{A}^*$, $i \in \mathbb{N}$. Show that

$$\mu^*\left(Q \cap \bigcup_i A_i\right) = \sum_{i=1}^\infty \mu^*(Q \cap A_i) \quad \text{for all } Q \subset X.$$

6.3. Consider on $\mathbb{R}$ the family Σ of all Borel sets which are symmetric w.r.t. the origin. Show that Σ is a σ-algebra. Is it possible to extend a pre-measure μ on Σ to a measure on $\mathscr{B}(\mathbb{R})$? If so, is this extension unique? (Problem 9.14 is a continuation of this problem.)

6.4. Completion (2). Recall from Problem 9.14 that a measure space $(X, \mathscr{A}, \mu)$ is complete if every subset of a μ-null set is a μ-null set (thus, in particular, measurable). Let $(X, \mathscr{A}, \mu)$

be a σ-finite measure space – i.e. there is an exhausting sequence $(A_i)_{i\in\mathbb{N}} \subset \mathscr{A}$ such that $\mu(A_i) < \infty$. As in the proof of Theorem 6.1 we write μ^* for the outer measure (6.1) – now defined using $\mathscr{A}$-coverings – and $\mathscr{A}^*$ for the σ-algebra defined by (6.2).

(i) Show that for every $Q \subset X$ there is some $A \in \mathscr{A}$ such that $\mu^*(Q) = \mu(A)$ and that $\mu(N) = 0$ for all $N \subset A \setminus Q$ with $N \in \mathscr{A}$.
[Hint: since μ^* is defined as an infimum, every Q with $\mu^*(Q) < \infty$ admits a sequence $B_k \in \mathscr{A}$ with $B_k \supset Q$ and $\mu(B_k) - \mu^*(Q) \leqslant 1/k$. If $\mu^*(Q) = \infty$, consider for each $i \in \mathbb{N}$ the set $Q \cap A_i$.]
(ii) Show that $(X, \mathscr{A}^*, \mu^*|_{\mathscr{A}^*})$ is a complete measure space.
(iii) Show that $(X, \mathscr{A}^*, \mu^*|_{\mathscr{A}^*})$ is the completion of $(X, \mathscr{A}, \mu)$ in the sense of Problem 4.15.

6.5. Let $(X, \mathscr{A}, \mu)$ be a finite measure space, $\mathscr{B} \subset \mathscr{A}$ a Boolean algebra (i.e. $X \in \mathscr{B}$, $\mathscr{B}$ is stable under the formation of finite unions, intersections and complements) and $m : \mathscr{B} \to [0, \infty)$ an additive set functions satisfying $0 \leqslant m(B) \leqslant \mu(B)$. Show that m is a pre-measure.

6.6. (i) Show that non-void open sets in $\mathbb{R}$ (resp. $\mathbb{R}^n$) have always strictly positive Lebesgue measure.
[Hint: let U be open. Find a small ball in U and inscribe a square.]
(ii) Does your answer to part (i) hold also for closed sets?

6.7. (i) Show that $\lambda^1((a,b)) = b - a$ for all $a, b \in \mathbb{R}$, $a \leqslant b$.
[Hint: approximate (a, b) by half-open intervals and use Proposition 4.3(vi), (vii).]
(ii) Let $H \subset \mathbb{R}^2$ be a hyperplane which is perpendicular to the x_1-direction (that is to say, H is a translate of the x_2-axis). Show that $H \in \mathscr{B}(\mathbb{R}^2)$ and $\lambda^2(H) = 0$.
[Hint: note that $H \subset y + \cup_{k\in\mathbb{N}} A_k$ for some y and $A_k = [-\epsilon 2^{-k}, \epsilon 2^{-k}) \times [-k, k)$.]
(iii) State and prove the $\mathbb{R}^n$-analogues of parts (i) and (ii).

6.8. Let $(X, \mathscr{A}, \mu)$ be a measure space such that all *singletons* $\{x\} \in \mathscr{A}$. A point x is called an *atom*, if $\mu\{x\} > 0$. A measure is called *non-atomic* or *diffuse*, if there are no atoms.

(i) Show that the one-dimensional Lebesgue measure λ^1 is diffuse.
(ii) Give an example of a non-diffuse measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.
(iii) Show that for a diffuse measure μ on $(X, \mathscr{A})$ all countable sets are null sets.
(iv) Show that every probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ can be decomposed into a sum of two measures $\mu + \nu$, where μ is diffuse and ν is a measure of the form $\nu = \sum_{i\in\mathbb{N}} \epsilon_i \delta_{x_i}$, $\epsilon_i > 0$, $x_i \in \mathbb{R}$.
[Hint: since $\mathbb{P}(\mathbb{R}) = 1$, there are at most k points $y_1^{(k)}, y_2^{(k)}, \ldots, y_k^{(k)}$ such that $\frac{1}{k-1} > P\{y_i^{(k)}\} \geqslant \frac{1}{k}$. Find by recursion (in k) all points satisfying such a relation. There are at most countably many of these $y_i^{(k)}$. Relabel them as $x_1, x_2, \ldots$. These are the atoms of $\mathbb{P}$. Now take $\epsilon_i = \mathbb{P}\{y_i\}$, define ν as stated and prove that ν and $\mathbb{P} - \nu$ are measures.]

6.9. A set $A \subset \mathbb{R}^n$ is called *bounded* if it can be contained in a ball $B_r(0) \supset A$ of finite radius r. A set $A \subset \mathbb{R}^n$ is called *pathwise connected* if we can go along a curve from any point $a \in A$ to any other point $a' \in A$ without ever leaving A.

(i) Construct an open and unbounded set in $\mathbb{R}$ with finite, strictly positive Lebesgue measure.
[Hint: try unions of ever smaller open intervals centred around $n \in \mathbb{N}$.]
(ii) Construct an open, unbounded and pathwise connected set in $\mathbb{R}^2$ with finite, strictly positive Lebesgue measure.
[Hint: try a union of adjacent, ever longer, ever thinner rectangles.]
(iii) Is there a pathwise connected, open and unbounded set in $\mathbb{R}$ with finite, strictly positive Lebesgue measure?

6.10. Let $\lambda := \lambda^1|_{[0,1]}$ be Lebesgue measure on $([0,1], \mathscr{B}[0,1])$. Show that for every $\epsilon > 0$ there is a dense open set $U \subset [0,1]$ with $\lambda(U) \leqslant \epsilon$.
[Hint: take an enumeration $(q_i)_{i\in\mathbb{N}}$ of $\mathbb{Q} \cap (0,1)$ and make each q_i the centre of a small open interval.]

6.11. Let $\lambda = \lambda^1$ be Lebesgue measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Show that $N \in \mathscr{B}(\mathbb{R})$ is a null set if, and only if, for every $\epsilon > 0$ there is an open set $U_\epsilon \supset N$ such that $\lambda(U_\epsilon) < \epsilon$.
[Hint: sufficiency is trivial, for necessity use λ^* constructed in Theorem 6.1 from $\lambda|_{\mathscr{O}}$ and observe that $\lambda|_{\mathscr{B}(\mathbb{R}^n)} = \lambda^*|_{\mathscr{B}(\mathbb{R}^n)}$. This gives the required open cover.]

6.12. **Borel–Cantelli lemma (1)** – the direct half. Prove the following theorem.
Theorem (Borel–Cantelli lemma). *Let* $(\Omega, \mathscr{A}, \mathbb{P})$ *be a probability space. For every sequence* $(A_i)_{i\in\mathbb{N}} \subset \mathscr{A}$ *we have*

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty \Longrightarrow \mathbb{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i\Big) = 0. \tag{6.11}$$

[Hint: use Proposition 4.3(vii) and the fact that $\mathbb{P}\big(\bigcup_{i\geqslant n} A_i\big) \leqslant \sum_{i\geqslant n} \mathbb{P}(A_i)$.]
Remark. This is the 'easy' or direct half of the so-called Borel–Cantelli lemma; for the more difficult part see Theorem 24.9. The condition $\omega \in \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$ means that ω happens to be in *infinitely many* of the A_i and the lemma gives a simple sufficient condition when certain events happen almost surely *not* infinitely often, i.e. only finitely often with probability one.

6.13. **Non-measurable sets (1).** Let μ be a measure on $\mathscr{A} = \{\emptyset, [0,1), [1,2), [0,2)\}$, $X = [0,2)$, such that $\mu([0,1)) = \mu([1,2)) = \frac{1}{2}$ and $\mu([0,2)) = 1$. Denote by μ^* and $\mathscr{A}^*$ the outer measure and σ-algebra which appear in the proof of Theorem 6.1.

(i) Find $\mu^*(a,b)$ and $\mu^*\{a\}$ for all $0 \leqslant a < b < 2$ if we use $\mathscr{S} = \mathscr{A}$ in Theorem 6.1.
(ii) Show that $(0,1), \{0\} \notin \mathscr{A}^*$.

6.14. **Non-measurable sets (2).** Consider on $X = \mathbb{R}$ the σ-algebra $\mathscr{A} := \{A \subset \mathbb{R} : A \text{ or } A^c \text{ is countable}\}$ from Example 3.3(v) and the measure $\gamma(A)$ from Example 4.5(ii), which is 0 or 1 according to A or A^c being countable. Denote by γ^* and $\mathscr{A}^*$ the outer measure and σ-algebra which appear in the proof of Theorem 6.1.

(i) Find γ^* if we use $\mathscr{S} = \mathscr{A}$ in Theorem 6.1.
(ii) Show that no set $B \subset \mathbb{R}$, such that both B and B^c are uncountable, is in $\mathscr{A}$ or in $\mathscr{A}^*$.

7

Measurable Mappings

In this chapter we consider maps $T: X \to X'$ between two measurable spaces $(X, \mathscr{A})$ and $(X', \mathscr{A}')$ which respect the measurable structures, that is σ-algebras, on X and X'. Such maps can be used to transport a given measure μ, defined on $(X, \mathscr{A})$, onto $(X', \mathscr{A}')$. We have already used this technique in Theorem 5.8, where we considered shifts of sets $A \rightsquigarrow x + A$, but probability theory is where this concept is truly fundamental: you use it whenever you speak of the 'distribution' of a 'random variable'.

Definition 7.1 Let $(X, \mathscr{A}), (X', \mathscr{A}')$ be measurable spaces. A map $T: X \to X'$ is called $\mathscr{A}/\mathscr{A}'$-*measurable* (or *measurable* unless this is too ambiguous) if the pre-image of every measurable set is a measurable set:

$$T^{-1}(A') \in \mathscr{A} \qquad \forall A' \in \mathscr{A}'. \tag{7.1}$$

Caution • Probabilists call a measurable map on a probability space a *random variable*.

- Often the (symbolic!) notation $T^{-1}(\mathscr{A}') := \{T^{-1}(A') : A' \in \mathscr{A}'\}$ and shorthand $T^{-1}(\mathscr{A}') \subset \mathscr{A}$ are used instead of (7.1).
- The notation $T: (X, \mathscr{A}) \to (X', \mathscr{A}')$ is commonly used to indicate measurability of the map T.
- A $\mathscr{B}(\mathbb{R}^n)/\mathscr{B}(\mathbb{R}^m)$ measurable map is usually called a *Borel* (*measurable*) *map*.

Using Definition 7.1 we can recast the measurability problem which we met in Theorem 5.8, (5.4). For the maps

$$\tau_x : \mathbb{R}^n \to \mathbb{R}^n, \; y \mapsto y - x \quad \text{and} \quad \tau_x^{-1} = \tau_{-x} : \mathbb{R}^n \to \mathbb{R}^n \; y \mapsto y + x$$

we have for all Borel sets $B \in \mathscr{B}(\mathbb{R}^n)$

$$x + B \in \mathscr{B}(\mathbb{R}^n) \iff \tau_{-x}(B) = \tau_x^{-1}(B) \in \mathscr{B}(\mathbb{R}^n),$$

i.e. τ_x is $\mathscr{B}(\mathbb{R}^n)/\mathscr{B}(\mathbb{R}^n)$-measurable or *Borel measurable*. In the proof of Theorem 5.6 we have used an *ad hoc* argument, reducing everything to the generator $\mathscr{J}$ of the σ-algebra – but this is good enough even in the most general case. The following lemma shows that measurability need be checked only for the sets of a generator.

Lemma 7.2 *Let $(X, \mathscr{A}), (X', \mathscr{A}')$ be measurable spaces and let $\mathscr{A}' = \sigma(\mathscr{G}')$. Then $T: X \to X'$ is $\mathscr{A}/\mathscr{A}'$-measurable if, and only if, $T^{-1}(\mathscr{G}') \subset \mathscr{A}$, i.e. if*

$$T^{-1}(G') \in \mathscr{A} \quad \forall G' \in \mathscr{G}'. \tag{7.2}$$

Proof If T is $\mathscr{A}/\mathscr{A}'$-measurable, we have $T^{-1}(\mathscr{G}') \subset T^{-1}(\mathscr{A}') \subset \mathscr{A}$, and (7.2) is obviously satisfied.

Conversely, consider the system $\Sigma' := \{A' \subset X' : T^{-1}(A') \in \mathscr{A}\}$. By (7.2), $\mathscr{G}' \subset \Sigma'$ and it is not difficult to see that Σ' is itself a σ-algebra since T^{-1} commutes with all set operations. [✍] Therefore,

$$\mathscr{A}' = \sigma(\mathscr{G}') \subset \sigma(\Sigma') = \Sigma' \implies T^{-1}(A') \in \mathscr{A} \qquad \forall A' \in \mathscr{A}'. \qquad \square$$

On a topological space $(X, \mathscr{O})$ – see page 18 – we usually consider the (topological) Borel σ-algebra $\mathscr{B}(X) := \sigma(\mathscr{O})$. The interplay between measurability and topology can be intricate. One of the simple and extremely useful aspects is the fact that *continuous maps are measurable*; let us check this for $\mathbb{R}^n$.

Example 7.3 Continuous maps $T: \mathbb{R}^m \to \mathbb{R}^n$ are $\mathscr{B}(\mathbb{R}^m)/\mathscr{B}(\mathbb{R}^n)$-measurable.

From calculus[1] we know that T is continuous if, and only if,

$$T^{-1}(V) \subset \mathbb{R}^m \quad \text{is open} \quad \forall \text{ open } V \subset \mathbb{R}^n. \tag{7.3}$$

Since the open sets $\mathscr{O}_{\mathbb{R}^m}$ in $\mathbb{R}^m$ generate the Borel σ-algebra $\mathscr{B}(\mathbb{R}^m)$, we can use (7.3) to deduce

$$T^{-1}(\mathscr{O}_{\mathbb{R}^n}) \subset \mathscr{O}_{\mathbb{R}^m} \subset \sigma(\mathscr{O}_{\mathbb{R}^m}) = \mathscr{B}(\mathbb{R}^m).$$

By Lemma 7.2, $T^{-1}(\mathscr{B}(\mathbb{R}^n)) \subset \mathscr{B}(\mathbb{R}^m)$, which means that T is measurable.

Caution Not every measurable map is continuous, e.g. $x \mapsto \mathbb{1}_{[-1,1]}(x)$ is clearly Borel measurable, but discontinuous at $x = \pm 1$.

Theorem 7.4 *Let $(X_i, \mathscr{A}_i)$, $i = 1, 2, 3$, be measurable spaces and $T: X_1 \to X_2$, $S: X_2 \to X_3$ be $\mathscr{A}_1/\mathscr{A}_2$- resp. $\mathscr{A}_2/\mathscr{A}_3$-measurable maps. Then $S \circ T: X_1 \to X_3$ is $\mathscr{A}_1/\mathscr{A}_3$-measurable.*

[1] See also Appendix B, Theorem B.1.

Proof For $A_3 \in \mathscr{A}_3$ we have

$$(S \circ T)^{-1}(A_3) = T^{-1}\big(\underbrace{S^{-1}(A_3)}_{\in \mathscr{A}_2}\big) \in T^{-1}(\mathscr{A}_2) \subset \mathscr{A}_1. \qquad \square$$

Often we find ourselves in a situation where $T: X \to X'$ is given and X' is equipped with a natural σ-algebra $\mathscr{A}'$ – e.g. if $X' = \mathbb{R}$ and $\mathscr{A}' = \mathscr{B}(\mathbb{R})$ – but no σ-algebra is specified in X. Then the following question arises: *is there a (smallest) σ-algebra on X which makes T measurable?* An obvious, but nevertheless useless, candidate is $\mathscr{P}(X)$, which renders every map measurable. [✍] From Example 3.3(vii) we know that $T^{-1}(\mathscr{A}')$ is a σ-algebra in X but we cannot remove a single set from it without endangering the measurability of T. [✍] Let us formalize this observation.

Definition 7.5 (and lemma) Let $(T_i)_{i\in I}$, $T_i : X \to X_i$, be arbitrarily many mappings from the same space X into measurable spaces $(X_i, \mathscr{A}_i)$. The smallest σ-algebra on X that makes all T_i simultaneously measurable [✍] is

$$\sigma(T_i : i \in I) := \sigma\Bigg(\bigcup_{i\in I} T_i^{-1}(\mathscr{A}_i)\Bigg). \tag{7.4}$$

We say that $\sigma(T_i : i \in I)$ is *generated by the family $(T_i)_{i\in I}$*.

Although $T_i^{-1}(\mathscr{A}_i)$ is a σ-algebra this is, in general, not true for $\bigcup_{i\in I} T_i^{-1}(\mathscr{A}_i)$ if $\#I > 1$; this explains why we have to use the σ-hull in (7.4).

Theorem 7.6 *Let $(X, \mathscr{A}), (X', \mathscr{A}')$ be measurable spaces and $T: X \to X'$ be an $\mathscr{A}/\mathscr{A}'$-measurable map. For every measure μ on $(X, \mathscr{A})$,*

$$\mu'(A') := \mu(T^{-1}(A')), \qquad A' \in \mathscr{A}', \tag{7.5}$$

defines a measure on $(X', \mathscr{A}')$.

Proof If $A' = \emptyset$, then $T^{-1}(\emptyset) = \emptyset$ and $\mu'(\emptyset) = \mu(\emptyset) = 0$. If $(A_n')_{n\in\mathbb{N}} \subset \mathscr{A}'$ is a sequence of mutually disjoint sets, then

$$\begin{aligned}\mu'\Big(\dot\bigcup_{n\in\mathbb{N}} A_n'\Big) &= \mu\Big(T^{-1}\Big(\dot\bigcup_{n\in\mathbb{N}} A_n'\Big)\Big) \overset{[✍]}{=} \mu\Big(\dot\bigcup_{n\in\mathbb{N}} T^{-1}(A_n')\Big)\\ &= \sum_{n\in\mathbb{N}} \mu\big(T^{-1}(A_n')\big) = \sum_{n\in\mathbb{N}} \mu'(A_n'). \end{aligned} \qquad \square$$

Definition 7.7 The measure $\mu'(\,\cdot\,)$ of Theorem 7.6 is called the *image measure* or *push forward* of μ under T and is denoted by $T(\mu)(\,\cdot\,)$ or $T_*\mu(\,\cdot\,)$ or $\mu \circ T^{-1}(\,\cdot\,)$.

Example 7.8 (i) In the proof of Theorem 5.8 we use $\lambda^n(x + B) = \lambda^n(\tau_x^{-1}(B)) = \tau_x(\lambda^n)(B)$, $B \in \mathscr{B}(\mathbb{R}^n)$.

(ii) Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and $\xi : \Omega \to \mathbb{R}$ be a random variable, i.e. an $\mathscr{A}/\mathscr{B}(\mathbb{R})$-measurable map. Then[2]

$$\xi(\mathbb{P})(A) = \mathbb{P}(\xi^{-1}(A)) = \mathbb{P}\{\omega : \xi(\omega) \in A\} = \mathbb{P}(\xi \in A)$$

is a probability measure, called the *law* or *distribution* of the random variable ξ.

(iii) More concretely, if $(\Omega, \mathscr{A}, \mathbb{P})$ describes throwing two fair dice, i.e. $\Omega := \{(i,k) : 1 \leqslant i, k \leqslant 6\}$, $\mathscr{A} = \mathscr{P}(\Omega)$ and $\mathbb{P}\{(i,k)\} = \frac{1}{36}$, we could ask for the total number of points thrown, $\xi : \Omega \to \{2, 3, \ldots, 12\}$, $\xi((i,k)) := i + k$, which is a measurable map. [✍] The law of ξ is then given in Table 7.1.

Table 7.1. *The law of* ξ

i	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(\xi = i)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

We close this section with some transformation formulae for Lebesgue measure. Recall that a matrix $T \in \mathbb{R}^{n \times n}$ is *orthogonal* if, and only if, $T^\top \cdot T = \mathrm{id}$. Orthogonal matrices preserve lengths and angles, i.e. we have for all $x, y \in \mathbb{R}^n$

$$\langle x, y\rangle = \langle Tx, Ty\rangle, \iff \|x\| = \|Tx\|, \tag{7.6}$$

where $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ and $\|x\|^2 = \langle x, x\rangle$ denote the usual Euclidean scalar product, resp. norm.

Theorem 7.9 *If* $T \in \mathbb{R}^{n \times n}$ *is an orthogonal matrix, then* $\lambda^n = T(\lambda^n)$.

Proof The matrix T induces a linear map $x \mapsto Tx$, $T(ax + by) = aTx + bTy$ for all $a, b \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$. The orthogonality relation (7.6) shows that T is an isometry,

$$\|Tx - Ty\| = \|T(x - y)\| = \|x - y\|,$$

and hence it is continuous and measurable by virtue of Example 7.3. In particular, the image measure $\mu(B) := \lambda^n(T^{-1}(B))$ is well-defined (by Theorem 7.6) and satisfies for all $x \in \mathbb{R}^n$

$$\mu(x + B) = \lambda^n(T^{-1}(x + B)) = \lambda^n(T^{-1}x + T^{-1}B) \stackrel{5.8}{=} \lambda^n(T^{-1}B) = \mu(B)$$

[2] We use the shorthand $\{\xi \in A\}$ for $\xi^{-1}(A)$ and $\mathbb{P}(\xi \in A)$ for $\mathbb{P}\{\xi \in A\}$.

and, again by Theorem 5.8, $\mu(B)=\kappa\lambda^n(B)$ for all $B\in\mathscr{B}(\mathbb{R}^n)$. To determine the constant κ we choose $B=B_1(0)$. Since $T\in\mathbb{R}^{n\times n}$ is orthogonal, (7.6) implies

$$B_1(0)=\{x:\|x\|<1\}=\{x:\|Tx\|<1\}=T^{-1}B_1(0)$$

and thus

$$\lambda^n(B_1(0))=\lambda^n(T^{-1}B_1(0))=\mu(B_1(0))=\kappa\lambda^n(B_1(0)).$$

As $0<\lambda^n(B_1(0))<\infty$, we have $\kappa=1$, and the theorem follows. □

Theorem 7.9 is a particular case of the following general *change-of-variable* formula for Lebesgue measure. Recall that a matrix $S\in\mathbb{R}^{n\times n}$ is invertible, if $\det S\neq 0$.

Theorem 7.10 *Let $S\in\mathbb{R}^{n\times n}$ be an invertible matrix. Then*

$$S(\lambda^n)=|\det S^{-1}|\lambda^n=|\det S|^{-1}\lambda^n. \tag{7.7}$$

Proof Since S is invertible, both S and S^{-1} are linear maps on $\mathbb{R}^n$, and as such they are continuous and measurable (Example 7.3). Set $\mu(B):=\lambda^n(S^{-1}(B))$ for $B\in\mathscr{B}(\mathbb{R}^n)$. Then we have for all $x\in\mathbb{R}^n$

$$\mu(x+B)=\lambda^n(S^{-1}(x+B))=\lambda^n(S^{-1}x+S^{-1}B)\overset{5.8}{=}\lambda^n(S^{-1}B)=\mu(B),$$

and from Theorem 5.8 we conclude that

$$\mu(B)=\mu\big([0,1)^n\big)\lambda^n(B)=\lambda^n\big(S^{-1}([0,1)^n)\big)\lambda^n(B).$$

From elementary geometry we know that $S^{-1}([0,1)^n)$ is a parallelepiped spanned by the vectors $S^{-1}e_i$, $i=1,2,\dots,n$, $e_i=(\underbrace{0,\dots,0,1}_{i},0,\dots)$. Its volume is

$$\mathrm{vol}^{(n)}\big(S^{-1}([0,1)^n)\big)=|\det S^{-1}|=|\det S|^{-1},$$

see also Appendix C. By Remark 6.8, $\mathrm{vol}^{(n)}=\lambda^n$ (at least on the Borel sets) and the proof is finished. □

Theorem 7.9 or Theorem 7.10 allows us to complete the characterization of Lebesgue measure announced earlier in Theorem 4.7. A *motion* is a linear transformation of the form

$$M_x=\tau_x\circ T,$$

where $\tau_x(y)=y-x$ is a translation and $T\in\mathbb{R}^{n\times n}$ is an orthogonal mapping ($T^\top\cdot T=\mathrm{id}_n$). In particular, congruent sets are connected by motions.

Corollary 7.11 *Lebesgue measure is invariant under motions: $\lambda^n = M(\lambda^n)$ for all motions M in $\mathbb{R}^n$. In particular, congruent sets have the same measure.*

Proof We know that M is of the form $\tau_x \circ T$. Since $\det T = \pm 1$, we get

$$M(\lambda^n) = \tau_x(T(\lambda^n)) \overset{7.10}{=} \tau_x(\lambda^n) \overset{7.10}{=} \lambda^n. \qquad \square$$

Problems

7.1. Use Lemma 7.2 to show that $\tau_x : y \mapsto y - x$, $x, y \in \mathbb{R}^n$, is $\mathscr{B}(\mathbb{R}^n)/\mathscr{B}(\mathbb{R}^n)$-measurable.

7.2. Show that Σ' defined in the proof of Lemma 7.2 is a σ-algebra.

7.3. Let $X = \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$. Show that

(i) $\mathscr{A} := \{A \subset \mathbb{Z} \mid \forall n > 0 : 2n \in A \iff 2n+1 \in A\}$ is a σ-algebra;

(ii) $T : \mathbb{Z} \to \mathbb{Z}$, $T(n) := n + 2$ is $\mathscr{A}/\mathscr{A}$-measurable and bijective, but T^{-1} is not measurable.

7.4. Let X be a set, let $(X_i, \mathscr{A}_i)$, $i \in I$, be arbitrarily many measurable spaces and let $T_i : X \to X_i$ be a family of maps.

(i) Show that for every $i \in I$ the smallest σ-algebra in X that makes T_i measurable is given by $T_i^{-1}(\mathscr{A}_i)$.

(ii) Show that $\sigma\big(\bigcup_{i \in I} T_i^{-1}(\mathscr{A}_i)\big)$ is the smallest σ-algebra in X that makes all T_i, $i \in I$, *simultaneously* measurable.

7.5. Let $(X, \mathscr{A})$ and $(X', \mathscr{A}')$ be measurable spaces and $T : X \to X'$.

(i) Show that $\mathbb{1}_{T^{-1}(A')}(x) = \mathbb{1}_{A'} \circ T(x) \quad \forall x \in X$.

(ii) T is measurable if, and only if, $\sigma(T) \subset \mathscr{A}$.

(iii) If T is measurable and ν is a finite measure on $(X, \mathscr{A})$, then $\nu \circ T^{-1}$ is a finite measure on $(X', \mathscr{A}')$. Does this remain true for σ-finite measures?

7.6. Let $T : X \to Y$ be any map. Show that $T^{-1}(\sigma(\mathscr{G})) = \sigma(T^{-1}(\mathscr{G}))$ holds for arbitrary families $\mathscr{G}$ of subsets of Y.

7.7. Let X be a set, let $(X_i, \mathscr{A}_i)$, $i \in I$, be arbitrarily many measurable spaces and let $T_i : X \to X_i$ be a family of maps. Show that a map f from a measurable space $(F, \mathscr{F})$ to $(X, \sigma(T_i : i \in I))$ is measurable if, and only if, all maps $T_i \circ f$ are $\mathscr{F}/\mathscr{A}_i$-measurable.

7.8. Use Problem 7.7 to show that a function $f : \mathbb{R}^n \to \mathbb{R}^m$, $x \mapsto (f_1(x), \dots, f_m(x))$ is measurable if, and only if, all coordinate maps $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \dots, m$, are measurable.
[Hint: show that the coordinate projections $x = (x_1, \dots, x_n) \mapsto x_i$ are measurable.]

7.9. Let $T : (X, \mathscr{A}) \to (X', \mathscr{A}')$ be a measurable map. Under which circumstances is the family of sets $T(\mathscr{A})$ a σ-algebra?

7.10. Use image measures to give a new proof of Problem 5.9, i.e. show that

$$\lambda^n(t \cdot B) = t^n \lambda^n(B) \qquad \forall B \in \mathscr{B}(\mathbb{R}^n),\ \forall t > 0.$$

7.11. Let $(X, \mathscr{A}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and let λ be one-dimensional Lebesgue measure.

(i) A point x with $\mu\{x\} > 0$ is an atom. Show that every measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ which has no atoms can be written as image measure of λ.
[Hint: μ has no atoms, so F_μ is continuous. Thus, $G = F^{-1}$ exists and can be made left-continuous. Finally $\mu[a, b) = F_\mu(b) - F_\mu(a) = \lambda(G^{-1}\{[a, b)\})$.]

(ii) Is (i) true for measures with atoms, say, $\mu = \delta_0$?
[Hint: determine $F_{\delta_0}^{-1}$. Is it measurable?]

7.12. Cantor's ternary set. Let $(X, \mathscr{A}) = \big([0,1], [0,1] \cap \mathscr{B}(\mathbb{R})\big)$, $\lambda = \lambda^1|_{[0,1]}$, and set $C_0 = [0,1]$. Remove the open middle third of C_0 to get $C_1 = J_1^0 \dot\cup J_1^1$. Remove the open middle thirds of J_1^i, $i = 0, 1$, to get $C_2 = J_2^{00} \dot\cup J_2^{01} \dot\cup J_2^{10} \dot\cup I_2^{11}$ and so forth.

(i) Make a sketch of C_0, C_1, C_2, C_3.

(ii) Prove that each C_n is compact. Conclude that $C := \bigcap_{n \in \mathbb{N}_0} C_n$ is non-void and compact.

(iii) The set C is known as the *Cantor set* or *Cantor's discontinuum*. It satisfies $C \cap \bigcup_{n \in \mathbb{N}} \bigcup_{k \in \mathbb{N}_0} \big(\frac{3k+1}{3^n}, \frac{3k+2}{3^n}\big) = \emptyset$.

(iv) Find the value of $\lambda(C_n)$ and show that $\lambda(C) = 0$.

(v) Show that C does not contain any open interval. Conclude that the interior (of the closure) of C is empty.
Remark. Sets with empty interior are called *nowhere dense*.

(vi) We can write $x \in [0,1)$ as a base-3 ternary fraction, i.e. $x = 0.x_1 x_2 x_3 \ldots$, where $x_i \in \{0,1,2\}$, which is short for $x = \sum_{i=1}^{\infty} x_i 3^{-i}$. (For example, $\frac{1}{3} = 0.1 = 0.02222\ldots$; note that this representation is not unique, [✍] which is important for this exercise.)
Show that $x \in C$ if, and only if, x has a ternary representation involving only 0s and 2s.
[Hint: the numbers in $\big(\frac{1}{3}, \frac{2}{3}\big)$, the first interval to be removed, are all of the form $0.1{*}{*}{*}\ldots$, i.e. they contain at least one '1', while in $\big[0, \frac{1}{3}\big]$ and $\big[\frac{2}{3}, 1\big]$ we have numbers of the form $0.0 * * * \ldots - 0.022222\ldots$ and $0.2 * * * \ldots - 0.2222\ldots$, respectively. The next step eliminates the $0.01 * * * \ldots$s and $0.21 * * * \ldots$s – etc.]

(vii) Use (vi) to show that C is not countable and even has the same cardinality as $[0,1]$. Nevertheless, $\lambda(C) = 0 \neq 1 = \lambda[0,1]$.

7.13. Let $\mathcal{E}, \mathscr{F} \subset \mathscr{P}(X)$ be two families of subsets of X. One *usually* uses the notation (as we do in this book)

$$\mathcal{E} \cup \mathscr{F} = \big\{A : A \in \mathcal{E} \text{ or } A \in \mathscr{F}\big\} \quad \text{and} \quad \mathcal{E} \cap \mathscr{F} = \big\{A : A \in \mathcal{E} \text{ and } A \in \mathscr{F}\big\}.$$

Let us, for this problem, also introduce the families

$$\mathcal{E} \Cup \mathscr{F} = \big\{E \cup F : E \in \mathcal{E},\ F \in \mathscr{F}\big\} \quad \text{and} \quad \mathcal{E} \Cap \mathscr{F} = \big\{E \cap F : E \in \mathcal{E},\ F \in \mathscr{F}\big\}.$$

Assume now that $\mathcal{E}$ and $\mathscr{F}$ are σ-Algebras.

(i) Show that $\mathcal{E} \Cup \mathscr{F} \supset \mathcal{E} \cup \mathscr{F}$ and $\mathcal{E} \Cap \mathscr{F} \supset \mathcal{E} \cup \mathscr{F}$.

(ii) Show that, in general, we have no equality in part (i).

(iii) Show that $\sigma(\mathcal{E} \Cup \mathscr{F}) = \sigma(\mathcal{E} \Cap \mathscr{F}) = \sigma(\mathcal{E} \cup \mathscr{F})$.

8

Measurable Functions

A *measurable function* is a measurable map $u: X \to \mathbb{R}$ from some measurable space $(X, \mathscr{A})$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Measurable functions will play a central rôle in the theory of integration. Recall that $u: X \to \mathbb{R}$ is $\mathscr{A}/\mathscr{B}(\mathbb{R})$-measurable[1] if

$$u^{-1}(B) \in \mathscr{A} \qquad \forall B \in \mathscr{B}(\mathbb{R}), \tag{8.1}$$

which is, due to Lemma 7.2, equivalent to

$$u^{-1}(G) \in \mathscr{A} \qquad \forall G \text{ from a generator } \mathscr{G} \text{ of } \mathscr{B}(\mathbb{R}). \tag{8.2}$$

As we have seen in Remark 3.9, $\mathscr{B}(\mathbb{R})$ is generated by all sets of the form $[a, \infty)$ (or (a, ∞) or $(-\infty, a)$ or $(-\infty, a]$) with $a \in \mathbb{R}$ or $a \in \mathbb{Q}$, and we need

$$u^{-1}([a, \infty)) = \{x \in X : u(x) \in [a, \infty)\} = \{x \in X : u(x) \geqslant a\} \in \mathscr{A}, \tag{8.3}$$

with similar expressions for the other types of intervals. Let us introduce the following useful shorthand notation:

$$\{u \geqslant v\} := \{x \in X : u(x) \geqslant v(x)\} \tag{8.4}$$

and $\{u > v\}, \{u \leqslant v\}, \{u < v\}, \{u = v\}, \{u \neq v\}, \{u \in B\}$, etc., which are defined in a similar fashion.

In this new notation measurability of functions reads as follows.

Lemma 8.1 *Let $(X, \mathscr{A})$ be a measurable space. The function $u: X \to \mathbb{R}$ is $\mathscr{A}/\mathscr{B}(\mathbb{R})$-measurable if, and only if, one, and hence all, of the following conditions hold:*

(i) $\{u \geqslant a\} \in \mathscr{A} \quad \forall a \in \mathbb{R}$ *or* $\mathbb{Q}$;

(ii) $\{u > a\} \in \mathscr{A} \quad \forall a \in \mathbb{R}$ *or* $\mathbb{Q}$;

(iii) $\{u \leqslant a\} \in \mathscr{A} \quad \forall a \in \mathbb{R}$ *or* $\mathbb{Q}$;

(iv) $\{u < a\} \in \mathscr{A} \quad \forall a \in \mathbb{R}$ *or* $\mathbb{Q}$.

[1] We will frequently drop the $\mathscr{B}(\mathbb{R})$ since $\mathbb{R}$ is naturally equipped with the Borel σ-algebra and just say that u is $\mathscr{A}$-measurable.

Proof Combine Remark 3.9 and Lemma 7.2. □

It is often helpful to use the values $+\infty$ and $-\infty$ in calculations. To do this properly, we have to consider the *extended real line* $\overline{\mathbb{R}} := [-\infty, +\infty]$. If we agree that $-\infty < x$ and $y < +\infty$ for all $x, y \in \mathbb{R}$, then $\overline{\mathbb{R}}$ inherits the ordering from $\mathbb{R}$ as well as the usual rules of addition and multiplication of elements from $\mathbb{R}$. The latter need to be augmented as shown in Table 8.1.

Table 8.1. *Addition and multiplication in* $\overline{\mathbb{R}}$*, where* $x, y \in \mathbb{R}$ *and* $a, b \in (0, \infty)$

$+$	0	x	$+\infty$	$-\infty$
0	0	x	$+\infty$	$-\infty$
y	y	$x+y$	$+\infty$	$-\infty$
$+\infty$	$+\infty$	$+\infty$	$+\infty$	$\nexists$
$-\infty$	$-\infty$	$-\infty$	$\nexists$	$-\infty$

$\cdot$	0	$\pm a$	$+\infty$	$-\infty$
0	0	0	0	0
$\pm b$	0	$a \cdot b$	$\pm\infty$	$\mp\infty$
$+\infty$	0	$\pm\infty$	$+\infty$	$-\infty$
$-\infty$	0	$\mp\infty$	$-\infty$	$+\infty$

Caution $\overline{\mathbb{R}}$ is not a field. Expressions of the form

$$\infty - \infty \quad \text{and} \quad \frac{\pm\infty}{\pm\infty}$$

must be avoided.[2]

The Borel σ-algebra $\mathscr{B}(\overline{\mathbb{R}})$ is defined by

$$B^* \in \mathscr{B}(\overline{\mathbb{R}}) \iff \begin{array}{l} B^* = B \cup S \quad \text{for some } B \in \mathscr{B}(\mathbb{R}) \text{ and} \\ S \in \{\emptyset, \{-\infty\}, \{+\infty\}, \{-\infty, +\infty\}\} \end{array} \tag{8.5}$$

and it is not hard to see that $\mathscr{B}(\overline{\mathbb{R}})$ is a σ-algebra whose trace w.r.t. $\mathbb{R}$ is $\mathscr{B}(\mathbb{R})$. [✍]

Lemma 8.2 $\mathscr{B}(\mathbb{R}) = \mathbb{R} \cap \mathscr{B}(\overline{\mathbb{R}}) \overset{def}{=} \{A \cap \mathbb{R} : A \in \mathscr{B}(\overline{\mathbb{R}})\}$.

Moreover,

Lemma 8.3 $\mathscr{B}(\overline{\mathbb{R}})$ *is generated by all sets of the form* $[a, \infty]$ (*or* $(a, \infty]$ *or* $[-\infty, a)$ *or* $[-\infty, a]$) *where* a *is from* $\mathbb{R}$ *or* $\mathbb{Q}$.

[2] Conventions are tricky. The rationale behind our definitions is to understand '$\pm\infty$' in every instance as the limit of some (possibly different each time) sequence, and '0' as a bona fide zero. Then $0 \cdot (\pm\infty) = 0 \cdot \lim_n a_n = \lim_n (0 \cdot a_n) = \lim_n 0 = 0$, while expressions of the type $\infty - \infty$ or $\frac{\pm\infty}{\pm\infty}$ become $\lim_n a_n - \lim_n b_n$ or $\frac{\lim_n a_n}{\lim_n b_n}$, where two sequences compete and do not lead to unique results.

Proof Set $\Sigma := \sigma(\{[a,\infty] : a \in \mathbb{R}\})$. Since

$$[a,\infty] = [a,\infty) \cup \{+\infty\} \quad \text{and} \quad [a,\infty) \in \mathscr{B}(\mathbb{R}),$$

we see that $[a,\infty] \in \mathscr{B}(\overline{\mathbb{R}})$ and $\Sigma \subset \mathscr{B}(\overline{\mathbb{R}})$. On the other hand,

$$[a,b) = [a,\infty] \setminus [b,\infty] \in \Sigma \quad \forall -\infty < a \leqslant b < \infty,$$

which means that $\mathscr{B}(\mathbb{R}) \subset \Sigma \subset \mathscr{B}(\overline{\mathbb{R}})$. Since also

$$\{+\infty\} = \bigcap_{j\in\mathbb{N}} [j,\infty], \quad \{-\infty\} = \bigcap_{j\in\mathbb{N}} [-\infty,-j) = \bigcap_{j\in\mathbb{N}} [-j,\infty]^c$$

we have $\{-\infty\}, \{+\infty\} \in \Sigma$, which entails that all sets of the form

$$B,\ B \cup \{+\infty\},\ B \cup \{-\infty\},\ B \cup \{-\infty,+\infty\} \in \Sigma, \qquad \forall B \in \mathscr{B}(\mathbb{R});$$

therefore, $\mathscr{B}(\overline{\mathbb{R}}) \subset \Sigma$.

The proofs for $a \in \mathbb{Q}$ and the other generating systems are similar. □

Definition 8.4 Let $(X,\mathscr{A})$ be a measurable space. We write $\mathcal{M} := \mathcal{M}(\mathscr{A})$ and $\mathcal{M}_{\overline{\mathbb{R}}} := \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$ for the families of (extended) real-valued measurable functions $u : (X,\mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and $u : (X,\mathscr{A}) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$.

Example 8.5 Let $(X,\mathscr{A})$ be a measurable space.

(i) The indicator function $f(x) := \mathbb{1}_A(x)$ is measurable if, and only if, $A \in \mathscr{A}$.
This follows easily from Lemma 8.1 and

$$\{\mathbb{1}_A > \lambda\} = \begin{cases} \emptyset, & \text{if } \lambda \geqslant 1, \\ A, & \text{if } 1 > \lambda \geqslant 0, \\ X, & \text{if } \lambda < 0. \end{cases}$$

(ii) Let $A_1, A_2, \ldots, A_M \in \mathscr{A}$ be mutually disjoint and $y_1, \ldots, y_M \in \mathbb{R}$. Then the function

$$f(x) := \sum_{j=m}^{M} y_m \mathbb{1}_{A_m}(x)$$

is measurable. This follows from Lemma 8.1 and the fact (Fig. 8.1) that

$$\{f > \lambda\} = \dot{\bigcup_{m : y_m > \lambda}} A_m \in \mathscr{A}.$$

The functions from Example 8.5(ii) are the building blocks for all measurable functions as well as for the definition of the integral.

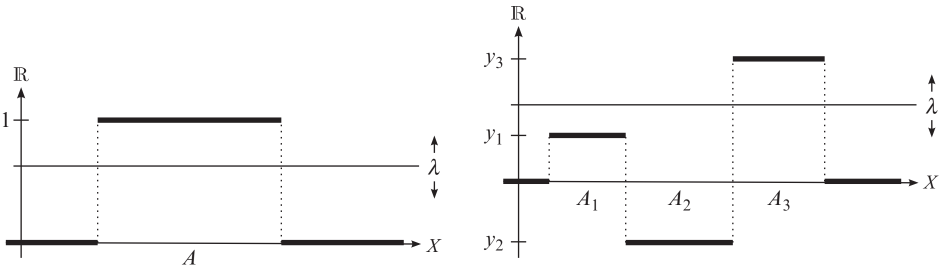


Fig. 8.1. Measurability of simple functions.

Definition 8.6 A *simple function* $f\colon X \to \mathbb{R}$ on a measurable space $(X, \mathscr{A})$ is a function of the form

$$f(x) = \sum_{m=1}^{M} y_m \mathbb{1}_{A_m}(x), \quad M \in \mathbb{N},\ y_m \in \mathbb{R},\ A_m \in \mathscr{A} \text{ disjoint.} \tag{8.6}$$

If we have

$$f(x) = \sum_{n=0}^{N} z_n \mathbb{1}_{B_n}(x), \quad N \in \mathbb{N},\ z_n \in \mathbb{R},\ B_n \in \mathscr{A},\ X = \bigcup_{n=0}^{N}{}^{\cdot}\, B_n, \tag{8.7}$$

then (8.7) is called a *standard representation* of f. The set of simple functions is denoted by $\mathcal{E}$ or $\mathcal{E}(\mathscr{A})$.

Caution The representations (8.6) and (8.7) are *not unique*.

Example 8.5 (continued)

(iii) A measurable function $h : X \to \mathbb{R}$ which attains only finitely many values is a simple function.

Indeed: the sets $\{h = \alpha\}$ with $\alpha \in h(X) = \{y_0, \dots, y_M\}$ are mutually disjoint and satisfy

$$\{h = \alpha\} = \{h \leqslant \alpha\} \setminus \{h < \alpha\} \in \mathscr{A} \quad \text{and} \quad \bigcup_{\alpha \in h(X)}{}^{\cdot}\, \{h = \alpha\} = X.$$

Thus $h(x) = \sum_{\alpha \in h(X)} \alpha \mathbb{1}_{\{h=\alpha\}}(x)$ is a standard representation.

Since every simple function attains only finitely many values, this shows that every simple function has at least one standard representation. In particular, $\mathcal{E}(\mathscr{A}) \subset \mathcal{M}(\mathscr{A})$ consists of measurable functions.

(iv) $f, g \in \mathcal{E}(\mathscr{A}) \implies f \pm g,\ f \cdot g \in \mathcal{E}(\mathscr{A})$.

Indeed: let $f = \sum_{m=0}^{M} y_m \mathbb{1}_{A_m}$ and $g = \sum_{n=0}^{N} z_n \mathbb{1}_{B_n}$ be standard representations of f and g.

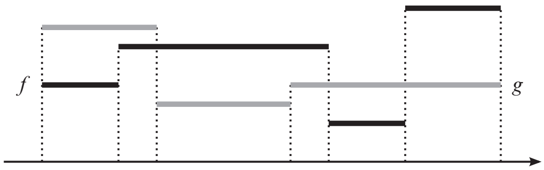


Fig. 8.2. For Example 8.5(iv).

It is not hard to see (use Fig. 8.2!) that

$$f \pm g = \sum_{m=0}^{M}\sum_{n=0}^{N}(y_m \pm z_n)\mathbb{1}_{A_m\cap B_n} \quad \text{and} \quad f\cdot g = \sum_{m=0}^{M}\sum_{n=0}^{N} y_m z_n \mathbb{1}_{A_m\cap B_n}$$

and that $(A_m \cap B_n) \cap (A_{m'} \cap B_{n'}) = \emptyset$ whenever $(m,n) \neq (m',n')$. After relabelling and merging the double indexation into a single index, this shows that $f \pm g$, $f \cdot g \in \mathcal{E}(\mathscr{A})$.

Notice that $(A_m \cap B_n)_{m,n}$ is the common refinement of the partitions $(A_m)_m$ and $(B_n)_n$ and that on each set $A_m \cap B_n$ the functions f and g do not change their respective values.

(v) $f \in \mathcal{E}(\mathscr{A}) \implies f^+, f^- \in \mathcal{E}(\mathscr{A})$. See Fig. 8.3.

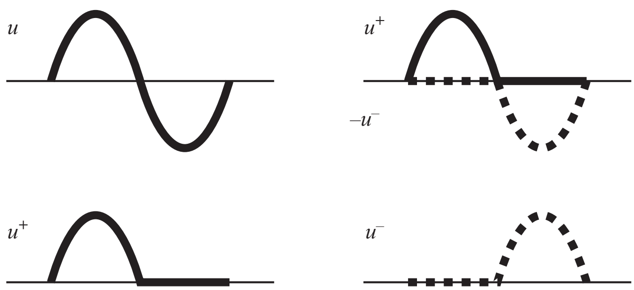


Fig. 8.3. The positive part, $u^+(x) := \max\{u(x), 0\}$, and the negative part, $u^-(x) := -\min\{u(x), 0\}$, of a function $u : X \to \overline{\mathbb{R}}$.

Obviously,

$$f = f^+ - f^- \quad \text{and} \quad |f| = f^+ + f^-. \tag{8.8}$$

(vi) $f \in \mathcal{E}(\mathscr{A}) \implies |f| \in \mathcal{E}(\mathscr{A})$.

Our next theorem reveals the fundamental rôle of simple functions.

Theorem 8.8 (sombrero lemma) *Let $(X, \mathscr{A})$ be a measurable space. Every positive $\mathscr{A}/\mathscr{B}(\overline{\mathbb{R}})$-measurable function $u : X \to [0, \infty]$ is the pointwise limit of an* *increasing sequence of simple functions $f_n \in \mathcal{E}(\mathscr{A})$, $f_n \geqslant 0$, i.e.*

$$u(x) = \sup_{n\in\mathbb{N}} f_n(x) = \lim_{n\to\infty} f_n(x), \quad f_1 \leqslant f_2 \leqslant f_3 \leqslant \cdots . \tag{8.9}$$

Proof Fix $n \in \mathbb{N}$ and define level sets

$$A_k^n := \begin{cases} \{k2^{-n} \leqslant u < (k+1)2^{-n}\} & \text{for } k = 0, 1, 2, \ldots, n2^n - 1, \\ \{u \geqslant n\} & \text{for } k = n2^n, \end{cases}$$

which slice up the graph of u *horizontally* as shown in the picture. The approximating simple functions are

$$f_n(x) := \sum_{k=0}^{n2^n} k2^{-n} \mathbb{1}_{A_k^n}(x)$$

and from Fig. 8.3 it is easy to see that

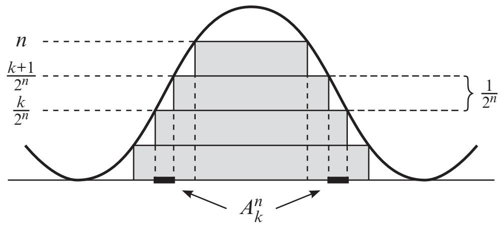


Fig. 8.4. The function u sits like a 'Mexican hat' over the approximating simple functions.

- $0 \leqslant f_n \leqslant f_{n+1} \leqslant u$ and $f_n \uparrow u$;
- $|f_n(x) - u(x)| \leqslant 2^{-n}$ if $x \in \{u < n\}$;
- $A_k^n = \{k2^{-n} \leqslant u\} \cap \{u < (k+1)2^{-n}\}$, $\{u \geqslant n\} \in \mathscr{A}$. □

The following extension of Theorem 8.8 is almost immediate.

Corollary 8.9 *Let $(X, \mathscr{A})$ be a measurable space. Every $\mathscr{A}/\mathscr{B}(\overline{\mathbb{R}})$-measurable function $u : X \to \overline{\mathbb{R}}$ is the pointwise limit of simple functions $f_n \in \mathcal{E}(\mathscr{A})$ such that $|f_n| \leqslant |u|$. If u is bounded, the limit is uniform.*

Proof Write $u = u^+ - u^-$ and observe that

$$\{u^+ > \lambda\} = \begin{cases} \{u > \lambda\} & \text{if } \lambda \geqslant 0, \\ X & \text{if } \lambda < 0. \end{cases}$$

Since $u^- = (-u)^+$, we have $\{u^\pm > \lambda\} \in \mathscr{A}$ for all $\lambda \in \overline{\mathbb{R}}$. Thus $u^\pm$ are positive measurable functions, and we can construct, as in Theorem 8.8, simple functions $g_n \uparrow u^+$ and $h_n \uparrow u^-$.

Clearly, $f_n := g_n - h_n \to u^+ - u^- = u$ and $|f_n| = g_n + h_n \leqslant u^+ + u^- = |u|$.

If u is bounded, say $|u(x)| \leqslant c$ for all x, then the proof of Theorem 8.8 shows that we have for all $x \in X$ and $n > c$

$$|f_n(x) - u(x)| \leqslant |g_n(x) - u^+(x)| + |h_n(x) - u^-(x)| \leqslant 2 \cdot 2^{-n}. \qquad \square$$

Corollary 8.10 *Let $(X, \mathscr{A})$ be a measurable space. If $u_n : X \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, are measurable functions, then so are*

$$\sup_{n\in\mathbb{N}} u_n, \quad \inf_{n\in\mathbb{N}} u_n, \quad \limsup_{n\to\infty} u_n, \quad \liminf_{n\to\infty} u_n$$

and, whenever it exists in $\overline{\mathbb{R}}$, $\lim_{n\to\infty} u_n$.

Before we prove Corollary 8.10, let us stress again that expressions of the type $\sup_{n\in\mathbb{N}} u_n$ or $u_n \to u$, etc. are always understood in a *pointwise, x-by-x sense*, i.e. they are short for $\sup_{n\in\mathbb{N}} u_n(x) := \sup\{u_n(x) : n \in \mathbb{N}\}$ or $\lim_{n\to\infty} u_n(x) = u(x)$ at each x (or for a specified range).

The infimum 'inf' and supremum 'sup' are familiar from calculus. Recall the useful formula

$$\inf_{n\in\mathbb{N}} u_n(x) = -\sup_{n\in\mathbb{N}}(-u_n(x)), \tag{8.10}$$

which allows us to express an inf as a sup, and vice versa. Recall also the definition of the *lower* resp. *upper limits* lim inf and lim sup,

$$\liminf_{n\to\infty} u_n(x) := \sup_{k\in\mathbb{N}} \left(\inf_{n\geqslant k} u_n(x) \right) = \lim_{k\to\infty} \left(\inf_{n\geqslant k} u_n(x) \right), \tag{8.11}$$

$$\limsup_{n\to\infty} u_n(x) := \inf_{k\in\mathbb{N}} \left(\sup_{n\geqslant k} u_n(x) \right) = \lim_{k\to\infty} \left(\sup_{n\geqslant k} u_n(x) \right); \tag{8.12}$$

more details can be found in Appendix A.

In the extended real line $\overline{\mathbb{R}}$ the upper and lower limits *always* exist – but they may attain the values $+\infty$ and $-\infty$ – and we have

$$\liminf_{n\to\infty} u_n(x) \leqslant \limsup_{n\to\infty} u_n(x). \tag{8.13}$$

Moreover, $\lim_{n\to\infty} u_n(x)$ exists [and is finite] if, and only if, upper and lower limits coincide, $\liminf_{n\to\infty} u_n(x) = \limsup_{n\to\infty} u_n(x)$ [and are finite]; in this case all three limits have the same value.

Proof of Corollary 8.10 We show that $\sup_n u_n$ and $(-1)u = -u$ (for a measurable function u) are again measurable. Observe that for all $a \in \mathbb{R}$

$$\left\{ \sup_{n\in\mathbb{N}} u_n > a \right\} = \bigcup_{n\in\mathbb{N}} \underbrace{\{u_n > a\}}_{\in\mathscr{A}} \in \mathscr{A}.$$

The inclusion '$\supset$' is trivial since $a < u_n(x) \leqslant \sup_{n\in\mathbb{N}} u_n(x)$ always holds; the direction '$\subset$' follows by contradiction: if $u_n(x) \leqslant a$ for all $n \in \mathbb{N}$, then also $\sup_{n\in\mathbb{N}} u_n(x) \leqslant a$. This proves the measurability of $\sup_{n\in\mathbb{N}} u_n$.

If u is measurable, we have for all $a \in \mathbb{R}$

$$\{-u > a\} = \{u < -a\} \in \mathscr{A},$$

which shows that $-u$ is also measurable.

The measurability of $\inf_{n\in\mathbb{N}} u_n$, $\liminf_{n\to\infty} u_n$ and $\limsup_{n\to\infty} u_n$ follows now from formulae (8.10)–(8.12), which can be written down in terms of $\sup_n$s and several multiplications by (-1). If $\lim_{n\to\infty} u_n$ exists, then it coincides with $\liminf_{n\to\infty} u_n = \limsup_{n\to\infty} u_n$ and inherits their measurability. $\square$

Corollary 8.11 *Let $u, v : X \to \overline{\mathbb{R}}$ be $\mathscr{A}/\mathscr{B}(\overline{\mathbb{R}})$-measurable functions. Then*

$$u \pm v, \quad u \cdot v, \quad u \vee v := \max\{u, v\}, \quad u \wedge v := \min\{u, v\} \tag{8.14}$$

are $\mathscr{A}/\mathscr{B}(\overline{\mathbb{R}})$-measurable functions (whenever they are defined). See Fig. 8.5.

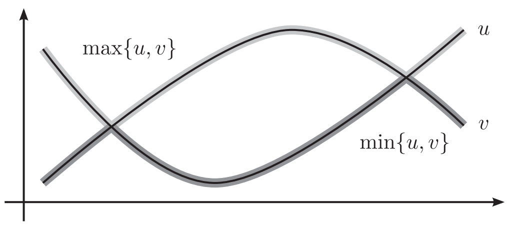


Fig. 8.5. The maximum and minimum of two functions is always meant pointwise, i.e.

$$(u \vee v)(x) = \max\{u(x), v(x)\} \overset{[✍]}{=} \tfrac{1}{2}\big(u(x) + v(x) + |u(x) - v(x)|\big),$$

$$(u \wedge v)(x) = \min\{u(x), v(x)\} \overset{[✍]}{=} \tfrac{1}{2}\big(u(x) + v(x) - |u(x) - v(x)|\big).$$

Proof of Corollary 8.11 If $u, v \in \mathcal{E}$ are simple functions, all functions in (8.14) are again simple [✍] and, therefore, measurable. For general $u, v \in \mathcal{M}_{\overline{\mathbb{R}}}$ choose sequences $(f_n)_{n\in\mathbb{N}}, (g_n)_{n\in\mathbb{N}} \subset \mathcal{E}$ of simple functions such that $f_n \to u$ and $g_n \to v$ for $n \to \infty$. The claim now follows from the usual rules for limits. $\square$

Corollary 8.12 *A function u is $\mathscr{A}/\mathscr{B}(\overline{\mathbb{R}})$-measurable if, and only if, the positive and negative parts $u^{\pm}$ are $\mathscr{A}/\mathscr{B}(\overline{\mathbb{R}})$-measurable.*

Corollary 8.13 *If $u, v : X \to \overline{\mathbb{R}}$ are $\mathscr{A}/\mathscr{B}(\overline{\mathbb{R}})$-measurable functions, then*

$$\{u < v\}, \quad \{u \leqslant v\}, \quad \{u = v\}, \quad \{u \neq v\} \in \mathscr{A}.$$

Let us finally show two results which are frequently used in applications. When reading this text for the first time, they can safely be omitted.

The first result elicits the structure of $\sigma(T)$-measurable functions.

∗Lemma 8.14 (factorization lemma) *Let $T:(X,\mathscr{A})\to(X',\mathscr{A}')$ be a measurable map.*

$$\left.\begin{array}{r} u:X\to\overline{\mathbb{R}}\ \textit{is}\\ \sigma(T)/\mathscr{B}(\overline{\mathbb{R}})\textit{-measurable}\end{array}\right\} \iff \left\{\begin{array}{l} u=w\circ T\ \textit{for a function}\\ w:(X',\mathscr{A}')\xrightarrow{\textit{measurable.}}(\overline{\mathbb{R}},\mathscr{B}(\overline{\mathbb{R}})).\end{array}\right.$$

Proof '⇐' Since T is $\sigma(T)$-measurable, we see that the composition

$$w\circ T:(X,\sigma(T))\xrightarrow[\text{measurable}]{T}(X',\mathscr{A}')\xrightarrow[\text{measurable}]{w}(\overline{\mathbb{R}},\mathscr{B}(\overline{\mathbb{R}}))$$

is measurable.

'⇒' Assume that u is $\sigma(T)$-measurable. We have to find a suitable function w.

Step 1. $u=\mathbb{1}_A$. We have

$$\begin{aligned} A\in\sigma(T) &\iff \exists A'\in\mathscr{A}':A=T^{-1}(A')\\ &\implies u=\mathbb{1}_A=\mathbb{1}_{T^{-1}(A')}=\mathbb{1}_{A'}\circ T.\end{aligned}$$

This shows that $w=\mathbb{1}_{A'}$.

Step 2. $u\in\mathcal{E}(\sigma(T))$. Using a standard representation of u, we find with Step 1 and linearity a corresponding $w\in\mathcal{E}(\mathscr{A}')$.

Step 3. $u\in\mathcal{M}_{\overline{\mathbb{R}}}(\sigma(T))$. By the sombrero lemma (Corollary 8.9) there is a sequence $f_n\in\mathcal{E}(\sigma(T))$ such that $u=\lim_{n\to\infty}f_n$. From Step 2 we find simple functions $w_n\in\mathcal{E}(\mathscr{A}')$ such that $f_n=w_n\circ T$. Setting $w:=\liminf_{n\to\infty}w_n$, we get $w\in\mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A}')$ and

$$w\circ T=\Big(\liminf_{n\to\infty}w_n\Big)\circ T=\lim_{n\to\infty}\overbrace{(w_n\circ T)}^{=f_n}=u. \qquad \square$$

The second result is a version of the monotone class theorem (see Problem 3.14) for functions.

∗Theorem 8.15 (monotone class theorem) *Let $\mathscr{G}\subset\mathscr{P}(X)$ be a $\cap$-stable family and $\mathcal{V}$ be a vector space of functions $u:X\to\mathbb{R}$ such that*

(i) *$\mathbb{1}\in\mathcal{V}$ and $\mathbb{1}_G\in\mathcal{V}$ for all $G\in\mathscr{G}$;*
(ii) *for every sequence $0\leqslant u_1\leqslant u_2\leqslant\cdots$, $u_n\in\mathcal{V}$, with $u(x):=\sup_{n\in\mathbb{N}}u_n(x)<\infty$ for all $x\in X$, we have $u\in\mathcal{V}$.*

Under these assumptions $\mathcal{V}\supset\mathcal{M}(\sigma(\mathscr{G}))$.

Proof Set $\mathscr{D}:=\{A\subset X:\mathbb{1}_A\in\mathcal{V}\}$. Let us show that $\mathscr{D}$ is a Dynkin system.

(D1) We have $X\in\mathscr{D}$ since $\mathbb{1}=\mathbb{1}_X\in\mathcal{V}$.
(D2) If $A\in\mathscr{D}$, then $\mathbb{1}_A\in\mathcal{V}$. Since $\mathcal{V}$ is a vector space containing the constant function $\mathbb{1}$, we get $\mathbb{1}_{A^c}=\mathbb{1}-\mathbb{1}_A\in\mathcal{V}$, and so $A^c\in\mathscr{D}$.

(D3) Let $(A_n)_{n\in\mathbb{N}} \subset \mathscr{D}$ be a sequence of mutually disjoint sets. By assumption, $\mathbb{1}_{A_n} \in \mathcal{V}$ and $u_n := \mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n}$ is an increasing sequence of positive functions in $\mathcal{V}$. By (ii), $\sup_{n\in\mathbb{N}} u_n = \mathbb{1}_{\bigcup_{n\in\mathbb{N}} \mathbb{1}_{A_n}} \in \mathcal{V}$, i.e. $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{D}$.

Since $\mathscr{G} \subset \mathscr{D}$ is $\cap$-stable, we know from Theorem 5.5 that $\sigma(\mathscr{G}) \subset \mathscr{D}$. This means, in particular, that all simple functions $\mathcal{E}(\sigma(\mathscr{G}))$ are contained in $\mathcal{V}$.

Since every positive, real-valued $u \in \mathcal{M}(\sigma(\mathscr{G}))$ can be approximated by an increasing sequence of $\sigma(\mathscr{G})$-measurable simple functions, see Theorem 8.8, we conclude with property (ii) that $\mathcal{M}^+(\sigma(\mathscr{G})) \subset \mathcal{V}$.

Finally, on decomposing $u \in \mathcal{M}(\sigma(\mathscr{G}))$ into its positive and negative parts, we get $\mathcal{M}(\sigma(\mathscr{G})) \subset \mathcal{V}$. □

Problems

8.1. Show *directly* that condition (i) of Lemma 8.1 is equivalent to one of the conditions (ii), (iii) and (iv).

8.2. Verify that $\mathscr{B}(\overline{\mathbb{R}})$ defined in (8.5) is a σ-algebra. Show that $\mathscr{B}(\mathbb{R}) = \mathbb{R} \cap \mathscr{B}(\overline{\mathbb{R}})$.

8.3. Let $(X, \mathscr{A})$ be a measurable space.

(i) Let $f, g : X \to \mathbb{R}$ be measurable functions. Show that for every $A \in \mathscr{A}$ the function $h(x) := f(x)$, if $x \in A$, and $h(x) := g(x)$, if $x \notin A$, is measurable.

(ii) Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions and let $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ such that $\bigcup_{n\in\mathbb{N}} A_n = X$. Suppose that $f_n|_{A_n\cap A_k} = f_k|_{A_n\cap A_k}$ for all $k, n \in \mathbb{N}$ and set $f(x) := f_n(x)$ if $x \in A_n$. Show that $f : X \to \mathbb{R}$ is measurable.

8.4. Let $(X, \mathscr{A})$ be a measurable space and let $\mathscr{B} \subsetneq \mathscr{A}$ be a sub-σ-algebra. Show that $\mathcal{M}(\mathscr{B}) \subsetneq \mathcal{M}(\mathscr{A})$.

8.5. Show that $f \in \mathcal{E}$ implies that $f^{\pm} \in \mathcal{E}$. Is the converse true?

8.6. Show that for every real-valued function $u = u^+ - u^-$ and $|u| = u^+ + u^-$.

8.7. Show that every continuous function $u : \mathbb{R} \to \mathbb{R}$ is $\mathscr{B}(\mathbb{R})/\mathscr{B}(\mathbb{R})$-measurable. [Hint: check that for continuous functions $\{u > \alpha\}$ is an open set.]

8.8. Show that $x \mapsto \max\{x, 0\}$ and $x \mapsto \min\{x, 0\}$ are continuous, and by virtue of Problem 8.7 or Example 7.3, measurable functions from $\mathbb{R} \to \mathbb{R}$. Conclude that on any measurable space $(X, \mathscr{A})$ positive and negative parts $u^{\pm}$ of a measurable function $u : X \to \mathbb{R}$ are measurable.

8.9. Let $(f_i)_{i\in I}$ be arbitrarily many maps $f_i : X \to \mathbb{R}$. Show that

(i) $\{\sup_i f_i > \lambda\} = \bigcup_i \{f_i > \lambda\}$; (ii) $\{\sup_i f_i < \lambda\} \subset \bigcap_i \{f_i < \lambda\}$;
(iii) $\{\sup_i f_i \geqslant \lambda\} \supset \bigcup_i \{f_i \geqslant \lambda\}$; (iv) $\{\sup_i f_i \leqslant \lambda\} = \bigcap_i \{f_i \leqslant \lambda\}$;
(v) $\{\inf_i f_i > \lambda\} \subset \bigcap_i \{f_i > \lambda\}$; (vi) $\{\inf_i f_i < \lambda\} = \bigcup_i \{f_i < \lambda\}$;
(vii) $\{\inf_i f_i \geqslant \lambda\} = \bigcap_i \{f_i \geqslant \lambda\}$; (viii) $\{\inf_i f_i \leqslant \lambda\} \supset \bigcup_i \{f_i \leqslant \lambda\}$.

8.10. Check that the approximating sequence $(f_n)_{n\in\mathbb{N}}$ for u in Theorem 8.8 consists of $\sigma(u)$-measurable functions.

8.11. Complete the proofs of Corollaries 8.12 and 8.13.

8.12. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and assume that $\mathscr{A}$ is generated by a Boolean algebra $\mathscr{G}$, i.e. a family of subsets of X which is stable under finite unions, intersections and complementation and contains $\emptyset$. Show that every $u \in \mathcal{M}(\mathscr{A})$ can be approximated by a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathcal{E}(\mathscr{G})$ such that $\lim_{n\to\infty} f_n(x) = u(x)$ for all $x \in N^c$, where $N \in \mathscr{A}$ has measure zero: $\mu(N) = 0$. (The set N may, and will in general, depend on u).

[Hint: use the result from Problem 5.12 and combine it with Theorem 8.8 or Theorem 8.15.]

Remark. If μ is Lebesgue measure on $X = \mathbb{R}$, $\mathscr{A} = \mathscr{B}(\mathbb{R})$, we can use $\mathscr{G}$ as the family of finite unions of all half-open intervals $[a, b)$, $-\infty \leqslant a \leqslant b \leqslant +\infty$. It is not hard to see that this is a Boolean algebra. The problem tells us that we can approximate any measurable function *up to a null set*(!) by step-functions (with finitely many steps) with basis sets of the form $[a, b)$. You should compare this result with the comparison results between the Lebesgue and Riemann integrals, see Chapter 12: in some sense this leads to a Riemann-sum approximation of a Lebesgue integral.

8.13. Let $u: \mathbb{R} \to \mathbb{R}$ be differentiable. Explain why u and $u' = du/dx$ are measurable.

8.14. Find $\sigma(u)$, i.e. the σ-algebra generated by u, for the following functions:

$$f, g, h: \mathbb{R} \to \mathbb{R}, \qquad \text{(i)}\quad f(x) = x; \qquad \text{(ii)}\quad g(x) = x^2; \qquad \text{(iii)}\quad h(x) = |x|;$$
$$F, G: \mathbb{R}^2 \to \mathbb{R}, \qquad \text{(iv)}\quad F(x, y) = x + y; \qquad \text{(v)}\quad G(x, y) = x^2 + y^2.$$

[Hint: under f, g, h the pre-images of intervals are (unions of) intervals, under F we get strips in the plane, under G annuli and discs.]

8.15. Consider $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and $u: \mathbb{R} \to \mathbb{R}$. Show that $\{x\} \in \sigma(u)$ for all $x \in \mathbb{R}$ if, and only if, u is injective.

8.16. Let λ be one-dimensional Lebesgue measure. Find $\lambda \circ u^{-1}$, if $u(x) = |x|$.

8.17. Let $E \in \mathscr{B}(\mathbb{R})$, $Q: E \to \mathbb{R}$, $Q(x) = x^2$, and $\lambda_E := \lambda(E \cap \cdot)$ (Lebesgue measure).

(i) Show that Q is $\mathscr{B}(E)/\mathscr{B}(\mathbb{R})$-measurable.
(ii) Find $\nu \circ Q^{-1}$ for $E = [0, 1]$, $\nu = \lambda_E$ and $E = [-1, 1]$, $\nu = \frac{1}{2}\lambda_E$.

8.18. Let $u: \mathbb{R} \to \mathbb{R}$ be measurable. Which of the following functions are measurable:

$$u(x - 2), \qquad e^{u(x)}, \qquad \sin(u(x) + 8), \qquad u''(x), \qquad \operatorname{sgn} u(x - 7)?$$

8.19. Check the following. In the proof of the factorization lemma (Lemma 8.14) we cannot, in general, replace $\liminf_n w_n$ by $\lim_n w_n$.

[Hint: find a sequence $(w_n)_n$ and a map T such that $(w_n \circ T)_n$ is convergent and $(w_n)_n$ divergent.]

8.20. One can show that there are non-Borel measurable sets $A \subset \mathbb{R}$, see Appendix G. Taking this fact for granted, show that measurability of $|u|$ does not, in general, imply measurability of u. (The converse is, of course, true: measurability of u always guarantees that of $|u|$.)

8.21. Show that every increasing function $u: \mathbb{R} \to \mathbb{R}$ is $\mathscr{B}(\mathbb{R})/\mathscr{B}(\mathbb{R})$-measurable. Under which additional condition(s) do we have $\sigma(u) = \mathscr{B}(\mathbb{R})$?

[Hint: show that $\{u < \lambda\}$ is an interval by distinguishing three cases: u is continuous and strictly increasing when passing the level λ; and u jumps over the level λ; and u is 'flat' at level λ. Draw a picture of these situations.]

8.22. Let $\mathscr{A} = \sigma(\mathscr{G})$ be a σ-algebra on X, where $\mathscr{G} = \{G_i : i \in \mathbb{N}\}$ is a countable generator. Let $g := \sum_{i=1}^{\infty} 2^{-i} \mathbb{1}_{G_i}$. Show that $\sigma(g) = \mathscr{A}$.

8.23. Show that any left- or right-continuous function $u: \mathbb{R} \to \mathbb{R}$ is Borel measurable.

8.24. Show that every linear map $f: \mathbb{R}^n \to \mathbb{R}^m$ is $\mathscr{B}(\mathbb{R}^n)/\mathscr{B}(\mathbb{R}^m)$-measurable. Provide an example in which measurability of linear functions may fail if we use the completions of the Borel σ-algebras.

[Hint: Think of suitable *subsets* of Borel null sets, see Problems 4.15 and 6.7.]

8.25. Let $(\Omega, \mathscr{A})$ be a measurable space and $\xi : \mathbb{R} \times \Omega \to \mathbb{R}$ be a map such that $\omega \mapsto \xi(t, \omega)$ is $\mathscr{A}/\mathscr{B}(\mathbb{R})$-measurable and $t \mapsto \xi(t, \omega)$ is left- (or right-)continuous. Show that the following functions are measurable:

$$t \mapsto \xi(t, \omega) \quad \text{and} \quad \omega \mapsto \sup_{t \in \mathbb{R}} \xi(t, \omega).$$

[Hint: approximate $t \mapsto \xi(t, \omega)$ by simple functions.]

8.26. Let $(X, \mathscr{A}, \mu)$ be a measure space and $(X, \overline{\mathscr{A}}, \bar{\mu})$ its completion (see Problem 4.15). Show that a function $\phi : X \to \mathbb{R}$ is $\overline{\mathscr{A}}/\mathscr{B}(\mathbb{R})$-measurable if, and only if, there are $\mathscr{A}/\mathscr{B}(\mathbb{R})$-measurable functions $f, g : X \to \mathbb{R}$ such that $f \leqslant \phi \leqslant g$ and $\mu\{f \neq g\} = 0$.
[Hint: use simple functions and the sombrero lemma.]

9

Integration of Positive Functions

Throughout this chapter $(X, \mathscr{A}, \mu)$ will be some measure space. Recall that $\mathcal{M}^+(\mathscr{A})$ [$\mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$] are the $\mathscr{A}$-measurable positive functions [with values in $\overline{\mathbb{R}}$] and $\mathcal{E}(\mathscr{A})$ [$\mathcal{E}^+(\mathscr{A})$] are the [positive] simple functions.

The fundamental idea of *integration* is to measure the area between the graph of a function and the abscissa. For a positive simple function $f \in \mathcal{E}^+(\mathscr{A})$ in standard representation[1] this is easily done:

$$\text{if} \quad f = \sum_{i=0}^{M} y_i \mathbb{1}_{A_i} \in \mathcal{E}^+(\mathscr{A}) \qquad \text{then} \ \sum_{i=0}^{M} y_i \mu(A_i) \tag{9.1}$$

should be the μ-area enclosed by the graph and the abscissa.

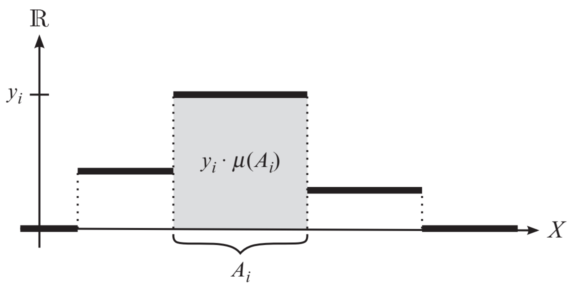


Fig. 9.1. The integral of a positive simple function.

There is only the problem that (9.1) might depend on the particular (standard) representation of f – and this should not happen.

[1] in the sense of Definition 8.6. By Example 8.5(iii) every $f \in \mathcal{E}(\mathscr{A})$ has a standard representation.

Lemma 9.1 *Let $\sum_{i=0}^{M} y_i \mathbb{1}_{A_i} = \sum_{k=0}^{N} z_k \mathbb{1}_{B_k}$ be two standard representations of the same function $f \in \mathcal{E}^+(\mathscr{A})$. Then*

$$\sum_{i=0}^{M} y_i \mu(A_i) = \sum_{k=0}^{N} z_k \mu(B_k).$$

Proof Since $A_0 \cupdot A_1 \cupdot \cdots \cupdot A_M = X = B_0 \cupdot B_1 \cupdot \cdots \cupdot B_N$, we get

$$A_i = \biguplus_{k=0}^{N} (A_i \cap B_k) \quad \text{and} \quad B_k = \biguplus_{i=0}^{M} (B_k \cap A_i).$$

Using the (finite) additivity of μ we see that

$$\sum_{i=0}^{M} y_i \mu(A_i) = \sum_{i=0}^{M} y_i \sum_{k=0}^{N} \mu(A_i \cap B_k) = \sum_{i=0}^{M} \sum_{k=0}^{N} y_i \mu(A_i \cap B_k) \tag{9.2}$$

(since all y_i are positive, the above sums always exist in $[0, \infty]$). Similarly,

$$\sum_{k=0}^{N} z_k \mu(B_k) = \sum_{k=0}^{N} z_k \sum_{i=0}^{M} \mu(A_i \cap B_k) = \sum_{k=0}^{N} \sum_{i=0}^{M} z_k \mu(A_i \cap B_k). \tag{9.3}$$

But $y_i = z_k$ whenever $A_i \cap B_k \neq \emptyset$, while for disjoint sets $A_i \cap B_k = \emptyset$ we have $\mu(A_i \cap B_k) = \mu(\emptyset) = 0$. Therefore,

$$y_i \mu(A_i \cap B_k) = z_k \mu(A_i \cap B_k) \qquad \forall (i, k),$$

and (9.2) and (9.3) have the same value. $\square$

Lemma 9.1 justifies the following definition based on (9.1).

Definition 9.2 Let $f = \sum_{i=0}^{M} y_i \mathbb{1}_{A_i} \in \mathcal{E}^+(\mathscr{A})$ be a simple function in standard representation. Then the number

$$I_\mu(f) := \sum_{i=0}^{M} y_i \mu(A_i) \in [0, \infty]$$

(which is independent of the representation of f) is called the (μ-)*integral* of f.

Properties 9.3 (of $I_\mu : \mathcal{E}^+(\mathscr{A}) \to [0, \infty]$) Let $f, g \in \mathcal{E}^+(\mathscr{A})$. Then

(i) $I_\mu(\mathbb{1}_A) = \mu(A) \qquad \forall A \in \mathscr{A}$;
(ii) $I_\mu(\lambda f) = \lambda I_\mu(f) \qquad \forall \lambda \geqslant 0$ (*positive homogeneous*);
(iii) $I_\mu(f + g) = I_\mu(f) + I_\mu(g)$ (*additive*);
(iv) $f \leqslant g \implies I_\mu(f) \leqslant I_\mu(g)$ (*monotone*).

Proof (i) and (ii) are obvious from the definition of I_μ.

(iii) Take standard representations

$$f = \sum_{i=0}^{M} y_i \mathbb{1}_{A_i} \quad \text{and} \quad g = \sum_{k=0}^{N} z_k \mathbb{1}_{B_k}$$

and observe that, as in Example 8.5(iv),

$$f + g = \sum_{i=0}^{M} \sum_{k=0}^{N} (y_i + z_k) \mathbb{1}_{A_i \cap B_k} \in \mathcal{E}^+(\mathscr{A})$$

is a standard representation of $f + g$. Thus

$$\begin{aligned}
I_\mu(f+g) &= \sum_{i=0}^{M} \sum_{k=0}^{N} (y_i + z_k) \mu(A_i \cap B_k) \\
&= \sum_{i=0}^{M} y_i \sum_{k=0}^{N} \mu(A_i \cap B_k) + \sum_{k=0}^{N} z_k \sum_{i=0}^{M} \mu(A_i \cap B_k) \\
&\overset{(9.2)}{\underset{(9.3)}{=}} \sum_{i=0}^{M} y_i \mu(A_i) + \sum_{k=0}^{N} z_k \mu(B_k) \\
&= I_\mu(f) + I_\mu(g).
\end{aligned}$$

(iv) If $f \leqslant g$, then $g = f + (g - f)$, where $g - f \in \mathcal{E}^+(\mathscr{A})$, see e.g. Example 8.5(iv). By part (iii) of this proof,

$$I_\mu(g) = I_\mu(f) + I_\mu(g - f) \geqslant I_\mu(f)$$

since $g - f$, hence $I_\mu(g - f)$, is positive. □

In Theorem 8.8 we saw that every $u \in \mathcal{M}^+(\mathscr{A})$ can be written as an increasing limit of simple functions; by Corollary 8.10, suprema of simple functions are again measurable, so that

$$u \in \mathcal{M}^+(\mathscr{A}) \iff u = \sup_{n \in \mathbb{N}} f_n, \quad f_n \in \mathcal{E}^+(\mathscr{A}), \quad f_n \leqslant f_{n+1} \leqslant \cdots .$$

We will use this to 'inscribe' simple functions (which we know how to integrate) below the graph of a positive measurable function u and exhaust the μ-area below u.

Definition 9.4 Let $(X, \mathscr{A}, \mu)$ be a measure space. The (μ-)*integral* of a positive function $u \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$ is given by

$$\int u\, d\mu := \sup\{I_\mu(g) : g \leqslant u,\ g \in \mathcal{E}^+(\mathscr{A})\} \in [0, \infty]. \tag{9.4}$$

If we need to emphasize the *integration variable*, we write $\int u(x)\, \mu(dx)$.[2]

The key observation is that the integral $\int \ldots\, d\mu$ extends I_μ, i.e. we have the following lemma.

Lemma 9.5 *For all $f \in \mathcal{E}^+(\mathscr{A})$ we have $\int f\, d\mu = I_\mu(f)$.*

Proof Let $f \in \mathcal{E}^+(\mathscr{A})$. Since $f \leqslant f$, f is an admissible function in the supremum appearing in (9.4), hence

$$I_\mu(f) \leqslant \sup\{I_\mu(g) : g \leqslant f,\ g \in \mathcal{E}^+(\mathscr{A})\} \overset{\text{def}}{=} \int f\, d\mu.$$

On the other hand, $\mathcal{E}^+(\mathscr{A}) \ni g \leqslant f$ implies that $I_\mu(g) \leqslant I_\mu(f)$ by Property 9.3(iv), and

$$\int f\, d\mu \overset{\text{def}}{=} \sup\{I_\mu(g) : g \leqslant f,\ g \in \mathcal{E}^+(\mathscr{A})\} \leqslant I_\mu(f). \qquad \square$$

The next result is the first of several *convergence theorems*. It shows, in particular, that we could have defined (9.4) using *any* increasing sequence $f_n \uparrow u$ of simple functions $f_n \in \mathcal{E}^+(\mathscr{A})$.

Theorem 9.6 (Beppo Levi) *Let $(X, \mathscr{A}, \mu)$ be a measure space. For an increasing sequence of functions $(u_n)_{n\in\mathbb{N}} \subset \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$, $0 \leqslant u_n \leqslant u_{n+1} \leqslant \cdots$, we have for the supremum $u := \sup_{n\in\mathbb{N}} u_n \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$ and*

$$\int \sup_{n\in\mathbb{N}} u_n\, d\mu = \sup_{n\in\mathbb{N}} \int u_n\, d\mu. \tag{9.5}$$

Note that we can write $\lim_{n\to\infty}$ instead of $\sup_{n\in\mathbb{N}}$ in (9.5) since the supremum of an increasing sequence is its limit. Moreover, (9.5) holds in $[0, +\infty]$, i.e. the case '$+\infty = +\infty$' is possible.

Proof of Theorem 9.6 We know from Corollary 8.10 that $u \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$.

[2] Some authors use the alternative notation $\int u(x) d\mu(x)$.

Step 1. Claim: $u, w \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$, $u \leqslant w \implies \int u\,d\mu \leqslant \int w\,d\mu$. If $u \leqslant w$, then every simple function with $f \leqslant u$ also satisfies $f \leqslant w$. Therefore,

$$\begin{aligned}\int u\,d\mu &= \sup\{I_\mu(f) : f \leqslant u,\, f \in \mathcal{E}^+(\mathscr{A})\}\\ &\leqslant \sup\{I_\mu(g) : g \leqslant w,\ g \in \mathcal{E}^+(\mathscr{A})\} = \int w\,d\mu.\end{aligned}$$

Step 2. Claim: $\sup_{n\in\mathbb{N}} \int u_n\,d\mu \leqslant \int \sup_{n\in\mathbb{N}} u_n\,d\mu$. According to Step 1, the integral is a monotone functional. Therefore,

$$\begin{aligned}\forall m : u_m \leqslant \sup_{n\in\mathbb{N}} u_n \overset{\text{Step 1}}{\implies} \forall m : \int u_m\,d\mu &\leqslant \underbrace{\int \sup_{n\in\mathbb{N}} u_n\,d\mu}_{\text{independent of } m}\\ \implies \sup_{m\in\mathbb{N}} \int u_m\,d\mu &\leqslant \int \sup_{n\in\mathbb{N}} u_n\,d\mu.\end{aligned}$$

Step 3. Claim: $f \leqslant u$, $f \in \mathcal{E}^+(\mathscr{A}) \implies I_\mu(f) \leqslant \sup_{n\in\mathbb{N}} \int u_n\,d\mu$. Pick some simple function $f \in \mathcal{E}^+(\mathscr{A})$ such that $f \leqslant u$ and fix $\alpha \in (0,1)$. Then

$$\begin{aligned}u = \sup_{n\in\mathbb{N}} u_n &\implies \forall x \quad \exists N(x,\alpha) \in \mathbb{N} \quad \forall n \geqslant N(x,\alpha) : \alpha f(x) \leqslant u_n(x)\\ &\implies B_n := \underbrace{\{x : \alpha f(x) \leqslant u_n(x)\}}_{\in\mathscr{A}} \uparrow_{n\to\infty} X\\ &\implies \alpha \mathbb{1}_{B_n} \cdot f \leqslant \mathbb{1}_{B_n} \cdot u_n \leqslant u_n,\end{aligned}$$

where we use the definition of the sets B_n (for the penultimate estimate) and the fact that $\mathbb{1}_{B_n} \leqslant 1$ (in the last estimate). For all simple functions $f = \sum_{m=0}^M y_m \mathbb{1}_{A_m}$ we see that

$$\alpha \sum_{m=0}^{M} y_m \mu(B_n \cap A_m) = I_\mu(\alpha \mathbb{1}_{B_n} \cdot f) \leqslant \int u_n\,d\mu \leqslant \underbrace{\sup_{m\in\mathbb{N}} \int u_m\,d\mu}_{\text{independent of } n}.$$

Since the right-hand side does not depend on n and $B_n \uparrow X$, we can let $n \to \infty$ and get $\mu(B_n \cap A_m) \to \mu(A_m)$. This shows that

$$\alpha I_\mu(f) = \alpha \sum_{m=0}^{M} y_m \mu(A_m) \leqslant \sup_{m\in\mathbb{N}} \int u_m\,d\mu \quad \forall \alpha \in (0,1).$$

Since the right-hand side does not depend on $\alpha \in (0,1)$, the claim follows as $\alpha \to 1$.

Step 4. In the estimate which we established in Step 3, we can go to the supremum over all $f \in \mathcal{E}^+(\mathscr{A})$ with $f \leqslant u$. This yields

$$\int u\,d\mu = \sup_{\mathcal{E}^+(\mathscr{A}) \ni f \leqslant u} I_\mu(f) \leqslant \sup_{n\in\mathbb{N}} \int u_n\,d\mu,$$

finishing the proof. □

One can see the next corollary just as a special case of Theorem 9.6. Its true meaning, however, is that it allows us to calculate the integral of a measurable function using *any* approximating *sequence* of simple functions – and this is a considerable simplification of the original definition (9.4).

Corollary 9.7 *Let* $u \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$. *Then*

$$\int u\,d\mu = \lim_{n\to\infty} \int f_n\,d\mu$$

holds for every increasing sequence $(f_n)_{n\in\mathbb{N}} \subset \mathcal{E}^+(\mathscr{A})$ *with* $\lim_{n\to\infty} f_n = u$.

Properties 9.8 (of the integral) Let $u, v \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$. Then

(i) $\displaystyle\int \mathbb{1}_A\,d\mu = \mu(A) \qquad \forall A \in \mathscr{A}$;

(ii) $\displaystyle\int \alpha u\,d\mu = \alpha \int u\,d\mu \qquad \forall \alpha \geqslant 0$ (*positive homogeneous*);

(iii) $\displaystyle\int (u+v)\,d\mu = \int u\,d\mu + \int v\,d\mu$ (*additive*);

(iv) $\displaystyle u \leqslant v \implies \int u\,d\mu \leqslant \int v\,d\mu$ (*monotone*).

Proof (i) follows from Property 9.3(i) and Lemma 9.5. (ii) and (iii) follow from the corresponding properties of I_μ, Corollary 9.7 and the usual rules for limits. (iv) has been proved in Step 1 of the proof of Theorem 9.6. □

Corollary 9.9 *Let* $(u_n)_{n\in\mathbb{N}} \subset \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$. *Then* $\sum_{n=1}^\infty u_n$ *is measurable and we have*

$$\int \sum_{n=1}^\infty u_n\,d\mu = \sum_{n=1}^\infty \int u_n\,d\mu \tag{9.6}$$

(*including the possibility* '$+\infty = +\infty$').

Proof Set $s_M := u_1 + u_2 + \cdots + u_M$ and use Property 9.8(iii) and Theorem 9.6. □

Example 9.10 Let $(X, \mathscr{A})$ be a measurable space.

(i) Let $\mu = \delta_y$ be the Dirac measure for some fixed $y \in X$. Then

$$\int u \, d\delta_y = \int u(x)\delta_y(dx) = u(y) \quad \forall u \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A}).$$

Indeed: for any $f \in \mathcal{E}^+(\mathscr{A})$ with standard representation $f = \sum_{n=0}^{M} \phi_n \mathbb{1}_{A_n}$, we know that $y \in X$ lies in exactly one of the A_n, say $y \in A_{n_0}$. Then

$$\int f(x)\delta_y(dx) = \int \sum_{n=0}^{M} \phi_n \mathbb{1}_{A_n}(x)\delta_y(dx) = \sum_{n=0}^{M} \phi_n \delta_y(A_n) = \phi_{n_0} = f(y).$$

Now take any sequence of simple functions $f_k \uparrow u$. By Corollary 9.7

$$\int u(x)\delta_y(dx) = \lim_{k\to\infty} \int f_k(x)\delta_y(dx) = \lim_{k\to\infty} f_k(y) = u(y).$$

(ii) Let $(X, \mathscr{A}, \mu) = \left(\mathbb{N}, \mathscr{P}(\mathbb{N}), \sum_{n=1}^{\infty} \alpha_n \delta_n\right)$. As we saw in Problem 4.7(ii), μ is indeed a measure and $\mu\{k\} = \alpha_k$. On the other hand, all $\mathscr{P}(\mathbb{N})$-measurable positive functions are of the form

$$u(k) = \sum_{n=1}^{\infty} u_n \mathbb{1}_{\{n\}}(k) \quad \forall k \in \mathbb{N}$$

for a suitable sequence $(u_n)_{n\in\mathbb{N}} \subset [0, \infty]$.[3] Thus, by Corollary 9.9,

$$\begin{aligned}\int u \, d\mu &= \int \sum_{n=1}^{\infty} u_n \mathbb{1}_{\{n\}} \, d\mu = \sum_{n=1}^{\infty} \int u_n \mathbb{1}_{\{n\}} \, d\mu \\ &= \sum_{n=1}^{\infty} u_n \mu\{n\} = \sum_{n=1}^{\infty} u_n \alpha_n.\end{aligned}$$

We close this chapter with another *convergence theorem*. It is due to P. Fatou and often called *Fatou's lemma*.

Theorem 9.11 (Fatou) *Let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$ be a sequence of positive measurable functions. Then $u := \liminf_{n\to\infty} u_n$ is measurable and*

$$\int \liminf_{n\to\infty} u_n \, d\mu \leqslant \liminf_{n\to\infty} \int u_n \, d\mu.$$

[3] This means that we can identify $\mathscr{P}(\mathbb{N})$-measurable functions $u : \mathbb{N} \to [0, \infty]$ and the arbitrary sequences $(u_n)_{n\in\mathbb{N}} \subset [0, \infty]$ by $u(k) = u_k$.

Proof Recall that $\liminf_{n\to\infty} u_n = \sup_{k\in\mathbb{N}} \inf_{n\geqslant k} u_n$ always exists in $\overline{\mathbb{R}}$; the measurability of lim inf was shown in Corollary 8.10. Applying Theorem 9.6 to the increasing sequence $\left(\inf_{n\geqslant k} u_n\right)_{k\in\mathbb{N}}$ – which is in $\mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$ by Corollary 8.10 – we find

$$\begin{aligned}\int \liminf_{n\to\infty} u_n \, d\mu &\overset{9.6}{=} \sup_{k\in\mathbb{N}} \left(\int \inf_{n\geqslant k} u_n \, d\mu \right) \\ &\overset{9.8(\text{iv})}{\leqslant} \sup_{k\in\mathbb{N}} \left(\inf_{\ell\geqslant k} \int u_\ell \, d\mu \right) \\ &= \liminf_{\ell\to\infty} \int u_\ell \, d\mu,\end{aligned}$$

where we use that $\inf_{n\geqslant k} u_n \leqslant u_\ell$ for all $\ell \geqslant k$ and the monotonicity of the integral, see Property 9.8(iv). □

Problems

9.1. Let $f: X \to \mathbb{R}$ be a positive simple function of the form $f(x) = \sum_{n=1}^m \xi_n \mathbb{1}_{A_n}(x)$, $\xi_n \geqslant 0$, $A_n \in \mathscr{A}$ – but not necessarily disjoint. Show that $I_\mu(f) = \sum_{n=1}^m \xi_n \mu(A_n)$.
[Hint: use additivity and positive homogeneity of I_μ.]

9.2. Let $(X, \mathscr{A}, \mu)$ be a measure space and $A_1, \dots, A_N \in \mathscr{A}$ such that $\mu(A_n) < \infty$. Then

$$\mu\left(\bigcup_{n=1}^N A_n\right) \geqslant \sum_{n=1}^N \mu(A_n) \; - \sum_{1\leqslant n<k\leqslant N} \mu(A_n \cap A_k).$$

[Hint: show first an inequality for indicator functions.]

9.3. Complete the proof of Properties 9.8.

9.4. Find an example showing that an '*increasing sequence of functions*' is, in general, different from a '*sequence of increasing functions*'.

9.5. Let $(X, \mathscr{A}, \mu)$ be a measure space. Show the following variant of Theorem 9.6. If $u_n \geqslant 0$ are measurable functions such that for some u we have

$$\exists K \in \mathbb{N} \quad \forall x : u_{n+K}(x) \uparrow u(x) \;\text{ as } n \to \infty,$$

then $u \geqslant 0$ is measurable and $\int u_n \, d\mu \uparrow \int u \, d\mu$.
Show that we cannot replace the above condition with

$$\forall x \quad \exists K \in \mathbb{N} : u_{n+K}(x) \uparrow u(x) \;\text{ as } n \to \infty.$$

9.6. Complete the proof of Corollary 9.9 and show that (9.6) is actually *equivalent* to (9.5) in Beppo Levi's theorem.

9.7. Let $(X, \mathscr{A}, \mu)$ be a measure space and $u \in \mathcal{M}^+(\mathscr{A})$. Show that $A \mapsto \int \mathbb{1}_A u \, d\mu$, $A \in \mathscr{A}$, is a measure.

9.8. Prove that *every* function $u: \mathbb{N} \to \mathbb{R}$ on $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$ is measurable.

9.9. Let $(X, \mathscr{A})$ be a measurable space and $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of measures thereon. Set, as in Example 9.10(ii), $\mu = \sum_{n\in\mathbb{N}} \mu_n$. By Problem 4.7(ii) this is again a measure. Show that

$$\int u\,d\mu = \sum_{n\in\mathbb{N}} \int u\,d\mu_n \quad \forall u \in \mathcal{M}^+(\mathscr{A}).$$

[Instructions: (1) consider $u = \mathbb{1}_A$; (2) consider $u = f \in \mathcal{E}^+$; and (3) approximate $u \in \mathcal{M}^+$ by an increasing sequence of simple functions and use Theorem 9.6. To interchange increasing limits/suprema use the hint to Problem 4.7(ii).]

9.10. Reverse Fatou lemma. Let $(X, \mathscr{A}, \mu)$ be a measure space and $(u_n)_{n\in\mathbb{N}} \subset \mathcal{M}^+(\mathscr{A})$. If $u_n \leqslant u$ for all $n \in \mathbb{N}$ and some $u \in \mathcal{M}^+(\mathscr{A})$ with $\int u\,d\mu < \infty$, then

$$\limsup_{n\to\infty} \int u_n\,d\mu \leqslant \int \limsup_{n\to\infty} u_n\,d\mu.$$

9.11. Fatou's lemma for measures. Let $(X, \mathscr{A}, \mu)$ be a measure space and let $(A_n)_{n\in\mathbb{N}}$, $A_n \in \mathscr{A}$, be a sequence of measurable sets. We set

$$\liminf_{n\to\infty} A_n := \bigcup_{k\in\mathbb{N}} \bigcap_{n\geqslant k} A_n \quad \text{and} \quad \limsup_{n\to\infty} A_n := \bigcap_{k\in\mathbb{N}} \bigcup_{n\geqslant k} A_n. \tag{9.7}$$

(i) Prove that $\mathbb{1}_{\liminf_{n\to\infty} A_n} = \liminf_{n\to\infty} \mathbb{1}_{A_n}$ and $\mathbb{1}_{\limsup_{n\to\infty} A_n} = \limsup_{n\to\infty} \mathbb{1}_{A_n}$.
[Hint: check first that $\mathbb{1}_{\bigcap_{n\in\mathbb{N}} A_n} = \inf_{n\in\mathbb{N}} \mathbb{1}_{A_n}$ and $\mathbb{1}_{\bigcup_{n\in\mathbb{N}} A_n} = \sup_{n\in\mathbb{N}} \mathbb{1}_{A_n}$.]

(ii) Prove that $\mu\Big(\liminf_{n\to\infty} A_n\Big) \leqslant \liminf_{n\to\infty} \mu(A_n)$.

(iii) Prove that $\limsup_{n\to\infty} \mu(A_n) \leqslant \mu\Big(\limsup_{n\to\infty} A_n\Big)$ if μ is a finite measure.

(iv) Provide an example showing that (iii) fails if μ is not finite.

9.12. Let $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ be a sequence of disjoint sets such that $\dot{\bigcup}_{n\in\mathbb{N}} A_n = X$. Show that for every $u \in \mathcal{M}^+(\mathscr{A})$

$$\int u\,d\mu = \sum_{n=1}^{\infty} \int \mathbb{1}_{A_n} u\,d\mu.$$

Use this to construct on a σ-finite measure space $(X, \mathscr{A}, \mu)$ a function w which satisfies $w(x) > 0$ for all $x \in X$ and $\int w\,d\mu < \infty$.

9.13. Kernels. Let $(X, \mathscr{A}, \mu)$ be a measure space. A map $N: X \times \mathscr{A} \to [0, \infty]$ is called a *kernel* if

$$\begin{aligned} &A \mapsto N(x, A) && \text{is a measure for every } x \in X, \\ &x \mapsto N(x, A) && \text{is a measurable function for every } A \in \mathscr{A}. \end{aligned}$$

(i) Show that $\mathscr{A} \ni A \mapsto \mu N(A) := \int N(x, A)\mu(dx)$ is a measure on $(X, \mathscr{A})$.

(ii) For $u \in \mathcal{M}^+(\mathscr{A})$ define $Nu(x) := \int u(y) N(x, dy)$. Show that $u \mapsto Nu$ is additive, positive homogeneous and $Nu(\,\cdot\,) \in \mathcal{M}^+(\mathscr{A})$.

(iii) Let μN be the measure introduced in (i). Show that $\int u\,d(\mu N) = \int Nu\,d\mu$ for all $u \in \mathcal{M}^+(\mathscr{A})$.

[Hint: consider in each part of this problem first indicator functions $u = \mathbb{1}_A$, then simple functions $u \in \mathcal{E}^+(\mathscr{A})$, and then approximate $u \in \mathcal{M}^+(\mathscr{A})$ by simple functions using Theorems 8.8 and 9.6.]

9.14. (Continuation of Problem 6.3) Consider on $\mathbb{R}$ the σ-algebra Σ of all Borel sets which are symmetric w.r.t. the origin. Set $A^+ := A \cap [0, \infty)$, $A^- := (-\infty, 0] \cap A$ and consider their symmetrizations $A_\sigma^\pm := A^\pm \cup (-A^\pm) \in \Sigma$. Show that for every $u \in \mathcal{M}^+(\Sigma)$ with $0 \leqslant u \leqslant 1$ and for every measure μ on $(\mathbb{R}, \Sigma)$ the set function

$$\mathscr{B}(\mathbb{R}) \ni A \mapsto \int u \mathbb{1}_{A_\sigma^+} \, d\mu + \int (1-u) \mathbb{1}_{A_\sigma^-} \, d\mu$$

is a measure on $\mathscr{B}(\mathbb{R})$ that extends μ.
Why does this not contradict the uniqueness Theorem 5.7 for measures?

10

Integrals of Measurable Functions

Throughout this chapter $(X, \mathscr{A}, \mu)$ will be a measure space. So far, we have constructed an integral for positive measurable functions starting from a measure. The key steps and ideas are shown in Fig. 10.1. We want to extend this integral to not necessarily positive measurable functions $u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$ by linearity. The fundamental observation is that

$$u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A}) \iff u = u^+ - u^-, \quad u^+, u^- \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$$

(see Corollary 8.12). This suggests the following definition.

Definition 10.1 A function $u: X \to \overline{\mathbb{R}}$ on a measure space $(X, \mathscr{A}, \mu)$ is said to be (μ-)*integrable*, if it is $\mathscr{A}/\mathscr{B}(\overline{\mathbb{R}})$-measurable and if $\int u^+ \,d\mu$, $\int u^- \,d\mu < \infty$. In this case we call

$$\int u \,d\mu := \int u^+ \,d\mu - \int u^- \,d\mu \in (-\infty, \infty) \tag{10.1}$$

the (μ-)*integral* of u. We write $\mathcal{L}^1(\mu)$ and $\mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ for the set of all real-valued and $\overline{\mathbb{R}}$-valued μ-integrable functions.

Remark 10.2 (i) As usual, we must discuss the well-definedness of (10.1). Let $u = f - g$ be a further representation of $u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$ as a difference of positive measurable functions. Then $\int u \,d\mu = \int f \,d\mu - \int g \,d\mu$ (unless we are in the situation '$\infty - \infty$').

Indeed: we can rewrite $u^+ - u^- = f - g$ as $u^+ + g = u^- + f \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$ and integrate this identity to get

$$\begin{aligned} &\int u^+ \,d\mu + \int g \,d\mu = \int f \,d\mu + \int u^- \,d\mu \\ \iff &\int u^+ \,d\mu - \int u^- \,d\mu = \int f \,d\mu - \int g \,d\mu. \end{aligned}$$

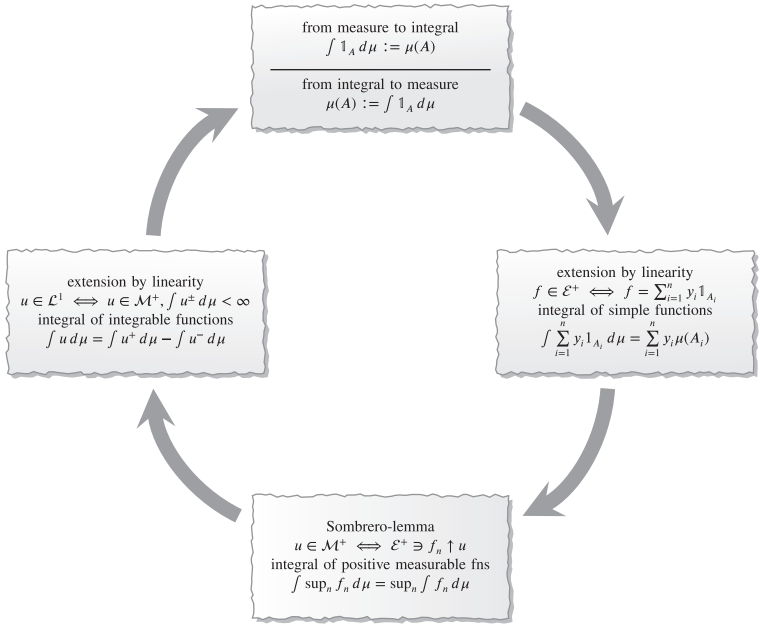


Fig. 10.1. *The way of integration: from measures to integrals and back.* Guided by the idea that the integral should be the area between the graph of a function and the x-axis, we define the integral for indicator functions of measurable sets $A \in \mathscr{A}$ as $\int \mathbb{1}_A \, d\mu = \mu(A)$ and extend it by linearity to linear combinations of indicators: the positive simple functions $\mathcal{E}^+(\mathscr{A})$ (there is an issue about well-definedness which is addressed in Lemma 9.1). Since all positive measurable functions can be obtained as increasing limits of simple functions (sombrero lemma, Theorem 8.8), we can define the integral of $u \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$ with increasing sequences of simple functions $f_n \uparrow u$. Integrating positive and negative parts of $u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$ separately, this integral extends, again by linearity, to functions with arbitrary sign.

(ii) If we need to exhibit the integration variable, we write

$$\int u \, d\mu = \int u(x) \mu(dx).^1$$

(iii) $\int u \, d\lambda^n$ is called an (n-dimensional) *Lebesgue integral* and $u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\lambda^n)$ is said to be *Lebesgue integrable*.[2] Traditionally one writes $\int u(x) dx$ or $\int u \, dx$ instead of the formally more correct $\int u \, d\lambda^n$. If we want to stress X or $\mathscr{A}$, etc., we will also write $\mathcal{L}^1_{\overline{\mathbb{R}}}(X)$ or $\mathcal{L}^1_{\overline{\mathbb{R}}}(\mathscr{A})$, etc.

[1] Some authors use the alternative notation $\int u(x) d\mu(x)$.

[2] the letter $\mathcal{L}$ is in honour of H. Lebesgue who was one of the pioneers of modern integration theory. If μ is other than λ^n, $\int \ldots \, d\mu$ is sometimes called the *abstract Lebesgue integral*.

(iv) In the definition of the integral for positive $u \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$ we did allow that $\int u\,d\mu = \infty$. Since we want to avoid the case '$\infty - \infty$' in (10.1), we impose the finiteness condition $\int u^{\pm}\,d\mu < \infty$. In particular, a positive function is said to be *integrable* only if the integral is finite:

$$u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu),\ u \geqslant 0 \iff u \in \mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A}) \quad \text{and} \quad \int u\,d\mu < \infty$$

(which is clear since for positive functions $u^+ = u$ and $u^- = 0$).

Caution Some authors call u μ-integrable if $\int u^+\,d\mu - \int u^-\,d\mu$ makes sense in $\overline{\mathbb{R}}$, i.e. whenever it is not of the form '$\infty - \infty$'. We will not use this convention.

Let us briefly summarize the most important integrability criteria.

Theorem 10.3 *Let $u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$. Then the following conditions are equivalent:*

(i) $u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$;
(ii) $u^+, u^- \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$;
(iii) $|u| \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$;
(iv) $\exists w \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu),\ w \geqslant 0$ *such that* $|u| \leqslant w$.

Proof (i)$\Leftrightarrow$(ii): this is just the definition of integrability.

(ii)$\Rightarrow$(iii): since $|u| = u^+ + u^-$, we can use the additivity of the integral on $\mathcal{M}^+_{\overline{\mathbb{R}}}(\mathscr{A})$, see Property 9.8(iii), to get $\int |u| d\mu = \int u^+\,d\mu + \int u^-\,d\mu < \infty$.

(iii)$\Rightarrow$(iv): take $w := |u|$.

(iv)$\Rightarrow$(i): we have to show that $u^{\pm} \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$. Since $u^{\pm} \leqslant |u| \leqslant w$ we find by the monotonicity of the integral of Property 9.8(iv) that $\int u^{\pm}\,d\mu \leqslant \int w\,d\mu < \infty$. $\square$

Now it is easy to extend Properties 9.8 of the integral to the set $\mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$.

Theorem 10.4 *Let $(X, \mathscr{A}, \mu)$ be a measure space and $u, v \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$, $\alpha \in \mathbb{R}$. Then*

(i) $\alpha u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ *and* $\int \alpha u\,d\mu = \alpha \int u\,d\mu$ (*homogeneous*);

(ii) $u + v \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ *and* $\int (u+v) d\mu = \int u\,d\mu + \int v\,d\mu$ (*additive*)
(*whenever $u + v$ is defined*);

(iii) $\min\{u, v\}, \max\{u, v\} \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ (*lattice property*);

(iv) $u \leqslant v \Longrightarrow \int u\,d\mu \leqslant \int v\,d\mu$ (monotone);

(v) $\left|\int u\,d\mu\right| \leqslant \int |u|d\mu$ (triangle inequality).

Proof There are principally two ways to prove this theorem: either we consider positive and negative parts for (i)–(v) and show that their integrals are finite, or we use Theorem 10.3(iii), (iv). Doing this we find the following.

(i) $|\alpha u| = |\alpha| \cdot |u| \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ by Property 9.8(ii). To see the integral formula we assume that $\alpha \leqslant 0$. Then $(\alpha u)^{\pm} = (-\alpha)(-u)^{\pm} = -\alpha u^{\mp}$ and the formula follows directly from Definition 10.1. The case where $\alpha \geqslant 0$ is analogous.

(ii) $|u+v| \leqslant |u| + |v| \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ by Property 9.8(iii). For the integral formula observe that $(u+v)^+ - (u+v)^- = u+v = (u^+ + v^+) - (u^- + v^-)$ and use Remark 10.2(i).

(iii) $|\max\{u,v\}| \leqslant |u| + |v| \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ and $|\min\{u,v\}| \leqslant |u| + |v| \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ by Property 9.8(iii).

(iv) If $u \leqslant v$, we find that $u^+ \leqslant v^+$ and $v^- \leqslant u^-$. Thus

$$\int u\,d\mu = \int u^+\,d\mu - \int u^-\,d\mu \overset{9.8(\text{iv})}{\leqslant} \int v^+\,d\mu - \int v^-\,d\mu = \int v\,d\mu.$$

(v) Using $\pm u \leqslant |u|$ we deduce from (iv) that

$$\begin{aligned}\left|\int u\,d\mu\right| &= \max\left\{\int u\,d\mu,\ -\int u\,d\mu\right\}\\ &\leqslant \max\left\{\int |u|d\mu, \int |-u|d\mu\right\} = \int |u|d\mu.\end{aligned}$$ □

Remark 10.5 If $u(x) \pm v(x)$ is defined in $\overline{\mathbb{R}}$ for all $x \in X$ – i.e. if we can exclude '$\infty - \infty$' – then Theorem 10.4(i), (ii) just say that the integral is *linear*:

$$\int (\alpha u + \beta v)d\mu = \alpha \int u\,d\mu + \beta \int v\,d\mu, \quad \alpha, \beta \in \mathbb{R}. \tag{10.2}$$

This is always true for real-valued $u, v \in \mathcal{L}^1(\mu)$, i.e. $\mathcal{L}^1(\mu)$ is a *vector space* with addition and scalar ($\mathbb{R}$) multiplication defined by

$$(u+v)(x) := u(x) + v(x), \quad (\alpha \cdot u)(x) := \alpha \cdot u(x),$$

and

$$\int \ldots\,d\mu : \mathcal{L}^1(\mu) \to \mathbb{R}, \quad u \mapsto \int u\,d\mu,$$

is a *positive linear functional*.

Example 10.6 Let us reconsider the examples from Example 9.10.

(i) On $(X, \mathscr{A}, \delta_y)$, $y \in X$ fixed, we have $\int u(x)\delta_y(dx) = u(y)$ and

$$u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\delta_y) \iff u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A}) \text{ and } |u(y)| < \infty.$$

(ii) On $\left(\mathbb{N}, \mathscr{P}(\mathbb{N}), \mu := \sum_{n=1}^{\infty} \alpha_n \delta_n\right)$ every $u : \mathbb{N} \to \mathbb{R}$ is measurable, see Problem 9.8. From Example 9.10(ii) we know that $\int |u| d\mu = \sum_{n=1}^{\infty} \alpha_n |u(n)|$, so that

$$u \in \mathcal{L}^1(\mu) \iff \sum_{n=1}^{\infty} \alpha_n |u(n)| < \infty.$$

If $\alpha_1 = \alpha_2 = \cdots = 1$, $\mathcal{L}^1(\mu)$ is called the set of *summable sequences* and customarily denoted by $\ell^1(\mathbb{N}) = \left\{(x_n)_{n\in\mathbb{N}} \subset \mathbb{R} : \sum_{n=1}^{\infty} |x_n| < \infty\right\}$. This space is important in functional analysis.

(iii) Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. Then every bounded measurable function ('random variable') $\xi \in \mathcal{M}(\mathscr{A})$, $C := \sup_{\omega\in\Omega} |\xi(\omega)| < \infty$, is integrable. This follows immediately from

$$\int |\xi| d\mathbb{P} \leqslant \int \sup_{\omega\in\Omega} |\xi(\omega)| \mathbb{P}(d\omega) = C \int \mathbb{P}(d\omega) = C < \infty.$$

Caution Not every $\mathbb{P}$-integrable function is bounded. [✍]

For $A \in \mathscr{A}$ and $u \in \mathcal{M}^{\pm}_{\overline{\mathbb{R}}}(\mathscr{A})$ [or $\mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$] we know from Example 8.5(i) and Corollary 8.11 [and Theorem 10.3(iv) using $|\mathbb{1}_A u| \leqslant |u|$] that $\mathbb{1}_A u$ is again measurable [or integrable].

Definition 10.7 Let $(X, \mathscr{A}, \mu)$ be a measure space and assume that $u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ or $u \in \mathcal{M}^{\pm}_{\overline{\mathbb{R}}}(\mathscr{A})$. Then

$$\int_A u \, d\mu := \int \mathbb{1}_A u \, d\mu = \int \mathbb{1}_A(x) u(x) \mu(dx), \quad \forall A \in \mathscr{A}.$$

Of course, $\int_X u \, d\mu = \int u \, d\mu$.

Lemma 10.8 *On the measure space* $(X, \mathscr{A}, \mu)$ *let* $u \in \mathcal{M}^+(\mathscr{A})$*. The set function*

$$\nu : A \mapsto \int_A u \, d\mu = \int \mathbb{1}_A u \, d\mu, \quad A \in \mathscr{A},$$

is a measure on $(X, \mathscr{A})$*. It is called the measure* with density (function) u with respect to μ. We write $\nu = u\,\mu$ or $d\nu = u\,d\mu$.

Proof This is left as an exercise. □

If ν has a density w.r.t. μ, one writes traditionally $d\nu/d\mu$ for the density function. This notation is to be understood in a purely symbolical way; this is motivated by the well-known *fundamental theorem of integral and differential calculus* (for Riemann integrals)

$$U(x) - U(a) = \int_a^x u(y)dy \implies \frac{dU}{dx} = u.$$

Using our notation $\lambda = \lambda^1$ and $\lambda(dx) = dx$ this becomes

$$\frac{d(u\lambda)}{d\lambda} = \frac{dU}{dx} = u.$$

If $u \geqslant 0$ is positive, then it induces a measure ν:

$$\nu[a, b) = U(b) - U(a) = \int_{[a,b)} u\, d\lambda$$

(see Problem 6.1). A more advanced discussion of derivatives can be found in Chapter 25, Theorem 25.20 and Appendix I, items I.16–I.19.

Problems

10.1. Prove Remark 10.5, i.e. prove the linearity of the integral.

10.2. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. Find a counterexample to the claim that *every* $\mathbb{P}$*-integrable function* $u \in \mathcal{L}^1(\mathbb{P})$ *is bounded*.
[Hint: you could try to take $\Omega = (0, 1)$, $\mathbb{P} = \lambda^1$ and show that $1/\sqrt{x}$ is Lebesgue integrable on $(0, 1)$ by finding a sequence of suitable simple functions that is below $1/\sqrt{x}$ on, say, $(1/m, 1)$ and then let $m \to \infty$ using Beppo Levi's theorem.]

10.3. Prove Lemma 10.8.

10.4. Let $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ be a sequence of mutually disjoint sets. Show that

$$u\mathbb{1}_{\bigcup_n A_n} \in \mathcal{L}^1(\mu) \iff u\mathbb{1}_{A_n} \in \mathcal{L}^1(\mu) \quad \text{and} \quad \sum_{n=1}^{\infty} \int_{A_n} |u| d\mu < \infty.$$

10.5. Let $(X, \mathscr{A}, \mu)$ be a measure space and $u \in \mathcal{M}(\mathscr{A})$. Show that

$$u \in \mathcal{L}^1(\mu) \iff \sum_{n\in\mathbb{Z}} 2^n \mu\left\{2^n \leqslant |u| < 2^{n+1}\right\} < \infty.$$

10.6. Let $(X, \mathscr{A}, \mu)$ be a finite measure space. Show that

(i) $u \in \mathcal{L}^1(\mu) \iff u \in \mathcal{M}(\mathscr{A})$ and $\displaystyle\sum_{n=0}^{\infty} \mu\{|u| \geqslant n\} < \infty$;

(ii) $\displaystyle\sum_{n=1}^{\infty} \mu\{|u| \geqslant n\} \leqslant \int |u| d\mu \leqslant \sum_{n=0}^{\infty} \mu\{|u| \geqslant n\}$.

(iii) The lower estimate in (ii) holds in an arbitrary measure space.

10.7. Generalized Fatou lemma. Assume that $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mu)$. Prove the following.

(i) If $u_n \geqslant v$ for all $n \in \mathbb{N}$ and some $v \in \mathcal{L}^1(\mu)$, then

$$\int \liminf_{n\to\infty} u_n \, d\mu \leqslant \liminf_{n\to\infty} \int u_n \, d\mu.$$

(ii) If $u_n \leqslant w$ for all $n \in \mathbb{N}$ and some $w \in \mathcal{L}^1(\mu)$, then

$$\limsup_{n\to\infty} \int u_n \, d\mu \leqslant \int \limsup_{n\to\infty} u_n \, d\mu.$$

(iii) Find examples that show that the upper and lower bounds in (i) and (ii) are necessary.

[Hint: mimic and scrutinize the proof of Fatou's lemma, especially when it comes to the application of Beppo Levi's theorem. What goes wrong if we do not have this upper/lower bound? Note that we have an 'invisible' $v = 0$ in Theorem 9.11.]

10.8. Independence (2). Let $(\Omega, \mathscr{A}, P)$ be a probability space and assume that the σ-algebras $\mathscr{B}, \mathscr{C} \subset \mathscr{A}$ are independent (see Problem 5.11). Show that $u \in \mathcal{M}^+(\mathscr{B})$ and $w \in \mathcal{M}^+(\mathscr{C})$ satisfy

$$\int uw \, dP = \int u \, dP \cdot \int w \, dP$$

and that for $u \in \mathcal{M}(\mathscr{B})$ and $w \in \mathcal{M}(\mathscr{C})$

$$uw \in \mathcal{L}^1(\mathscr{A}) \iff u \in \mathcal{L}^1(\mathscr{B}) \quad \text{and} \quad w \in \mathcal{L}^1(\mathscr{C}).$$

Find an example proving that this fails if $\mathscr{B}$ and $\mathscr{C}$ are not independent.
[Hint: start with simple functions and use Beppo Levi's theorem.]

10.9. Integrating complex functions. Let $(X, \mathscr{A}, \mu)$ be a measure space. For a complex number $z = x + iy \in \mathbb{C}$ we write $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$ for the real and imaginary parts, respectively; $\mathscr{O}_\mathbb{C}$ denotes the usual (Euclidean) topology given by the norm $|z| = \sqrt{x^2 + y^2}$. Finally, set $g : \mathbb{C} \to \mathbb{R}^2$, $z = x + iy \mapsto (x, y)$.

(i) $\mathscr{C} := g^{-1}(\mathscr{B}(\mathbb{R}^2)) := \{g^{-1}(B) : \mathscr{B}(\mathbb{R}^2)\}$ coincides with $\mathscr{B}(\mathbb{C}) := \sigma(\mathscr{O}_\mathbb{C})$.
[Hint: $\mathscr{O}_\mathbb{C}$ is generated by the balls in $\mathbb{C}$.]

(ii) A map $h : X \to \mathbb{C}$ is $\mathscr{A}/\mathscr{C}$-measurable if, and only if, $\operatorname{Re} h$ and $\operatorname{Im} h$ are $\mathscr{A}/\mathscr{B}(\mathbb{R})$-measurable.
[Hint: the maps $z \mapsto (\operatorname{Re} z, \operatorname{Im} z)$ and $(x, y) \mapsto z = x + iy$ are continuous.]

A function $h : E \to \mathbb{C}$ is called μ-integrable, if $\operatorname{Re} h$ and $\operatorname{Im} h$ are μ-integrable; we write $\mathcal{L}^1_\mathbb{C}(\mu)$ for the complex μ-integrable functions. The integral is extended by linearity, i.e. $\int h \, d\mu := \int \operatorname{Re} h \, d\mu + i \int \operatorname{Im} h \, d\mu$. Show the following.

(iii) $h \mapsto \int h \, d\mu$ is a $\mathbb{C}$-linear map on $\mathcal{L}^1_\mathbb{C}(\mu)$.

(iv) $\operatorname{Re} \int h \, d\mu = \int \operatorname{Re} h \, d\mu$ and $\operatorname{Im} \int h \, d\mu = \int \operatorname{Im} h \, d\mu$.

(v) $\left| \int h \, d\mu \right| \leqslant \int |h| \, d\mu$.
[Hint: since $\int h \, d\mu \in \mathbb{C}$ there is some $\theta \in (-\pi, \pi]$ such that $e^{i\theta} \int h \, d\mu \geqslant 0$.]

(vi) $\mathcal{L}^1_\mathbb{C}(\mu) = \{h : E \to \mathbb{C} : h \text{ is } \mathscr{A}/\mathscr{C}\text{-measurable and } |h| \in \mathcal{L}^1_\mathbb{R}(\mu)\}$.

11

Null Sets and the 'Almost Everywhere'

Let $(X, \mathscr{A}, \mu)$ be a measure space. We are going to discuss sets and properties which are negligible for the measure μ.

Definition 11.1 A (μ-)*null set* $N \in \mathscr{N}_\mu$ is a measurable set $N \in \mathscr{A}$ satisfying

$$N \in \mathscr{N}_\mu \iff N \in \mathscr{A} \quad \text{and} \quad \mu(N) = 0. \tag{11.1}$$

If a property $\Pi = \Pi(x)$ is true for all $x \in X$ apart from some x contained in a null set $N \in \mathscr{N}_\mu$, we say that $\Pi(x)$ *holds for* (μ-)*almost all* (*a.a.*) $x \in X$ or that Π *holds* (μ-)*almost everywhere* (*a.e.*). In other words,

$$\Pi \quad \text{holds a.e.} \quad \iff \quad \{x : \Pi(x) \quad \text{fails}\} \ \subset \ N \in \mathscr{N}_\mu,$$

but we do not *a priori* require that the set $\{\Pi \text{ fails}\}$ is itself measurable. Typically we are interested in properties $\Pi(x)$ of the type: $u(x) = v(x)$, $u(x) \leqslant v(x)$, etc. and we say, for example,

$$u = v \quad \text{a.e.} \quad \iff \quad \{x : u(x) \neq v(x)\} \quad \text{is (contained in) a } \mu\text{-null set.}$$

Caution The assertions 'u enjoys a property Π a.e.' and 'u is a.e. equal to v which satisfies Π everywhere' are, in general, *far apart*; see Problem 11.8.

Theorem 11.2 *Let $u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$ be a measurable function on a measure space $(X, \mathscr{A}, \mu)$. Then*

(i) $\displaystyle\int |u| \, d\mu = 0 \iff |u| = 0 \quad a.e. \iff \mu\{u \neq 0\} = 0,$

(ii) $\mathbb{1}_N u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ *for all* $N \in \mathscr{N}_\mu$ *and* $\displaystyle\int_N u \, d\mu = 0.$

Proof Let us begin with (ii). Obviously, $\min\{|u|, n\} \uparrow |u|$ as $n \uparrow \infty$. By Beppo Levi's theorem, Theorem 9.6, we find

$$\begin{aligned}\int \mathbb{1}_N |u| \, d\mu &\overset{9.6}{=} \sup_{n\in\mathbb{N}} \int \mathbb{1}_N \min\{|u|, n\} \, d\mu \leqslant \sup_{n\in\mathbb{N}} \int n \mathbb{1}_N \, d\mu \\ &= \sup_{n\in\mathbb{N}} \left(n \int \mathbb{1}_N \, d\mu \right) = \sup_{n\in\mathbb{N}} \big(\underbrace{n\mu(N)}_{=0} \big) = 0.\end{aligned}$$

Since $\int_N u \, d\mu := \int \mathbb{1}_N u \, d\mu$, the integrability of $\mathbb{1}_N u$ follows from Theorem 10.3 and, therefore, $\left|\int_N u \, d\mu\right| = \left|\int \mathbb{1}_N u \, d\mu\right| \overset{10.4(\mathrm{v})}{\leqslant} \int \mathbb{1}_N |u| \, d\mu = 0$.

The second equivalence in (i) is clear since, due to the measurability of u, the set $\{u \neq 0\}$ is not just a subset of a null set, but measurable, and hence a proper null set. In order to see the '$\Leftarrow$' of the first equivalence, we use (ii) with $N = \{u \neq 0\}$:

$$\begin{aligned}\int |u| \, d\mu &= \int_{\{|u|\neq 0\}} |u| \, d\mu + \int_{\{|u|=0\}} |u| \, d\mu \\ &= \int_{\{|u|\neq 0\}} |u| \, d\mu + \int_{\{|u|=0\}} 0 \, d\mu \overset{(\mathrm{ii})}{=} 0.\end{aligned}$$

For '$\Rightarrow$' we use the so-called *Markov inequality*: for $A \in \mathscr{A}$ and $c > 0$ we have

$$\begin{aligned}\mu(\{|u| \geqslant c\} \cap A) &= \int \mathbb{1}_{\{|u|\geqslant c\}\cap A}(x) \mu(dx) \\ &= \int_A \frac{c}{c} \mathbb{1}_{\{|u|\geqslant c\}}(x) \mu(dx) \\ &\leqslant \frac{1}{c} \int_A |u(x)| \mathbb{1}_{\{|u|\geqslant c\}}(x) \mu(dx) \\ &\leqslant \frac{1}{c} \int_A |u(x)| \mu(dx),\end{aligned} \tag{11.2}$$

and for $A = X$ this inequality implies that

$$\begin{aligned}\mu\{|u| > 0\} &\overset{[\text{✍}]}{=} \mu\Big(\bigcup_{n\in\mathbb{N}} \{|u| \geqslant \tfrac{1}{n}\} \Big) \underset{(\mathrm{viii})}{\overset{4.3}{\leqslant}} \sum_{n\in\mathbb{N}} \mu\{|u| \geqslant \tfrac{1}{n}\} \\ &\leqslant \sum_{n\in\mathbb{N}} \Big(n \underbrace{\int |u| \, d\mu}_{=0} \Big) = 0.\end{aligned}$$

$\square$

Corollary 11.3 *Let $u, v \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$ such that $u = v$ μ-almost everywhere. Then*

(i) $u, v \geqslant 0 \implies \int u\,d\mu = \int v\,d\mu$;[1]

(ii) $u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu) \implies v \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ *and* $\int u\,d\mu = \int v\,d\mu$.

Proof Since u, v are measurable, $N := \{u \neq v\} \in \mathscr{N}_\mu$. Therefore (i) follows from

$$\begin{aligned}\int u\,d\mu &= \int_{N^c} u\,d\mu + \int_N u\,d\mu \\ &\overset{11.2(ii)}{=} \int_{N^c} v\,d\mu + 0 \qquad\qquad \text{(use that } u = v \text{ on } N^c\text{)} \\ &\overset{11.2(ii)}{=} \int_{N^c} v\,d\mu + \int_N v\,d\mu = \int v\,d\mu.\end{aligned}$$

For (ii) we observe first that $u = v$ a.e. implies that $u^\pm = v^\pm$ a.e. and then apply (i) to positive and negative parts: $\int v^\pm\,d\mu = \int u^\pm\,d\mu < \infty$; the claim follows. □

Corollary 11.4 *If $u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$ and $v \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$, $v \geqslant 0$, then*

$$|u| \leqslant v \quad \textit{a.e.} \quad \implies \quad u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu).$$

Proof We have $u^\pm \leqslant |u| \leqslant v$ a.e., and by Corollary 11.3 $\int u^\pm\,d\mu \leqslant \int v\,d\mu < \infty$. This shows that u is integrable. □

Proposition 11.5 (Markov inequality) *For all $u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$, $A \in \mathscr{A}$ and $c > 0$*

$$\mu(\{|u| \geqslant c\} \cap A) \leqslant \frac{1}{c} \int_A |u|\,d\mu, \tag{11.3}$$

and if $A = X$, in particular,

$$\mu\{|u| \geqslant c\} \leqslant \frac{1}{c} \int |u|\,d\mu. \tag{11.4}$$

Proof See (11.2) in the proof of Theorem 11.2(i). □

Corollary 11.6 *If $u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$, then u is almost everywhere $\mathbb{R}$-valued. In particular, we can find a version $\tilde{u} \in \mathcal{L}^1(\mu)$ such that $\tilde{u} = u$ a.e. and $\int \tilde{u}\,d\mu = \int u\,d\mu$.*

Proof Set $N := \{|u| = \infty\} = \{u = +\infty\} \cup \{u = -\infty\} \in \mathscr{A}$. Now

$$N = \bigcap_{n \in \mathbb{N}} \{|u| \geqslant n\}$$

[1] Including, possibly, $+\infty = +\infty$.

and by the continuity of measures proposition 4.3(vii)[2] and the Markov inequality we get

$$\mu(N) = \lim_{n\to\infty} \mu\{|u| \geqslant n\} \leqslant \lim_{n\to\infty} \left(\frac{1}{n} \underbrace{\int |u| d\mu}_{<\infty} \right) = 0.$$

The function $\tilde{u} := \mathbb{1}_{N^c} u$ is real-valued, measurable and coincides outside N with u. From Corollary 11.3 we deduce that $\tilde{u}$ is integrable with $\int \tilde{u}\, d\mu = \int u\, d\mu$. $\square$

Corollary 11.6 allows us to identify (up to null sets) functions from $\mathcal{L}^1_{\overline{\mathbb{R}}}$ and $\mathcal{L}^1$. Since $\mathcal{L}^1$ is a much nicer space – it is a vector space and we need not take any precautions when adding functions, etc. – we will work from now on only with $\mathcal{L}^1$. The corresponding statements for $\mathcal{L}^1_{\overline{\mathbb{R}}}$ are then easily derived.

We close this chapter with a technique which will be useful in many applications later on.

⁎Corollary 11.7 *Let $\mathscr{B} \subset \mathscr{A}$ be a sub-σ-algebra.*

(i) *If $u, w \in \mathcal{L}^1(\mathscr{B})$ and if $\int_B u\, d\mu = \int_B w\, d\mu$ for all $B \in \mathscr{B}$, then $u = w$ μ-a.e.*
(ii) *If $u, w \in \mathcal{M}^+(\mathscr{B})$ and if $\int_B u\, d\mu = \int_B w\, d\mu$ for all $B \in \mathscr{B}$, then $u = w$ μ-a.e. under the additional assumption that $\mu|_{\mathscr{B}}$ is σ-finite.*[3]

Proof (i) Since u and w are $\mathscr{B}$-measurable, we have that $B \cap \{u \geqslant w\}$ *and that* $B \cap \{u < w\} \in \mathscr{B}$ for all $B \in \mathscr{B}$. Thus

$$\int_B |u - w| d\mu = \int_{B\cap\{u\geqslant w\}} (u - w) d\mu + \int_{B\cap\{u<w\}} (w - u) d\mu, \tag{11.5}$$

while

$$\int_{B\cap\{u\geqslant w\}} (u - w) d\mu = \left(\int_{B\cap\{u\geqslant w\}} u\, d\mu - \int_{B\cap\{u\geqslant w\}} w\, d\mu \right) = 0 \tag{11.6}$$

by the linearity of the integral and our assumption. The other term on the right-hand side of (11.5) can be treated similarly, and the conclusion follows from Theorem 9.6(i).

(ii) If u, w are positive measurable functions, we cannot use the linearity of the integral in (11.6) as this may yield an expression of the form '$\infty - \infty$'. We can avoid this by an approximation procedure which relies on the fact that $\mu|_{\mathscr{B}}$ is σ-finite. Pick a sequence $(B_n)_{n\in\mathbb{N}} \subset \mathscr{B}$ with $B_n \uparrow X$ and $\mu(B_n) < \infty$. Then the sets $C_n := B \cap \{u \geqslant w\} \cap B_n \cap \{u \leqslant n\}$ are in $\mathscr{B}$ and have finite

[2] This is applicable since μ is finite on $\{|u| \geqslant 1\}$, say [✍].
[3] That is there exists an exhausting sequence $(B_n)_n \subset \mathscr{B}$ with $B_n \uparrow X$ and $\mu(B_n) < \infty$.

μ-measure. Furthermore, the function $(u-w)\mathbb{1}_{C_n}$ is positive and increases towards $(u-w)\mathbb{1}_{B\cap\{u\geqslant w\}}$, and $u\mathbb{1}_{C_n}, w\mathbb{1}_{C_n}$ are integrable. [✍] By Beppo Levi's theorem (Theorem 9.6),

$$\int_{B\cap\{u\geqslant w\}}(u-w)d\mu = \sup_{n\in\mathbb{N}}\int_{C_n}(u-w)d\mu = \sup_{n\in\mathbb{N}}\underbrace{\left(\int_{C_n}ud\mu - \int_{C_n}w\,d\mu\right)}_{=0\quad\forall n\in\mathbb{N}} = 0.$$

A similar argument applies to the other term in (11.5) and the claim follows. □

Problem 11.10 below shows that the σ-finiteness of $\mathscr{B}$ in Corollary 11.7(iii) is really needed.

Problems

11.1. True or false: if $f\in\mathcal{L}^1$ we can change f on a set N of measure zero (e.g. by

$$\tilde{f}(x) := \begin{cases} f(x) & \text{if } x\notin N,\\ \beta & \text{if } x\in N,\end{cases}$$

where $\beta\in\overline{\mathbb{R}}$ is any number) and $\tilde{f}$ is still integrable, even $\int f\,d\mu = \int \tilde{f}\,d\mu$.

11.2. Every countable set is a λ^1-null set. Use the Cantor ternary set C (see Problem 7.12) to illustrate that the converse is not true. What happens if we change λ^1 to λ^2?

11.3. Prove the following variants of the Markov inequality Proposition 11.5. For all $\alpha, c > 0$ and whenever the expressions involved make sense/are finite,

(i) $\mu\{|u|>c\}\leqslant \frac{1}{c}\int |u|d\mu$;

(ii) $\mu\{|u|>c\}\leqslant \frac{1}{c^p}\int |u|^p\,d\mu$ for all $0<p<\infty$;

(iii) $\mu\{|u|\geqslant c\}\leqslant \frac{1}{\phi(c)}\int \phi(|u|)d\mu$ for $\phi:\mathbb{R}_+\to\mathbb{R}_+$ increasing;

(iv) $\mu\{u\geqslant \alpha\int u d\mu\}\leqslant \frac{1}{\alpha}$ for $u\geqslant 0$;

(v) $\mu\{|u|<c\}\leqslant \frac{1}{\psi(c)}\int \psi(|u|)\,d\mu$ for $\psi:\mathbb{R}_+\to\mathbb{R}_+$ decreasing;

(vi) $\mathbb{P}\big(|\xi-\mathbb{E}\xi|\geqslant \alpha\sqrt{\mathbb{V}\xi}\big)\leqslant \frac{1}{\alpha^2}$, where $(\Omega,\mathscr{A},\mathbb{P})$ is a probability space, ξ is a random variable (i.e. a measurable function $\xi:\Omega\to\mathbb{R}$), $\mathbb{E}\xi = \int \xi\,d\mathbb{P}$ is the expectation or mean value and $\mathbb{V}\xi = \int(\xi-\mathbb{E}\xi)^2\,dP$ is the variance.
Remark. This is *Chebyshev's inequality*.

11.4. Show that $\int |u|^p\,d\mu<\infty$ implies that $|u|$ is a.e. real-valued (in the sense $(-\infty,\infty)$-valued!). Is this still true if we have $\int \arctan(u)d\mu<\infty$?

11.5. **Completion (3).** Let $(X,\overline{\mathscr{A}},\bar{\mu})$ be the completion of $(X,\mathscr{A},\mu)$, see Problems 4.15, and 6.4.

(i) Show that for every $f^*\in\mathcal{E}^+(\overline{\mathscr{A}})$ there are $f,g\in\mathcal{E}^+(\mathscr{A})$ with $f\leqslant f^*\leqslant g$ and $\mu(f\neq g)=0$ as well as $\int f\,d\mu = \int f^*\,d\bar{\mu} = \int g\,d\mu$.

(ii) Show that $u^*:X\to\overline{\mathbb{R}}$ is $\overline{\mathscr{A}}$-measurable if, and only if, there exist $\mathscr{A}$-measurable functions $u,w:X\to\overline{\mathbb{R}}$ with $u\leqslant u^*\leqslant w$ and $u=w$ μ-a.e.

(iii) Show that $u^* \in \mathcal{L}^1(\bar{\mu})$, then u, w from (ii) can be chosen from $\mathcal{L}^1(\mu)$ such that $\int u\,d\mu = \int u^*\,d\bar{\mu} = \int w\,d\mu$.

[Hint: (i) use Problem 4.15(v). (ii) for '$\Rightarrow$' consider $\{u^* > \alpha\}$ and use Problem 4.15(v). The other direction is harder. For this consider first step functions using again Problem 4.15(v) and then general functions by monotone convergence. (iii) by Problem 4.15(iii), $\mu = \bar{\mu}$ on $\mathscr{A}$, and thus $\int f\,d\mu = \int f\,d\bar{\mu}$ for $\mathscr{A}$-measurable f.]

11.6. Completion (4). Inner measure and outer measure. Let $(X, \mathscr{A}, \mu)$ be a *finite* measure space. Define for every $E \subset X$ the outer resp. inner measure

$$\mu^*(E) := \inf\{\mu(A) : A \in \mathscr{A},\ A \supset E\},$$
$$\mu_*(E) := \sup\{\mu(A) : A \in \mathscr{A},\ A \subset E\}.$$

(i) Show that for all $E, F \subset X$

$$\mu_*(E) \leqslant \mu^*(E), \qquad \mu_*(E) + \mu^*(E^c) = \mu(X),$$
$$\mu^*(E \cup F) \leqslant \mu^*(E) + \mu^*(F), \qquad \mu_*(E) + \mu_*(F) \leqslant \mu_*(E \cupdot F).$$

(ii) For every $E \subset X$ there exist sets $E_*, E^* \in \mathscr{A}$ such that $\mu(E_*) = \mu_*(E)$ and $\mu(E^*) = \mu^*(E)$.
[Hint: use the definition of '∞' to find sets $E^n \supset E$ with $\mu(E^n) - \mu^*(E) \leqslant \frac{1}{n}$ and consider $\bigcap_n E^n \in \mathscr{A}$.]

(iii) Show that $\mathscr{A}^* := \{E \subset X : \mu_*(E) = \mu^*(E)\}$ is a σ-algebra and that it is the completion of $\mathscr{A}$ w.r.t. μ. Conclude, in particular, that $\mu^*|_{\mathscr{A}^*} = \mu_*|_{\mathscr{A}^*} = \bar{\mu}$ if $\bar{\mu}$ is the completion of μ.

11.7. Let $(X, \mathscr{A}, \mu)$ be a measure space and assume that $u \in \mathcal{M}(\mathscr{A})$ and $u = w$ almost everywhere w.r.t. μ. When can we say that $w \in \mathcal{M}(\mathscr{A})$?

11.8. 'a.e.' is a tricky business. When working with 'a.e.' properties one has to be extremely careful. For example, the assertions 'u is continuous a.e.' and 'u is a.e. equal to an (everywhere) continuous function' are *far apart*! Illustrate this by considering the functions $u = \mathbb{1}_{\mathbb{Q}}$ and $u = \mathbb{1}_{[0,\infty)}$.

11.9. Let μ be a σ-finite measure on the measurable space $(X, \mathscr{A})$. Show that there exists a finite measure P on $(X, \mathscr{A})$ such that $\mathscr{N}_\mu = \mathscr{N}_P$, i.e. μ and P have the same null sets.

11.10. Construct an example showing that for $u, w \in \mathcal{M}^+(\mathscr{B})$ the equality $\int_B u\,d\mu = \int_B w\,d\mu$ for all $B \in \mathscr{B}$ does not necessarily imply that $u = w$ almost everywhere.
[Hint: in view of Corollary 11.7 $\mu|_{\mathscr{B}}$ cannot be σ-finite. Consider on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ the measure $\mu = m\lambda^1$, where $m = \mathbb{1}_{\{|x| \leqslant 1\}} + \infty \mathbb{1}_{\{|x| > 1\}}$, $u \equiv 1$ and $w = \mathbb{1}_{\{|x| \leqslant 1\}} + 2\mathbb{1}_{\{|x| > 1\}}$. Then all Borel subsets of $\{|x| > 1\}$ have μ-measure either 0 or $+\infty$, thus $\int_B u\,d\mu = \int_B w\,d\mu$ for all $B \in \mathscr{B}(\mathbb{R})$ while $\mu(u \neq w) = \infty$.]

11.11. Show the following extension of Corollary 11.7. Let $\mathscr{C} \subset \mathscr{P}(X)$ be a $\cap$-stable generator of $\mathscr{A}$ which contains a sequence $C_n \uparrow X$ such that $\mu(C_n) < \infty$. For all $u, w \in \mathcal{L}^1(\mu)$ we have

$$\int_C u\,d\mu = \int_C w\,d\mu \quad \forall C \in \mathscr{C} \iff u = w \quad \mu\text{-a.e.}$$

[Hint: use the uniqueness theorem for measures for $C \mapsto \int_C (u^\pm + w^\mp)\,d\mu$ and Corollary 11.7.]

11.12. Egorov's theorem. Let $(X, \mathscr{A}, \mu)$ be a finite measure space and $f_n : X \to \mathbb{R}$, $n \in \mathbb{N}$, a sequence of measurable functions. Prove the following assertions.

(i) $C_f := \{x \in X : f(x) = \lim_n f_n(x) \text{ exists}\}$
$$= \bigcap_{k=1}^{\infty} \bigcup_{\ell=1}^{\infty} \bigcap_{m=\ell}^{\infty} \bigcap_{n=\ell}^{\infty} \left\{ |f_m - f_n| \leqslant \frac{1}{k} \right\}.$$

(ii) Assume that $\mu(X \setminus C_f) = 0$, i.e. $f_n(x) \to f(x)$ for all x outside a null set. Then we have for the sets $A_n^k := \bigcup_{\ell=1}^{n} \bigcap_{m=\ell}^{\infty} \{ |f_m - f| \leqslant \frac{1}{k} \}$
$$\forall \epsilon > 0 \ \ \forall k \in \mathbb{N} \ \ \exists n(k, \epsilon) \in \mathbb{N} \ : \ \mu(X \setminus A_{n(k,\epsilon)}^k) \leqslant \epsilon 2^{-k}.$$

(iii) **Theorem** (Egorov). Let $(X, \mathscr{A}, \mu)$ be a finite measure space and $(f_n)_{n \in \mathbb{N}}$ a sequence of measurable functions which converges pointwise (everywhere or outside a null set) to a function f. Then there is for every $\epsilon > 0$ a set $A_\epsilon \in \mathscr{A}$, $\mu(X \setminus A_\epsilon) \leqslant \epsilon$, such that $\lim_{n\to\infty} \sup_{x \in A_\epsilon} |f_n(x) - f(x)| = 0$, i.e. $f_n \to f$ uniformly on A_ϵ.

(iv) The finiteness of μ is essential for Egorov's theorem.
[Hint: use the counting measure on $\mathbb{N}$ or Lebesgue's measure on $\mathbb{R}$.]

12

Convergence Theorems and Their Applications

Throughout this chapter $(X, \mathscr{A}, \mu)$ will be a measure space. One of the shortcomings of the Riemann integral is the fact that we do not have sufficiently general results that allow us to interchange limits and integrals – typically one has to assume uniform convergence for this. This has to do with the fact that the set of Riemann integrable functions is somewhat limited, see Theorem 12.9. The classical example of a non-Riemann integrable function is *Dirichlet's jump function* $x \mapsto \mathbb{1}_{\mathbb{Q}\cap[0,1]}(x)$: its upper function is $\mathbb{1}_{[0,1]}$ while the lower function is $0 \cdot \mathbb{1}_{[0,1]}$. [✍]

For the Lebesgue integral on $\mathcal{M}^{+}_{\overline{\mathbb{R}}}(\mathscr{A})$ we have already seen more powerful convergence results in the form of Beppo Levi's theorem (Theorem 9.6) or Fatou's lemma (Theorem 9.11). They can deal with Dirichlet's jump function: for any enumeration of $\mathbb{Q} = \{q_n : n \in \mathbb{N}\}$ we get

$$\begin{aligned}\int \mathbb{1}_{\mathbb{Q}\cap[0,1]}\, d\lambda^1 &= \int \sup_{N\in\mathbb{N}} \mathbb{1}_{\{q_1,\dots,q_N\}\cap[0,1]}\, d\lambda^1 \\ &\stackrel{9.6}{=} \sup_{N\in\mathbb{N}} \int \mathbb{1}_{\{q_1,\dots,q_N\}\cap[0,1]}\, d\lambda^1 \\ &= \sup_{N\in\mathbb{N}} \underbrace{\lambda^1\{q_n \in [0,1] : 1 \leqslant n \leqslant N\}}_{=0} = 0.\end{aligned}$$

In this chapter we study systematically convergence theorems for $\mathcal{L}^1(\mu)$ and some of their most important applications. The first result is a generalization of Beppo Levi's theorem (Theorem 9.6).

Theorem 12.1 (monotone convergence) *Let $(X, \mathscr{A}, \mu)$ be a measure space.*

(i) *Let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mu)$ be an increasing sequence of integrable functions $u_1 \leqslant u_2 \leqslant \cdots$ with limit $u := \sup_{n\in\mathbb{N}} u_n = \lim_{n\to\infty} u_n$. Then $u \in \mathcal{L}^1(\mu)$ if, and* *only if,* $\sup_{n\in\mathbb{N}} \int u_n \, d\mu < +\infty$*, in which case*

$$\sup_{n\in\mathbb{N}} \int u_n \, d\mu = \int \sup_{n\in\mathbb{N}} u_n \, d\mu.$$

(ii) *Let* $(v_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mu)$ *be a decreasing sequence of integrable functions* $v_1 \geqslant v_2 \geqslant \cdots$ *with limit* $v := \inf_{n\in\mathbb{N}} v_n = \lim_{n\to\infty} v_n$*. Then* $v \in \mathcal{L}^1(\mu)$ *if, and only if,* $\inf_{n\in\mathbb{N}} \int v_n \, d\mu > -\infty$*, in which case*

$$\inf_{n\in\mathbb{N}} \int v_n \, d\mu = \int \inf_{n\in\mathbb{N}} v_n \, d\mu.$$

Proof Obviously, (i) implies (ii) as $u_n := -v_n$ fulfils all the assumptions of (i). To see (i), we remark that $u_n - u_1 \in \mathcal{L}^1(\mu)$ defines an increasing sequence of *positive* functions

$$0 \leqslant u_n - u_1 \leqslant u_{n+1} - u_1 \leqslant \cdots,$$

for which we may use the Beppo Levi theorem (Theorem 9.6):

$$0 \leqslant \sup_{n\in\mathbb{N}} \int (u_n - u_1) d\mu = \int \sup_{n\in\mathbb{N}} (u_n - u_1) d\mu. \tag{12.1}$$

Assume that $u \in \mathcal{L}^1(\mu)$. Since the 'sup' in (12.1) is an increasing limit, we get

$$\begin{aligned} \sup_{n\in\mathbb{N}} \int u_n \, d\mu &= \int (u - u_1) d\mu + \int u_1 \, d\mu \\ &\overset{10.2}{=} \int u \, d\mu - \int u_1 \, d\mu + \int u_1 \, d\mu = \int u \, d\mu < \infty. \end{aligned}$$

Conversely, if $\sup_{n\in\mathbb{N}} \int u_n \, d\mu < \infty$, we see from (12.1) that $u - u_1 \in \mathcal{L}^1(\mu)$ and, as $u_1 \in \mathcal{L}^1(\mu)$, $u = (u - u_1) + u_1 \in \mathcal{L}^1(\mu)$ by (10.2). Therefore, (12.1) implies

$$\int u \, d\mu = \int (u - u_1) d\mu + \int u_1 \, d\mu = \sup_{n\in\mathbb{N}} \int u_n \, d\mu < \infty. \qquad \square$$

One of the most useful and versatile convergence theorems is the following.

Theorem 12.2 (Lebesgue; dominated convergence) *Let* $(X, \mathscr{A}, \mu)$ *be a measure space and* $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mu)$ *be a sequence of functions such that*

(a) $|u_n(x)| \leqslant w(x)$ *for all* $n \in \mathbb{N}, x \in X$ *and some* $w \in \mathcal{L}^1(\mu)$*,*
(b) $u(x) = \lim_{n\to\infty} u_n(x)$ *exists in* $\overline{\mathbb{R}}$ *for all* $x \in X$*,*

then $u \in \mathcal{L}^1(\mu)$ *and we have*

(i) $\displaystyle\lim_{n\to\infty} \int |u_n - u| d\mu = 0;$

(ii) $\displaystyle\lim_{n\to\infty} \int u_n \, d\mu = \int \lim_{n\to\infty} u_n \, d\mu = \int u \, d\mu.$

Proof From $|u_n| \leqslant w$ we get $|u| = \lim_{n\to\infty} |u_n| \leqslant w$, and $u \in \mathcal{L}^1(\mu)$ by Corollary 11.4(iv). Therefore,

$$\left|\int u_n\, d\mu - \int u\, d\mu\right| = \left|\int (u_n - u) d\mu\right| \overset{10.4(\mathrm{v})}{\leqslant} \int |u_n - u|\, d\mu,$$

which means that (i) implies (ii). Since

$$|u_n - u| \leqslant |u_n| + |u| \leqslant 2w \quad \forall n \in \mathbb{N},$$

we get $2w - |u_n - u| \geqslant 0$, and Fatou's lemma (Theorem 9.11) tells us that

$$\begin{aligned}\int 2w\, d\mu &= \int \liminf_{n\to\infty}(2w - |u_n - u|) d\mu \\ &\leqslant \liminf_{n\to\infty} \int (2w - |u_n - u|) d\mu \\ &= \int 2w\, d\mu - \limsup_{n\to\infty} \int |u_n - u| d\mu.^1\end{aligned}$$

Thus $0 \leqslant \liminf_{n\to\infty} \int |u_n - u| d\mu \leqslant \limsup_{n\to\infty} \int |u_n - u| d\mu \leqslant 0$, and consequently $\lim_{n\to\infty} \int |u_n - u| d\mu = 0$. □

Remark 12.3 (i) We can replace in Theorem 12.2(a), (b) the expression 'for all $x \in X$' with 'for μ-almost all $x \in X$'. If we do this,

$$N := \left\{x : \lim_{n\to\infty} u_n(x) \text{ does not exist}\right\} \cup \bigcup_{n\in\mathbb{N}} \{x : |u_n(x)| > w(x)\}$$

is a measurable null set [✍] since the functions u_n and w are measurable. The functions $u\mathbb{1}_{N^c}$ and $u_n\mathbb{1}_{N^c}$ satisfy the assumptions of Theorem 12.2 for *all* $x \in X$ and with the majorant $w\mathbb{1}_{N^c}$; their integrals coincide with those of u and u_n.

(ii) The assumption of uniform boundedness by a fixed majorant

$$|u_n| \leqslant w \quad \forall n \in \mathbb{N} \quad \text{and some} \quad w \in \mathcal{L}^1(\mu) \tag{12.2}$$

is essential for Theorem 12.2. To see this, consider $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda^1)$ and set

$$u_n(x) := n\mathbb{1}_{\left[0,\frac{1}{n}\right]}(x) \xrightarrow[n\to\infty]{} \infty\mathbb{1}_{\{0\}}(x) \overset{\text{a.e.}}{=} 0,$$

whereas $\int u_n\, d\lambda = n\frac{1}{n} = 1 \neq 0 = \int \infty\mathbb{1}_{\{0\}}\, d\lambda$.

[1] Recall that $\liminf_{n\to\infty}(-x_n) = -\limsup_{n\to\infty} x_n$.

(iii) Lebesgue's theorem gives merely sufficient – but easily verifiable – conditions for the interchange of limits and integrals; the ultimate version of such a result with necessary and sufficient conditions will be given in the form of Vitali's convergence theorem (Theorem 22.7) in Chapter 22 below.

Application 1: Parameter-Dependent Integrals

Assume that $u(t,x)$ is a function depending on $x \in X$ and on some parameter $t \in (a,b)$. Often this dependence is 'smooth' (e.g. continuous, differentiable …) and we will now ask when integration in x preserves this smoothness, i.e. when is

$$U(t) := \int u(t,x)\mu(dx), \quad t \in (a,b),$$

again 'smooth' (e.g. continuous, differentiable …) as a function of t? Obviously this is a problem which involves the interchange of limits and integration. As usual, we work in some measure space $(X, \mathscr{A}, \mu)$.

Theorem 12.4 (continuity lemma) *Let $\emptyset \neq (a,b) \subset \mathbb{R}$ be a non-degenerate open interval and $u : (a,b) \times X \to \mathbb{R}$ be a function satisfying*

(a) *$x \mapsto u(t,x)$ is in $\mathcal{L}^1(\mu)$ for every fixed $t \in (a,b)$;*
(b) *$t \mapsto u(t,x)$ is continuous for every fixed $x \in X$;*
(c) *$|u(t,x)| \leqslant w(x)$ for all $(t,x) \in (a,b) \times X$ and some $w \in \mathcal{L}^1(\mu)$.*

Then the function $U : (a,b) \to \mathbb{R}$ given by

$$t \mapsto U(t) := \int u(t,x)\mu(dx) \tag{12.3}$$

is continuous.

Proof Let us, first of all, note that (12.3) is well-defined thanks to assumption (a). We will show that for any $t \in (a,b)$ and every sequence $(t_n)_{n\in\mathbb{N}} \subset (a,b)$ with $\lim_{n\to\infty} t_n = t$ we have $\lim_{n\to\infty} U(t_n) = U(t)$. This proves continuity of U at the point t.

Because of (b), $u(\cdot\,,x)$ is continuous and, therefore,

$$u_n(x) := u(t_n,x) \xrightarrow[n\to\infty]{} u(t,x) \quad \text{and} \quad |u_n(x)| \leqslant w(x) \quad \forall x \in X.$$

Thus we can use Lebesgue's dominated convergence theorem, and conclude

$$\begin{aligned}\lim_{n\to\infty} U(t_n) &= \lim_{n\to\infty} \int u(t_n, x)\mu(dx)\\ &= \int \lim_{n\to\infty} u(t_n, x)\mu(dx)\\ &= \int u(t, x)\mu(dx) = U(t).\end{aligned}$$ □

A very similar consideration leads to the following theorem.

Theorem 12.5 (differentiability lemma) *Let $\emptyset \neq (a, b) \subset \mathbb{R}$ be a non-degenerate open interval and $u: (a, b) \times X \to \mathbb{R}$ be a function satisfying*

(a) *$x \mapsto u(t, x)$ is in $\mathcal{L}^1(\mu)$ for every fixed $t \in (a, b)$;*
(b) *$t \mapsto u(t, x)$ is differentiable for every fixed $x \in X$;*
(c) *$|\partial_t u(t, x)| \leqslant w(x)$ for all $(t, x) \in (a, b) \times X$ and some $w \in \mathcal{L}^1(\mu)$.*

Then the function $U: (a, b) \to \mathbb{R}$ given by

$$t \mapsto U(t) := \int u(t, x)\mu(dx) \tag{12.4}$$

is differentiable and its derivative is

$$\frac{d}{dt}U(t) = \frac{d}{dt}\int u(t, x)\mu(dx) = \int \frac{\partial}{\partial t}u(t, x)\mu(dx). \tag{12.5}$$

Proof Let $t \in (a, b)$ and fix some sequence $(t_n)_{n\in\mathbb{N}} \subset (a, b)$ such that $t_n \neq t$ and $\lim_{n\to\infty} t_n = t$. Set

$$u_n(x) := \frac{u(t_n, x) - u(t, x)}{t_n - t} \xrightarrow[n\to\infty]{} \frac{\partial}{\partial t}u(t, x);$$

this shows, in particular, that $x \mapsto \frac{\partial}{\partial t}u(t, x)$ is measurable. By virtue of the mean value theorem of differential calculus and (c) there is some intermediate value $\theta = \theta(n, x) \in (a, b)$ such that

$$|u_n(x)| = \left|\frac{\partial}{\partial t}u(t, x)\Big|_{t=\theta}\right| \leqslant w(x) \quad \forall n \in \mathbb{N}.$$

Thus $u_n \in \mathcal{L}^1(\mu)$, and the sequence $(u_n)_{n\in\mathbb{N}}$ satisfies all conditions of the dominated convergence theorem (Theorem 12.2). Finally,

$$\begin{aligned} U'(t) &= \lim_{n\to\infty} \frac{U(t_n) - U(t)}{t_n - t} = \lim_{n\to\infty} \int \frac{u(t_n, x) - u(t, x)}{t_n - t} \mu(dx) \\ &= \lim_{n\to\infty} \int u_n(x)\mu(dx) \\ &\overset{12.2}{=} \int \lim_{n\to\infty} u_n(x)\mu(dx) \\ &= \int \frac{\partial}{\partial t} u(t, x)\, \mu(dx). \end{aligned}$$

□

Below we will give examples of how to apply the continuity and differentiability lemmas.

Application 2: Riemann vs. Lebesgue Integration

From here to the end of this chapter we use $(X, \mathscr{A}, \mu) = (\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$. One motivation for Lebesgue to develop his approach to integration was the existence of certain inadequacies of the Riemann integral. On the other hand, the notions of Lebesgue and Riemann integrals should be compatible for a large class of functions, so that we can evaluate integrals 'as usual'. As it turns out, we need Lebesgue's approach to integration if we want to characterize the set of functions which are Riemann integrable. We will now discuss these connections.

Let us briefly recall the definition of the *Riemann integral* (see Appendix I for a complete discussion). Consider on the finite interval $[a, b] \subset \mathbb{R}$ the partition

$$\Pi := \{a = t_0 < t_1 < \cdots < t_k = b\}, \quad k = k(\Pi),$$

and introduce the *lower* resp. *upper Darboux sums*

$$S_\Pi[u] := \sum_{i=1}^{k(\Pi)} m_i(t_i - t_{i-1}), \qquad m_i := \inf_{x\in[t_{i-1},t_i]} u(x),$$

$$S^\Pi[u] := \sum_{i=1}^{k(\Pi)} M_i(t_i - t_{i-1}), \qquad M_i := \sup_{x\in[t_{i-1},t_i]} u(x).$$

See Fig. 12.1

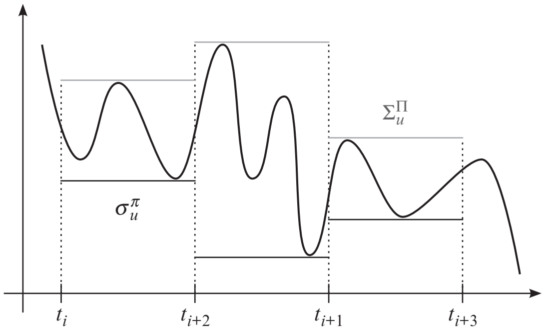


Fig. 12.1. Upper and lower functions for the Riemann integral.

Definition 12.6 A bounded function $u: [a,b] \to \mathbb{R}$ is said to be *Riemann integrable* if the values

$$\underline{\int} u := \sup_{\Pi} S_{\Pi}[u] = \inf_{\Pi} S^{\Pi}[u] =: \overline{\int} u$$

(sup, inf range over all finite partitions Π of $[a,b]$) coincide and are finite. Their common value is called the *Riemann integral* of u and denoted by $(R)\int_a^b u(x)dx$ or $\int_a^b u(x)dx$.

What is going on here? First of all, it is not difficult to see that lower [upper] Darboux sums increase [decrease] if we add points to the partition Π [✍], i.e. the sup [inf] in Definition 12.6 makes sense.

Moreover, to $S_{\Pi}[u]$ and $S^{\Pi}[u]$ there correspond simple functions, namely σ_u^{Π} and Σ_u^{Π} given by

$$\sigma_u^{\Pi}(x) = \sum_{i=1}^{k} m_i \mathbb{1}_{[t_{i-1},t_i)}(x) \quad \text{and} \quad \Sigma_u^{\Pi}(x) = \sum_{i=1}^{k} M_i \mathbb{1}_{[t_{i-1},t_i)}(x)$$

which satisfy $\sigma_u^{\Pi}(x) \leqslant u(x) \leqslant \Sigma_u^{\Pi}(x)$ and which increase resp. decrease as Π refines.

Remark 12.7 The above construction gives the ‘usual’ integral which is often introduced as the anti-derivative. Unfortunately, this notion of integration is somewhat insufficient. Nice general convergence theorems (such as monotone or dominated convergence) hold only under unnatural restrictions or are not available at all. Moreover, it cannot deal with functions of the type $x \mapsto \mathbb{1}_{\mathbb{Q}\cap[0,1]}(x)$: the smallest upper function Σ^{Π} is $\mathbb{1}_{[0,1]}$, while the largest lower function σ_{Π} is identically 0. [✍] Thus the Riemann integral of $\mathbb{1}_{\mathbb{Q}\cap[0,1]}$ does not exist, whereas by Theorem 11.2(ii) the Lebesgue integral $\int \mathbb{1}_{\mathbb{Q}\cap[0,1)}\,d\lambda = 0$. See Fig. 12.2.

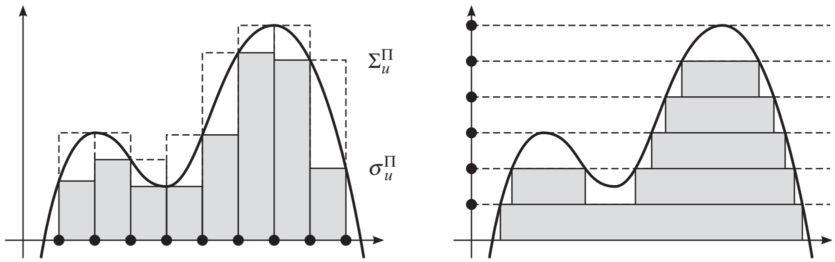


Fig. 12.2. Comparison of the Lebesgue and Riemann integrals: the Riemann sums partition the domain of the function without taking into account the shape of the function, thus slicing up the area under the function *vertically*. Lebesgue's approach is exactly the opposite: the domain is partitioned according to the values of the function at hand, leading to a *horizontal* decomposition of the area.

There is a beautifully simple connection with Lebesgue integrals which characterizes at the same time the class of Riemann integrable functions. It may come as a surprise that one needs the notion of Lebesgue null sets to understand Riemann's integral completely.

Theorem 12.8 *Let $u : [a,b] \to \mathbb{R}$ be a measurable and Riemann integrable function. Then we have*

$$u \in \mathcal{L}^1(\lambda) \quad \textit{and} \quad \int_{[a,b]} u\,d\lambda = (R)\int_a^b u(x)dx.$$

Proof As u is Riemann integrable, we find a sequence of partitions $\Pi(i)$ of $[a,b]$ such that

$$\lim_{i\to\infty} S_{\Pi(i)}[u] = \underline{\int} u = \overline{\int} u = \lim_{i\to\infty} S^{\Pi(i)}[u].$$

Without loss of generality we may assume that $\Pi(i) \subset \Pi(i+1) \subset \cdots$ – otherwise we use the increasing sequence $\Pi(1) \cup \cdots \cup \Pi(i)$, where we also observe that the lower [upper] Riemann sums increase [decrease] as the partitions refine. The corresponding simple functions $\sigma_u^{\Pi(i)}$ and $\Sigma_u^{\Pi(i)}$ converge monotonically towards

$$\sigma_u := \sup_{i\in\mathbb{N}} \sigma_u^{\Pi(i)} \leqslant u \leqslant \inf_{i\in\mathbb{N}} \Sigma_u^{\Pi(i)} =: \Sigma_u,$$

and from the monotone convergence theorem (Theorem 12.1) we conclude that

$$\underline{\int} u = \lim_{i\to\infty} S_{\Pi(i)}[u] = \lim_{i\to\infty} \int_{[a,b]} \sigma_u^{\Pi(i)}\, d\lambda = \int_{[a,b]} \sigma_u \, d\lambda \tag{12.6}$$

as well as

$$\overline{\int} u = \lim_{i\to\infty} S^{\Pi(i)}[u] = \lim_{i\to\infty} \int_{[a,b]} \Sigma_u^{\Pi(i)}\, d\lambda = \int_{[a,b]} \Sigma_u \, d\lambda. \tag{12.7}$$

In other words $\sigma_u, \Sigma_u \in \mathcal{L}^1(\lambda)$. Since u is Riemann integrable,

$$\int_{[a,b]} \underbrace{(\Sigma_u - \sigma_u)}_{\geqslant 0} d\lambda = \int_{[a,b]} \Sigma_u \, d\lambda - \int_{[a,b]} \sigma_u \, d\lambda = \overline{\int} u - \underline{\int} u = 0,$$

which implies by Theorem 11.2(i) that $\Sigma_u = \sigma_u$ Lebesgue a.e. Thus

$$\{u \neq \Sigma_u\} \cup \{u \neq \sigma_u\} \subset \{\sigma_u \neq \Sigma_u\} \in \mathcal{N}_\lambda, \tag{12.8}$$

and by Corollary 11.3(ii) we conclude that u is Lebesgue integrable. $\square$

The proof of Theorem 12.8 shows a bit more: if u is Riemann integrable but not necessarily Borel measurable, then there is some $\Sigma_u \in \mathcal{L}^1(\lambda)$ such that $u = \Sigma_u$ Lebesgue a.e. and $(R) \int_a^b u(x)dx = \int_{[a,b]} \Sigma_u \, d\lambda$.

Theorem 12.9 *Let $u : [a,b] \to \mathbb{R}$ be a bounded function. The function u is Riemann integrable if, and only if, the points in (a,b) where u is discontinuous are a (subset of a) Borel measurable null set.*

Caution Theorem 12.9 is often phrased in the following way: *f is Riemann integrable if, and only if, f is (Lebesgue) a.e. continuous*. Although correct, this is a dangerous way of putting things since one is led to read this statement (incorrectly) as 'if $f = \phi$ a.e. with $\phi \in C[a,b]$, then f is Riemann integrable'. That this is wrong is easily seen from $f = \mathbb{1}_{\mathbb{Q}\cap[a,b]}$ and $\phi \equiv 0$; see Problem 11.8 and 12.33.

Proof of Theorem 12.9 '$\Rightarrow$': Assume that u is Riemann integrable and let $\Pi(i)$, σ_u and Σ_u be as in the proof of Theorem 12.8. Using the very definition of sup and inf we find for every $\epsilon > 0$ and all $x \in X$ some $n(\epsilon, x) \in \mathbb{N}$ and some indices $t_{n_0-1}, t_{n_0} \in \Pi(n(\epsilon, x))$ such that the following conditions hold:

(a) $x \in [t_{n_0-1}, t_{n_0}]$,
(b) $\left|\sigma_u^{\Pi(i)}(x) - \sigma_u(x)\right| + \left|\Sigma_u^{\Pi(i)}(x) - \Sigma_u(x)\right| \leqslant \epsilon$ for all $i \geqslant n_{\epsilon,x}$.

For any $x \in [a,b] \setminus \bigcup_{n\in\mathbb{N}} \Pi(n)$ and all $y \in (t_{n_0-1}, t_{n_0})$ we get

$$\begin{aligned}|u(x) - u(y)| \leqslant M_{n_0} - m_{n_0} &= \Sigma_u^{\Pi(n_\epsilon,x)}(x) - \sigma_u^{\Pi(n_\epsilon,x)}(x)\\ &\leqslant \epsilon + |\Sigma_u(x) - \sigma_u(x)|.\end{aligned}$$

As in the proof of Theorem 12.8 we conclude from the Riemann integrability of u that $\{\Sigma_u \neq \sigma_u\}$ is a λ-null set, and so

$$\{x : u \text{ is not continuous at } x\} \subset \underbrace{\bigcup_{n\in\mathbb{N}} \Pi(n)}_{\text{countable set}} \cup \overbrace{\{\Sigma_u \neq \sigma_u\}}^{\lambda\text{-null set}} \in \mathscr{N}_\lambda.$$

'$\Leftarrow$': Assume now that $D := \{x : u(x) \text{ is not continuous}\}$ is a subset of a λ-null set. For every $x \in D$ and any partition $\Pi \subset [a,b]$ we can find some $k = k(x,\Pi) \in \mathbb{N}$ such that $x \in [t_{k-1}, t_k]$. Consequently,

$$\Sigma_u(x) - \sigma_u(x) \leqslant M_k - m_k \xrightarrow[|\Pi| := \max_i |t_i - t_{i-1}| \downarrow 0]{u \text{ continuous at } x} 0.$$

This shows that $\{\Sigma_u = \sigma_u\} \supset \{x : u \text{ continuous at } x\}$ or, equivalently, that $\{\Sigma_u \neq \sigma_u\} \subset \{x : u \text{ not continuous at } x\}$. In particular, the measurable set $\{\Sigma_u \neq \sigma_u\}$ is a subset of a null set, and we get

$$\underline{\int} u \overset{\text{def}}{=} \int \Sigma_u \, d\lambda = \int \sigma_u \, d\lambda \overset{\text{def}}{=} \overline{\int} u,$$

proving that u is Riemann integrable. $\square$

Remark 12.10 The set of continuity points $\{x \in [a,b] : u \text{ is continuous at } x\}$ of any function $u : [a,b] \to \mathbb{R}$ is a Borel set, see Appendix G.

Improper Riemann Integrals

Although the Lebesgue integral extends the *proper* Riemann integral, there is a further extension of the Riemann integral which cannot be captured by Lebesgue's theory. Recall that a measurable function u is Lebesgue integrable if, and only if, $|u|$ has finite Lebesgue integral. This means that the Lebesgue integral does not respect sign-changes and cancellations. The following *improper Riemann integral*, however, enjoys this property. Here we consider only improper Riemann integrals of the form

$$(R)\int_0^\infty u(x)dx := \lim_{a\to\infty} (R)\int_0^a u(x)dx, \tag{12.9}$$

provided that the limit exists (see Appendix I for other types of improper integrals). Here is a typical situation where improper Riemann and Lebesgue integrals coincide.

Corollary 12.11 *Let $u : [0,\infty) \to \mathbb{R}$ be a measurable function which is Riemann integrable for every interval $[0,N]$, $N \in \mathbb{N}$. Then $u \in \mathcal{L}^1[0,\infty)$ if, and only if,*

$$\lim_{N\to\infty} (R) \int_0^N |u(x)|dx < \infty. \tag{12.10}$$

In this case, $(R)\int_0^\infty u(x)dx = \int_{[0,\infty)} u\,d\lambda$.

Proof Using Theorem 12.8 we see that Riemann integrability of u implies Riemann integrability of $u^\pm$. [✍] Moreover,

$$(R)\int_0^N u^\pm(x)dx = \int_{[0,N]} u^\pm(x)\lambda(dx) = \int u^\pm \mathbb{1}_{[0,N]}\,d\lambda. \tag{12.11}$$

If u is Riemann integrable and satisfies (12.9) and (12.10), the limit $N\to\infty$ of the left-hand side of (12.11) exists and guarantees that the right-hand side has also a finite limit. The monotone convergence theorem, Theorem 12.1, together with Theorem 10.3(ii) shows that $u \in \mathcal{L}^1[0,\infty)$.

Conversely, if u is Lebesgue integrable, then so are $u^\pm, u\mathbb{1}_{[0,a]}$ and $u^\pm\mathbb{1}_{[0,a]}$ for every $a > 0$. Since u is Riemann integrable over each interval $[0,N]$, we see from Theorem 12.9 that u and $u^\pm$ are Riemann integrable over each interval $[0,a]$. The monotone convergence theorem (Theorem 12.1) shows that for *every* increasing sequence $a_i \uparrow \infty$

$$\lim_{i\to\infty} \int u^\pm \mathbb{1}_{[0,a_i]}\,d\lambda = \int_{[0,\infty)} u^\pm\,d\lambda < \infty,$$

which yields that the limits (12.9), (12.10) exist. □

The following example – taken from the theory of Fourier series – highlights a typical situation where the improper Riemann integral exists, but the Lebesgue integral doesn't.

Example 12.12 Lebesgue integration does not allow cancellations, but improper Riemann integrals do. More precisely: the limit (12.9) can make sense even if $\lim_{N\to\infty}(R)\int_0^N |u(x)|dx = \infty$.

The function $x \mapsto s(x) := \frac{\sin x}{x}$, $x \in (0,\infty)$, is improperly Riemann integrable but not Lebesgue integrable.

Indeed: for $a > 0$ we can find $N = N(a) \in \mathbb{N}$ such that $N\pi \leqslant a < (N+1)\pi$. Thus

$$\int_0^\infty \frac{\sin x}{x}\,dx = \lim_{a\to\infty}\left(\int_0^{N\pi} \frac{\sin x}{x}\,dx + \int_{N\pi}^a \frac{\sin x}{x}\,dx\right)$$
$$= \lim_{N\to\infty} \sum_{i=0}^{N-1} \underbrace{\int_{i\pi}^{(i+1)\pi} \frac{\sin x}{x}\,dx}_{=:\,a_i},$$

where we use

$$\left|\lim_{a\to\infty} \int_{N\pi}^a \frac{\sin x}{x}\,dx\right| \leqslant \lim_{N\to\infty} \int_{N\pi}^{(N+1)\pi} \left|\frac{\sin x}{x}\right|\,dx \leqslant \lim_{N\to\infty} \frac{\pi}{N\pi} = 0.$$

Observe that the a_i have alternating signs since

$$a_i = \int_{i\pi}^{(i+1)\pi} \frac{\sin x}{x}\,dx = \int_0^\pi \frac{\sin(y+i\pi)}{y+i\pi}\,dy = (-1)^i \int_0^\pi \frac{\sin y}{y+i\pi}\,dy$$

both as Riemann and as Lebesgue integrals, by Theorem 12.8. Furthermore,

$$|a_i| = \int_0^\pi \frac{\sin y}{y+i\pi}\,dy \leqslant \int_0^\pi \frac{\sin y}{y+iy}\,dy = \frac{1}{i+1}\int_0^\pi \frac{\sin y}{y}\,dy,$$

and also

$$|a_i| = \int_0^\pi \frac{\sin y}{y+i\pi}\,dy \geqslant \overbrace{\int_0^\pi \frac{\sin y}{y+(i+1)\pi}\,dy}^{=\,|a_{i+1}|}$$
$$\geqslant \int_0^\pi \frac{\sin y}{\pi+(i+1)\pi}\,dy = \frac{2}{(i+2)\pi}.$$

Since the function $y \mapsto \frac{\sin y}{y}$ is continuous and has a finite limit as $y \downarrow 0$, [✍] we see that $C := \int_0^\pi \frac{\sin y}{y}\,dy < \infty$, so that

$$\frac{2/\pi}{i+2} \leqslant |a_{i+1}| \leqslant |a_i| \leqslant \frac{C}{i+1}.$$

This and Leibniz's convergence test prove that the alternating series $\sum_{i=0}^\infty a_i$ converges conditionally but not absolutely, i.e. we get a finite improper Riemann integral, but the Lebesgue integral does not exist.

Examples

We have seen in this chapter that the Lebesgue integral provides very powerful tools which justify the interchange of limits and integrals. On the other hand, the Riemann theory is quite handy when it comes to calculating the primitive (anti-derivative) of some concrete integrand. Theorem 12.9 tells us when we can switch between these two notions.

Example 12.13 *Let* $f_\alpha(x) := x^\alpha$, $x > 0$ *and* $\alpha \in \mathbb{R}$. *Then*

(i) $f_\alpha \in \mathcal{L}^1(0,1) \iff \alpha > -1$.
(ii) $f_\alpha \in \mathcal{L}^1[1,\infty) \iff \alpha < -1$.

We show only (i); the proof of (ii) is very similar. Since f_α is continuous, it is Borel measurable, and since $f_\alpha \geqslant 0$ it is enough to evaluate $\int_{(0,1)} f_\alpha \, d\lambda$. We find

$$\begin{aligned}
\int_{(0,1)} x^\alpha \, dx &\overset{9.6}{=} \lim_{n\to\infty} \int x^\alpha \mathbb{1}_{[1/n,1)}(x)\lambda(dx) \\
&\overset{12.9}{=} \lim_{n\to\infty} (R)\int_{1/n}^{1} x^\alpha \, dx \\
&= \lim_{n\to\infty} \left[\frac{x^{\alpha+1}}{\alpha+1}\right]_{1/n}^{1} \\
&= \lim_{n\to\infty} \left(\frac{1}{\alpha+1} - \frac{1}{n^{\alpha+1}(\alpha+1)}\right),
\end{aligned}$$

and the last limit is finite if, and only if, $\alpha > -1$.

Example 12.14 *The function* $f(x) := x^\alpha e^{-\beta x}$, $x > 0$, *is Lebesgue integrable over* $(0,\infty)$ *for all* $\alpha > -1$ *and* $\beta > 0$.

Indeed: measurability of f follows from its continuity. Using the exponential series, we find for all $N \in \mathbb{N}$ and $x > 0$

$$\frac{(\beta x)^N}{N!} \leqslant \sum_{n=0}^{\infty} \frac{(\beta x)^n}{n!} = e^{\beta x} \implies e^{-\beta x} \leqslant \frac{N!}{\beta^N} x^{-N}.$$

As $e^{-\beta x} \leqslant 1$ for $x > 0$, we obtain the following majorization:

$$f(x) = x^\alpha e^{-\beta x} \leqslant \underbrace{x^\alpha \mathbb{1}_{(0,1)}(x)}_{\substack{\in \mathcal{L}^1(0,1),\text{ if } \alpha > -1 \\ \text{by Example 12.13(i)}}} + \underbrace{\frac{N!}{\beta^N} x^{\alpha-N} \mathbb{1}_{[1,\infty)}(x)}_{\substack{\in \mathcal{L}^1[1,\infty),\text{ if } \alpha - N < -1 \\ \text{by Example 12.13(ii)}}} \in \mathcal{L}^1(0,\infty), \tag{12.12}$$

and $f \in \mathcal{L}^1(0,\infty)$ follows from Theorem 10.3(iv).

Example 12.15 (Euler's gamma function) *The parameter-dependent integral*

$$\Gamma(t) := \int_{(0,\infty)} x^{t-1} e^{-x} \, \lambda(dx), \qquad t > 0 \tag{12.13}$$

is called the gamma function. It has the following properties:

(i) Γ *is continuous;*

(ii) Γ *is arbitrarily often differentiable* (see Problem 12.27(i));

(iii) $t\Gamma(t) = \Gamma(t+1)$, *in particular* $\Gamma(n+1) = n!$ (see Problem 12.27(ii));

(iv) $\ln \Gamma(t)$ *is convex* (see Problem 12.27(iii)).

Example 12.14 shows that the gamma function is well-defined for all $t > 0$. We prove (i) and (ii) first for every interval (a, b) where $0 < a < b < \infty$. Since both continuity and differentiability are local properties, i.e. they need only be checked in a small neighbourhood of each point, (i) and (ii) follow for the half-line if we let $a \to 0$ and $b \to \infty$.

(i) We apply the continuity lemma, Theorem 12.4. Set $u(t, x) := x^{t-1}e^{-x}$. We have already seen in Example 12.14 that $u(t, \cdot) \in \mathcal{L}^1(0, \infty)$ for all $t > 0$; the continuity of $u(\cdot, x)$ is clear and all that remains is to find a uniform (for $t \in (a, b)$) dominating function. An argument similar to (12.12) gives for $N > b + 1$

$$\begin{aligned} x^{t-1}e^{-x} &\leqslant x^{t-1}\mathbb{1}_{(0,1)}(x) + N!x^{t-1-N}\mathbb{1}_{[1,\infty)}(x) \\ &\leqslant x^{a-1}\mathbb{1}_{(0,1)}(x) + N!x^{b-1-N}\mathbb{1}_{[1,\infty)}(x) \\ &\leqslant x^{a-1}\mathbb{1}_{(0,1)}(x) + N!x^{-2}\mathbb{1}_{[1,\infty)}(x). \end{aligned}$$

The expression on the right does not depend on t, and it is integrable according to Example 12.13. (Note that $N = N(b)$ depends on the fixed interval (a, b), but not on t.) This shows that $\Gamma(t) = \int_{(0,\infty)} u(t, x)\lambda(dx)$ is continuous for all $t \in (a, b)$.

(ii) We apply the differentiability lemma, Theorem 12.5. The integrand $u(t, \cdot)$ is integrable, and $u(\cdot, x)$ is differentiable for fixed $x > 0$. In fact,

$$\frac{\partial u(t, x)}{\partial t} = \frac{\partial}{\partial t}x^{t-1}e^{-x} = x^{t-1}e^{-x}\ln x.$$

We still have to show that $\frac{\partial}{\partial t}u(t, x)$ has an integrable majorant uniformly for all $t \in (a, b)$. First we observe that $\ln x \leqslant x$, thus

$$\left|\frac{\partial}{\partial t}u(t, x)\right| \leqslant x^t e^{-x} \leqslant x^b e^{-x} \quad \forall a < t < b,\ x \geqslant 1.$$

For $0 < x < 1$ we use $|\ln x| = \ln(1/x)$, so that

$$\left|\frac{\partial}{\partial t}u(t, x)\right| = x^{t-1}e^{-x}\ln\left(\frac{1}{x}\right) \leqslant x^{a-1}e^{-x}\ln\left(\frac{1}{x}\right) \quad \forall a < t < b,\ 0 < x < 1,$$

and, since $a>0$, we find some $\epsilon>0$ with $a-\epsilon-1>-1$, so that

$$\left|\frac{\partial}{\partial t}u(t,x)\right| \leqslant x^{a-1-\epsilon}e^{-x}\underbrace{x^{\epsilon}\ln\left(\frac{1}{x}\right)}_{\to 0 \text{ as}^2\, x\to 0} \leqslant Cx^{a-1-\epsilon}e^{-x} \quad \forall a<t<b,\ 0<x<1.$$

Combining these calculations, we arrive at

$$\left|\frac{\partial}{\partial t}u(t,x)\right| \leqslant Cx^{a-1-\epsilon}e^{-x}\mathbb{1}_{(0,1)}(x)+x^{b}e^{-x}\mathbb{1}_{[1,\infty)}(x) \quad \forall a<t<b,$$

which is an integrable majorant (by Examples 12.13 and 12.14) that is independent of $t\in(a,b)$. This shows that $\Gamma(t)$ is differentiable on (a,b), with derivative

$$\Gamma'(t)=\int_{(0,\infty)} x^{t-1}e^{-x}\ln x\,\lambda(dx), \qquad t\in(a,b).$$

A similar calculation proves that $\frac{d^n}{dt^n}\Gamma(t)$ exists for every $n\in\mathbb{N}$; see Problem 12.27.

Problems

12.1. Adapt the proof of Theorem 12.2 to show that any sequence $(u_n)_{n\in\mathbb{N}}\subset\mathcal{M}(\mathscr{A})$ with $\lim_{n\to\infty}u_n(x)=u(x)$ and $|u_n|\leqslant g$ for some $g\geqslant 0$ with $g^p\in\mathcal{L}^1(\mu)$ satisfies

$$\lim_{n\to\infty}\int |u_n-u|^p\,d\mu=0.$$

[Hint: mimic the proof of Theorem 12.2 using $|u_n-u|^p\leqslant(|u_n|+|u|)^p\leqslant 2^pg^p$.]

12.2. Give an alternative proof of Theorem 12.2(ii) using the generalized Fatou theorem from Problem 10.7.

12.3. Prove the following result of W. H. Young [60]; among statisticians it is also known as *Pratt's lemma*, see J. W. Pratt [38].

Theorem (Young; Pratt). *Let $(f_k)_k$, $(g_k)_k$ and $(G_k)_k$ be sequences of integrable functions on a measure space $(X,\mathscr{A},\mu)$. If*

(a) $f_k(x)\to f(x)$, $g_k(x)\to g(x)$, $G_k(x)\to G(x)$ *for all* $x\in X$,
(b) $g_k(x)\leqslant f_k(x)\leqslant G_k(x)$ *for all* $k\in\mathbb{N}$ *and all* $x\in X$,
(c) $\int g_k\,d\mu\to\int g\,d\mu$ *and* $\int G_k\,d\mu\to\int G\,d\mu$ *with* $\int g\,d\mu$ and $\int G\,d\mu$ *finite,*

then $\lim_{k\to\infty}\int f_k\,d\mu=\int f\,d\mu$ *and* $\int f\,d\mu$ *is finite.*

Explain why this generalizes Lebesgue's dominated convergence theorem, Theorem 12.2(ii).

[2] To see this, use $\lim_{x\to 0}x^{\epsilon}\ln\left(\frac{1}{x}\right)\overset{x=\exp(-t)}{=}\lim_{t\to\infty}e^{-\epsilon t}t=0$ if $\epsilon>0$.

12.4. Let $(u_n)_{n\in\mathbb{N}}$ be a sequence of integrable functions on $(X, \mathscr{A}, \mu)$. Show that, if $\sum_{n=1}^{\infty} \int |u_n|\, d\mu < \infty$, the series $\sum_{n=1}^{\infty} u_n$ converges a.e. to a real-valued function $u(x)$, and that in this case

$$\int \sum_{n=1}^{\infty} u_n\, d\mu = \sum_{n=1}^{\infty} \int u_n\, d\mu.$$

[Hint: use Corollary 9.9 to see that the series $\sum_n u_n$ converges absolutely for almost all $x \in X$. The rest is then dominated convergence.]

12.5. Let $(u_n)_{n\in\mathbb{N}}$ be a sequence of positive integrable functions on a measure space $(X, \mathscr{A}, \mu)$. Assume that the sequence decreases to 0: $u_1 \geqslant u_2 \geqslant u_3 \geqslant \cdots$ and $u_n \downarrow 0$. Show that $\sum_{n=1}^{\infty} (-1)^n u_n$ converges and is integrable, and that the integral is given by

$$\int \sum_{n=1}^{\infty} (-1)^n u_n\, d\mu = \sum_{n=1}^{\infty} (-1)^n \int u_n\, d\mu.$$

[Hint: mimic the proof of the Leibniz test for alternating series.]

12.6. Find a sequence of integrable functions $(u_n)_{n\in\mathbb{N}}$ with $u_n(x) \to u(x)$ for all x and an integrable function u but such that $\lim_{n\to\infty} \int u_n\, d\mu \neq \int u\, d\mu$. Does this contradict Lebesgue's dominated convergence theorem (Theorem 12.2)?

12.7. Let μ be a finite measure on $([0,\infty), \mathscr{B}[0,\infty))$. Find the limit $\lim\limits_{r\to\infty} \int_{[0,\infty)} e^{-rx} \mu(dx)$.

12.8. Let λ denote Lebesgue measure on $\mathbb{R}^n$.

(i) Let $u \in \mathcal{L}^1(\lambda)$ and $K \subset \mathbb{R}^n$ be a compact (i.e. closed and bounded) set. Show that $\lim_{|x|\to\infty} \int_{K+x} |f| d\lambda = 0$.

(ii) Let u be uniformly continuous and $|f|^p \in \mathcal{L}^1(\lambda)$ for some $p > 0$. Show that $\lim_{|x|\to\infty} f(x) = 0$.

12.9. Let λ denote Lebesgue measure on $\mathbb{R}^n$ and $u \in \mathcal{L}^1(\lambda)$.

(i) For every $\epsilon > 0$ there is a set $B \in \mathscr{B}(\mathbb{R}^n)$, $\lambda(B) < \infty$ with $\sup_B |u| < \infty$ and $\int_{B^c} |u| d\lambda < \epsilon$.

(ii) Use (i) to show that $\lim_{\lambda(B)\to 0} \int_B |u| d\lambda = 0$.

12.10. Let $(X, \mathscr{A}, \mu)$ be a measure space and $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mu)$ be a uniformly convergent sequence.

(i) If $\mu(X) < \infty$, then $\lim_n \int u_n\, d\mu = \int \lim_n u_n\, d\mu$.

(ii) Assume that $u = \lim_n u_n \in \mathcal{L}^1(\mu)$ and $\lim_n \int u_n\, d\mu$ exists. It is true or false that $\lim_n \int u_n\, d\mu = \int u\, d\mu$?

12.11. Let $u \in \mathcal{L}^1(0,1)$ be positive and monotone. Find the limit

$$\lim_{n\to\infty} \int_0^1 u(t^n) dt.$$

12.12. Let $u \in \mathcal{L}^1(0,1)$. Find the limit

$$\lim_{n\to\infty} \int_0^1 t^n f(t) dt.$$

12.13. Show that

$$\int_0^\infty \frac{\sin t}{e^t - 1}\, dt = \sum_{n=1}^{\infty} \frac{1}{n^2+1}.$$

[Hint: Use the geometric series to express $(e^{-t}-1)^{-1}$, observe that $\sin t = \operatorname{Im} e^{it}$ and use Problem 10.9.]

12.14. Let $u:\mathbb{R}\to\mathbb{R}$ be a Borel measurable function and assume that $x\mapsto e^{\lambda x}u(x)$ is integrable for each $\lambda\in\mathbb{R}$. Show that for all $z\in\mathbb{C}$

$$\int_{\mathbb{R}} e^{zx}u(x)dx = \sum_{n=0}^{\infty}\frac{z^n}{n!}\int_{\mathbb{R}} x^n u(x)dx.$$

12.15. Let $(X,\mathscr{A},\mu)$ be a measure space and $u\in\mathcal{L}^1(\mu)$. Show that for every $\epsilon>0$ there is some $\delta>0$ such that

$$A\in\mathscr{A},\ \mu(A)<\delta \implies \left|\int_A u\,d\mu\right| \leqslant \int_A |u|d\mu < \epsilon.$$

12.16. Let λ be one-dimensional Lebesgue measure. Show that for every integrable function u, the *integral function* or *primitive*

$$x\mapsto \int_{(0,x)} u(t)\lambda(dt), \quad x>0,$$

is continuous. What happens if we exchange λ for a general measure μ?

12.17. Consider the functions

$$\text{(i)}\quad u(x)=\frac{1}{x},\quad x\in[1,\infty); \qquad \text{(ii)}\quad v(x)=\frac{1}{x^2},\quad x\in[1,\infty);$$
$$\text{(iii)}\quad w(x)=\frac{1}{\sqrt{x}},\quad x\in(0,1]; \qquad \text{(iv)}\quad y(x)=\frac{1}{x},\quad x\in(0,1];$$

and check whether they are Lebesgue integrable in the regions given – what would happen if we consider $\left[\frac{1}{2},2\right]$ instead?
[Hint: consider first $u_k=u\mathbb{1}_{[1,k]}$, resp., $w_k=w\mathbb{1}_{[1/k,1]}$, etc. and use monotone convergence and the fact that Riemann and Lebesgue integrals coincide if both exist.]

12.18. Show that the function $\mathbb{R}\ni x\mapsto \exp(-x^\alpha)$ is $\lambda^1(dx)$-integrable over the set $[0,\infty)$ for every $\alpha>0$.
[Hint: find integrable majorants u resp. w if $0\leqslant x\leqslant 1$ resp. $1<x<\infty$ and glue them together by $u\mathbb{1}_{[0,1]}+w\mathbb{1}_{(1,\infty)}$ to get an overall integrable upper bound.]

12.19. Show that for every parameter $\alpha>0$ the function

$$x\mapsto \left(\frac{\sin x}{x}\right)^3 e^{-\alpha x}$$

is integrable over $(0,\infty)$ and that the integral is continuous as a function of the parameter.
[Hint: find piecewise integrable majorants like in Problem 12.18; use the continuity lemma.]

12.20. Show that the function

$$G:\mathbb{R}\to\mathbb{R}, \qquad G(x):=\int_{\mathbb{R}\setminus\{0\}}\frac{\sin(tx)}{t(1+t^2)}\,dt$$

is differentiable and find $G(0)$ and $G'(0)$. Use a limit argument, integration by parts for $\int_{(-n,n)}\dots dt$ and the formula $t\frac{\partial}{\partial t}\sin(tx)=x\frac{\partial}{\partial x}\sin(tx)$ to show that

$$xG'(x)=\int_{\mathbb{R}}\frac{2t\sin(tx)}{(1+t^2)^2}\,dt.$$

12.21. Denote by λ one-dimensional Lebesgue measure. Prove that

(i) $\displaystyle\int_{(1,\infty)} e^{-x}\ln(x)\lambda(dx) = \lim_{k\to\infty}\int_{(1,k)}\left(1-\frac{x}{k}\right)^k \ln(x)\lambda(dx)$,

(ii) $\displaystyle\int_{(0,1)} e^{-x}\ln(x)\lambda(dx) = \lim_{k\to\infty}\int_{(0,1)}\left(1-\frac{x}{k}\right)^k \ln(x)\lambda(dx)$.

12.22. Denote by λ Lebesgue measure on $\mathbb{R}$ and set

$$F(t) := \int_{(0,\infty)} e^{-x}\frac{t}{t^2+x^2}\,\lambda(dx), \quad t>0.$$

Show that $F(0+) = \lim_{t\downarrow 0} F(t) = \pi/2$.

Remark. This exercise shows that 'naive' interchange of integration and limit may lead to wrong results: $F(0+)$ is *not* zero!

12.23. Let μ be a measure on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$, let $u:\mathbb{R}\to\mathbb{C}$ be a measurable function (see Problem 10.9) and denote by dx one-dimensional Lebesgue measure. Find conditions on μ and u which guarantee that the so-called *Fourier transforms*

$$\widehat{\mu}(\xi) := \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ix\xi}\,\mu(dx) \quad \text{and} \quad \widehat{u}(\xi) := \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ix\xi}u(x)dx$$

exist resp. are continuous resp. are n times differentiable.

12.24. Let $\phi\in\mathcal{L}^1([0,1],dx)$ and define $f(t) := \int_{[0,1]}|\phi(x)-t|dx$. Show that

(i) f is continuous,

(ii) f is differentiable at $t\in\mathbb{R}$ if, and only if, $\lambda\{\phi=t\}=0$.

12.25. Let $f(t) := \int_0^\infty x^{-2}\sin^2 x\, e^{-tx}\,dx$, $t\geqslant 0$.

(i) Show that f is continuous on $[0,\infty)$ and twice differentiable on $(0,\infty)$.

(ii) Find f'' and work out the limits $\lim_{t\to\infty} f(t)$ and $\lim_{t\to\infty} f'(t)$.

(iii) Use (i) and (ii) to obtain a simple expression for f.

12.26. Show that $\int_0^\infty x^n e^{-x}\,dx = n!$ for all $n\in\mathbb{N}$.

[Hint: Show that $\int_0^\infty e^{-xt}\,dx = 1/t$, $t>0$, and differentiate this identity.]

12.27. **Euler's gamma function.** Show that the function

$$\Gamma(t) := \int_{(0,\infty)} e^{-x}x^{t-1}\,dx, \quad t>0,$$

has the following properties.

(i) It is m-times differentiable with $\Gamma^{(m)}(t) = \int_{(0,\infty)} e^{-x}x^{t-1}(\log x)^m\,dx$.

[Hint: take $t\in(a,b)$ and use induction in m. Note that $|e^{-x}x^{t-1}(\log x)^m| \leqslant x^{m+t-1}e^{-x} \leqslant Mx^{-2}$ for $x\geqslant 1$, and $\leqslant M'x^{\delta-1}$ for $x<1$ and some $\delta>0$ because $\lim_{x\to 0} x^{a-\delta}|\log x|^m = 0$ – use, e.g. the substitution $x=e^{-y}$.]

(ii) It satisfies $\Gamma(t+1) = t\Gamma(t)$.

[Hint: use integration by parts for $\int_{1/n}^n \ldots dt$ and let $n\to\infty$.]

(iii) It is *logarithmically convex*, i.e. $t\mapsto\ln\Gamma(t)$ is convex.

[Hint: check that $\frac{d^2}{dt^2}\ln\Gamma(t)\geqslant 0$. Convexity is discussed in detail in Chapter 13, pp. 123–124.]

12.28. Denote by λ one-dimensional Lebesgue measure on the interval $(0,1)$.

(i) Show that for all $k\in\mathbb{N}_0$ one has

$$\int_{(0,1)} (x\ln x)^k\lambda(dx) = (-1)^k\left(\frac{1}{k+1}\right)^{k+1}\Gamma(k+1).$$

(ii) Use (i) to conclude that

$$\int_{(0,1)} x^{-x} \lambda(dx) = \sum_{k=1}^{\infty} k^{-k}.$$

[Hint: note that $x^{-x} = e^{-x \ln x}$ and use the exponential series.]

12.29. Show that $x \mapsto x^n f(u,x)$, $f(u,x) = e^{ux}/(e^x + 1)$, $0 < u < 1$, is integrable over $\mathbb{R}$ and that $g(u) := \int x^n f(u,x)dx$, $0 < u < 1$, is arbitrarily often differentiable.

12.30. Calculate the following limit:

$$\lim_{n\to\infty} \int_0^1 \frac{1 + nx^2}{(1 + x^2)^n}\, dx.$$

12.31. Let

$$f(t) = \int_0^\infty \arctan\Big(\frac{t}{\sinh x}\Big)\, dx, \quad t > 0,$$

where $\sinh x = \frac{1}{2}(e^x - e^{-x})$.

(i) Show that f is differentiable on $(0, \infty)$, but $f'(0+)$ does not exist.
(ii) Find closed expressions for f', $f(0)$ and $\lim_{t\to\infty} f(t)$.

12.32. Moment generating function. Let X be a positive random variable on the probability space $(\Omega, \mathscr{A}, \mathbb{P})$. The function $\phi_X(t) := \int e^{-tX}\, d\mathbb{P}$ is called *moment generating function*. Show that ϕ_X is m-times differentiable at $t = 0+$ if the absolute mth moment $\int |X|^m\, d\mathbb{P}$ exists. If this is the case, the following formulae hold.

(i) $M_k := \displaystyle\int X^k\, d\mathbb{P} = (-1)^k \frac{d^k}{dt^k}\phi_X(t)\Big|_{t=0+}$ for all $0 \leqslant k \leqslant m$.

(ii) $\phi_X(t) = \displaystyle\sum_{k=0}^{m} \frac{M_k}{k!}(-1)^k t^k + o(t^m).$ ($f(t) = o(t^m)$ means $\lim_{t\to 0} f(t)/t^m = 0$.)

(iii) $\displaystyle\left|\phi_X(t) - \sum_{k=0}^{m-1} \frac{M_k}{k!}(-1)^k t^k\right| \leqslant \frac{|t|^m}{m!} \int |X|^m\, d\mathbb{P}.$

(iv) If $\int |X|^k\, d\mathbb{P} < \infty$ for all $k \in \mathbb{N}$, then

$$\phi_X(t) = \sum_{k=0}^{\infty} \frac{M_k}{k!}(-1)^k t^k$$

for all t within the convergence radius of the series.

12.33. Consider the functions $u(x) = \mathbb{1}_{\mathbb{Q}\cap[0,1]}$ and $v(x) = \mathbb{1}_{\{n^{-1}:n\in\mathbb{N}\}}(x)$. Prove or disprove the following statements.

(i) The function u is 1 on the rationals and 0 otherwise. Thus u is continuous everywhere except for the set $\mathbb{Q} \cap [0, 1]$. Since this is a null set, u is a.e. continuous, hence Riemann integrable by Theorem 12.9.
(ii) The function v is 0 everywhere but for the values $x = 1/n$, $n \in \mathbb{N}$. Thus v is continuous everywhere except for a countable set, i.e. a null set, and v is a.e. continuous, and hence Riemann integrable by Theorem 12.9.
(iii) The functions u and v are Lebesgue integrable and $\int u\, d\lambda = \int v\, d\lambda = 0$.
(iv) The function u is not Riemann integrable.

12.34. Construct a sequence of functions $(u_n)_{n\in\mathbb{N}}$ which are Riemann integrable but converge to a limit $u_n \to u$ which is not Riemann integrable.
[Hint: consider e.g. $u_n = \mathbb{1}_{\{q_1,q_2,\dots,q_n\}}$, where $(q_n)_n$ is an enumeration of $\mathbb{Q}$.]

12.35. Assume that $u\colon [0,\infty) \to \mathbb{R}$ is positive and improperly Riemann integrable. Show that u is also Lebesgue integrable.

12.36. **Fresnel integrals.** Show that the following improper Riemann integrals exist:

$$\int_0^\infty \sin x^2 \, dx \quad \text{and} \quad \int_0^\infty \cos x^2 \, dx.$$

Do they exist as Lebesgue integrals?

Remark. The above integrals have the value $\frac{1}{2}\sqrt{\frac{\pi}{2}}$. This can be proved by methods from complex analysis: Cauchy's theorem or the residue theorem.

12.37. **Frullani's integral.** Let $f\colon (0,\infty) \to \mathbb{R}$ be a continuous function such that $\lim_{x\to 0} f(x) = m$ and $\lim_{x\to\infty} f(x) = M$. Show that the two-sided improper Riemann integral

$$\lim_{\substack{r\to 0\\ s\to\infty}} \int_r^s \frac{f(bx) - f(ax)}{x}\, dx = (M-m)\ln\left(\frac{b}{a}\right)$$

exists for all $a, b > 0$. Does this integral have a meaning as a Lebesgue integral?
[Hint: use the mean value theorem for integrals, Corollary I.12.]

13

The Function Spaces $\mathcal{L}^p$

Throughout this chapter $(X, \mathscr{A}, \mu)$ will be some measure space. We will now discuss functions whose (absolute) pth power is integrable. More precisely, we are interested in the sets

$$\mathcal{L}^p(\mu) := \left\{u : X \to \mathbb{R} : u \in \mathcal{M}(\mathscr{A}),\ \int |u|^p \, d\mu < \infty\right\}, \qquad p \in [1, \infty),$$
$$\mathcal{L}^\infty(\mu) := \left\{u : X \to \mathbb{R} : u \in \mathcal{M}(\mathscr{A}),\ \exists c > 0,\ \mu\{|u| \geqslant c\} = 0\right\}, \quad p = \infty.^1$$

By definition, $u \in \mathcal{L}^\infty(\mu)$ is μ-a.e. bounded by some constant $c > 0$ which may depend on u.

As usual, we suppress μ if the choice of measure is clear, and we write $\mathcal{L}^p(X)$ or $\mathcal{L}^p(\mathscr{A})$ if we want to stress the underlying space or σ-algebra. It is convenient to have the following notation:

$$\begin{aligned} \|u\|_p &:= \left(\int |u(x)|^p \, \mu(dx)\right)^{1/p}, \qquad p \in [1, \infty), \\ \|u\|_\infty &:= \inf\left\{c > 0 : \mu\{|u| \geqslant c\} = 0\right\}, \quad p = \infty.^2 \end{aligned} \tag{13.1}$$

Clearly, $u \in \mathcal{L}^p(\mu)$ if, and only if, $u \in \mathcal{M}(\mathscr{A})$ and $\|u\|_p < \infty$. It is no accident that the notation $\|\cdot\|_p$ resembles the symbol for a *norm*: indeed, we have, because of Theorem 11.2(i),

$$\|u\|_p = 0 \iff u = 0 \quad \text{a.e.}, \tag{13.2}$$

[1] Problem 13.21 shows that $\mathcal{L}^\infty$ is the limit of $\mathcal{L}^p$ as $p \to \infty$.
[2] Problem 13.21 shows that $\|\cdot\|_\infty$ is the limit of $\|\cdot\|_p$ as $p \to \infty$.

and for all $\alpha \in \mathbb{R}$

$$\|\alpha u\|_p = \left(\int |\alpha u|^p \, d\mu \right)^{1/p} = \left(|\alpha|^p \int |u|^p \, d\mu \right)^{1/p} = |\alpha| \|u\|_p, \tag{13.3}$$

$$\|\alpha u\|_\infty = \inf\left\{c : \mu\left\{|u| \geqslant \tfrac{c}{|\alpha|}\right\} = 0\right\} = \inf\{|\alpha| C : \mu\{|u| \geqslant C\} = 0\} = |\alpha| \|u\|_\infty.$$

The triangle inequality for $\|\cdot\|_p$ and deeper results on $\mathcal{L}^p$ depend on the following elementary inequality.

Lemma 13.1 (Young's inequality) *Let $p, q \in (1, \infty)$ be conjugate numbers, i.e. $\frac{1}{p} + \frac{1}{q} = 1$, hence $q = \frac{p}{p-1}$. Then*

$$AB \leqslant \frac{A^p}{p} + \frac{B^q}{q} \tag{13.4}$$

holds for all $A, B \geqslant 0$; equality occurs if, and only if, $B = A^{p-1}$.

Proof There are various different methods to prove (13.4) but probably the most intuitive one is through Fig. 13.1. If we add the areas below the graph of the function $\xi \mapsto \xi^{p-1}$ and below the graph of $\eta \mapsto \eta^{q-1}$ (shaded, turn Fig. 13.1 counterclockwise by 90°), we see that their combined area is greater than the area of the rectangle with sides $\overrightarrow{OA}$ and $\overrightarrow{OB}$,

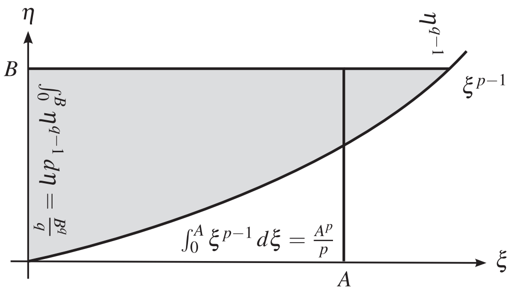


Fig. 13.1. For proof of Young's inequality.

$$\frac{A^p}{p} + \frac{B^q}{q} = \int_0^A \xi^{p-1} \, d\xi + \int_0^B \eta^{q-1} \, d\eta \geqslant AB.$$

Equality holds if, and only if, the shaded area to the right of A vanishes, i.e. if $B = A^{p-1}$. □

We can now prove the following fundamental inequality.

Theorem 13.2 (Hölder's inequality) *Assume that $u \in \mathcal{L}^p(\mu)$ and $v \in \mathcal{L}^q(\mu)$, where $p, q \in [1, \infty]$ are conjugate numbers: $\frac{1}{p} + \frac{1}{q} = 1$.[3] Then $uv \in \mathcal{L}^1(\mu)$, and the following inequality holds:*

$$\left| \int uv \, d\mu \right| \leqslant \int |uv| d\mu \leqslant \|u\|_p \cdot \|v\|_q. \tag{13.5}$$

[3] Setting $1/\infty = 0$, we see that $p = 1$ and $q = \infty$ are also conjugate numbers.

Equality occurs in the second inequality if, and only if, $|u(x)|^p/\|u\|_p^p = |v(x)|^q/\|v\|_q^q$ *a.e.* $(1 < p, q < \infty)$, *resp.,* $v|_{\{u\neq 0\}} \equiv \|v\|_\infty$ *a.e.* $(p = 1, q = \infty)$.

Proof The first inequality of (13.5) follows directly from Theorem 10.4(v). For the other inequality we distinguish between two cases.

Case 1. $p \in (1, \infty)$ *and* $q = p/(p-1) \in (1, \infty)$. Using (13.4) with

$$A := \frac{|u(x)|}{\|u\|_p} \quad \text{and} \quad B := \frac{|v(x)|}{\|v\|_q}$$

we get

$$\frac{|u(x)v(x)|}{\|u\|_p\|v\|_q} \leqslant \frac{|u(x)|^p}{p\|u\|_p^p} + \frac{|v(x)|^q}{q\|v\|_q^q}.$$

Integrating both sides of this inequality over x yields

$$\frac{\int |u(x)v(x)|\mu(dx)}{\|u\|_p\|v\|_q} \leqslant \frac{\|u\|_p^p}{p\|u\|_p^p} + \frac{\|v\|_q^q}{q\|v\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1.$$

Equality can happen only if we have equality in (13.4). Because of our choice of A and B, this means that $|v(x)|/\|v\|_q = \big(|u(x)|/\|u\|_p|\big)^{p-1}$ a.e. Raising both sides to the qth power gives $|v(x)|^q/\|v\|_q^q = |u(x)|^p/\|u\|_p^p$ since $(p-1)q = p$.

Case 2. $p = 1$ *and* $q = \infty$. As $\|v\|_\infty \geqslant |v|$ a.e., we have

$$\int |uv|d\mu \leqslant \int |u|\|v\|_\infty \, d\mu = \|u\|_1\|v\|_\infty. \qquad \square$$

Hölder's inequality with $p = q = 2$ is usually called the *Cauchy–Schwarz inequality*.

Corollary 13.3 (Cauchy–Schwarz inequality) *Let* $u, v \in \mathcal{L}^2(\mu)$. *Then* $uv \in \mathcal{L}^1(\mu)$ *and*

$$\int |uv|d\mu \leqslant \|u\|_2 \cdot \|v\|_2. \tag{13.6}$$

Equality occurs if, and only if, $|u(x)|^2/\|u\|_2^2 = |v(x)|^2/\|v\|_2^2$ *a.e.*

Another consequence of Hölder's inequality is the Minkowski or triangle inequality for $\|\cdot\|_p$.

Corollary 13.4 (Minkowski's inequality) *Let* $u, v \in \mathcal{L}^p(\mu)$, $p \in [1, \infty]$. *Then the sum* $u + v \in \mathcal{L}^p(\mu)$ *and*

$$\|u + v\|_p \leqslant \|u\|_p + \|v\|_p. \tag{13.7}$$

Proof *Case 1.* $p \in [1, \infty)$. Since

$$|u + v|^p \leqslant (|u| + |v|)^p \leqslant 2^p \max\{|u|^p, |v|^p\} \leqslant 2^p(|u|^p + |v|^p),$$

we get that $|u+v|^p \in \mathcal{L}^1(\mu)$ or $u+v \in \mathcal{L}^p(\mu)$. Now

$$\begin{aligned}\int |u+v|^p \, d\mu &= \int |u+v| \cdot |u+v|^{p-1} \, d\mu \\ &\leqslant \int |u| \cdot |u+v|^{p-1} \, d\mu + \int |v| \cdot |u+v|^{p-1} \, d\mu \\ &\qquad \text{(if } p=1 \text{ the proof stops here } \quad \Box\text{)} \\ &\overset{13.2}{\leqslant} \|u\|_p \cdot \left\| |u+v|^{p-1} \right\|_q + \|v\|_p \cdot \left\| |u+v|^{p-1} \right\|_q.\end{aligned}$$

Dividing both sides by $\| |u+v|^{p-1} \|_q$ proves our claim since

$$\left\| |u+v|^{p-1} \right\|_q = \left(\int |u+v|^{(p-1)q} \, d\mu \right)^{1/q} = \left(\int |u+v|^p \, d\mu \right)^{1-1/p},$$

where we also use that $q = p/(p-1)$.

Case 2. $p = \infty$. If $|u| < c$ and $|v| < C$ a.e., then $|u+v| \leqslant |u| + |v| < c + C$ a.e. Thus, $\|u+v\|_\infty \leqslant c + C$ and the claim follows by minimizing over all admissible constants c and C. $\Box$

Remark 13.5 (i) Formulae (13.3) and (13.7) imply

$$u, v \in \mathcal{L}^p(\mu) \implies \alpha u + \beta v \in \mathcal{L}^p(\mu) \quad \forall \alpha, \beta \in \mathbb{R},$$

which shows that $\mathcal{L}^p(\mu)$ is a vector space.

(ii) Formulae (13.2), (13.3) and (13.7) show that $\|\cdot\|_p$ is a *semi-norm* for $\mathcal{L}^p(\mu)$: the definiteness of a norm is not fulfilled since

$$\|u\|_p = 0 \quad \text{implies only that} \quad u(x) = 0 \text{ for almost every } x$$

but not for all x. There is a standard recipe to fix this. Since $\mathcal{L}^p$-functions can be altered on null sets without affecting their integration behaviour, we introduce the following *equivalence relation*: we call $u, v \in \mathcal{L}^p(\mu)$ *equivalent* if they differ on at most a μ-null set, i.e.

$$u \sim v \iff \{u \neq v\} \in \mathcal{N}_\mu.$$

The *quotient space* $L^p(\mu) := \mathcal{L}^p(\mu)/_\sim$ consists of all equivalence classes of $\mathcal{L}^p$-functions. If $[u]_p \in L^p(\mu)$ denotes the equivalence class induced by the function $u \in \mathcal{L}^p(\mu)$, it is not hard to see that

$$[\alpha u + \beta v]_p = \alpha [u]_p + \beta [v]_p \quad \text{and} \quad [uv]_1 = [u]_p [v]_q$$

hold, turning $L^p(\mu)$ into a bona fide vector space with the canonical *norm*

$$\|[u]_p\|_p := \inf\{\|w\|_p : w \in \mathcal{L}^p, \ w \sim u\}$$

for quotient spaces. Fortunately, $\|[u]_p\|_p = \|u\|_p$ and later on we will often follow the usual abuse of notation and identify $[u]$ with u.

(iii) All results of this chapter are still valid for $\overline{\mathbb{R}}$-valued functions. Indeed, if $f \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A})$ and $\int |f|^p \, d\mu < \infty$, then

$$\begin{aligned}\mu\{|f| = \infty\} = \mu\{|f|^p = \infty\} = \mu\Bigg(\bigcap_{n\in\mathbb{N}} \{|f|^p > n\}\Bigg) \\ &\overset{4.3}{\underset{(\text{vii})}{=}} \lim_{n\to\infty} \mu\{|f|^p > n\} \\ &\overset{11.5}{\leqslant} \lim_{n\to\infty} \frac{1}{n^p} \int |f|^p \, d\mu = 0,\end{aligned}$$

by the Markov inequality. This means, however, that f is a.e. $\mathbb{R}$-valued, so sums and products of such functions are always defined outside a μ-null set. In particular, there is no need to distinguish between the classes $L^p(\mu)$ $(= L^p_{\mathbb{R}}(\mu))$ and $L^p_{\overline{\mathbb{R}}}(\mu)$.

Convergence in $\mathcal{L}^p$ and completeness

We will need the concept of *convergence* of a sequence in the space $\mathcal{L}^p(\mu)$. A sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$ is said to be *convergent* in the space $\mathcal{L}^p(\mu)$ with *limit* $\mathcal{L}^p\text{-}\lim_{n\to\infty} u_n = u$ if, and only if,

$$\lim_{n\to\infty} \|u_n - u\|_p = 0.$$

Remember, however, that $\mathcal{L}^p$-limits are only almost everywhere unique. If u, w are both $\mathcal{L}^p$-limits of the same sequence $(u_n)_{n\in\mathbb{N}}$, we have

$$\|u - w\|_p \overset{13.4}{\leqslant} \lim_{n\to\infty} \big(\|u - u_n\|_p + \|u_n - w\|_p\big) = 0,$$

implying only $u = w$ almost everywhere.

We call $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$ an ($\mathcal{L}^p$-)*Cauchy sequence* if

$$\forall \epsilon > 0 \quad \exists N_\epsilon \in \mathbb{N} \quad \forall k, n \geqslant N_\epsilon : \|u_n - u_k\|_p < \epsilon.$$

Note that these definitions reduce convergence in $\mathcal{L}^p(\mu)$ to convergence questions of the semi-norm $\|\cdot\|_p$ in $\mathbb{R}^+$. This means that, apart from uniqueness, many formal properties of limits in $\mathbb{R}$ carry over to $\mathcal{L}^p$ – most of them even with the same proofs!

Caution Pointwise convergence of a sequence $u_n(x) \to u(x)$ of $\mathcal{L}^p$-functions $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p$ does not guarantee convergence in $\mathcal{L}^p$ – but in view of Lebesgue's dominated convergence theorem (Theorem 12.2), the additional condition that

$$|u_n(x)| \leqslant w(x) \quad \text{for some function } w \in \mathcal{L}^p(\mu)$$

is sufficient since $|u_n - u|^p \leqslant (|u_n| + |u|)^p \leqslant 2^p w^p$ and $|u_n(x) - u(x)| \to 0$. [✍]

Clearly, a convergent sequence $(u_n)_{n\in\mathbb{N}}$ is also a Cauchy sequence,

$$\|u_n - u_k\|_p \leqslant \|u_n - u\|_p + \|u - u_k\|_p < 2\epsilon \quad \forall k, n \geqslant N_\epsilon;$$

the converse of this assertion is also true, but much more difficult to prove. We start with a simple observation.

Lemma 13.6 *For any sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$, $p \in [1, \infty)$, of positive functions $u_n \geqslant 0$ we have*

$$\left\| \sum_{n=1}^{\infty} u_n \right\|_p \leqslant \sum_{n=1}^{\infty} \|u_n\|_p. \tag{13.8}$$

Proof Repeated applications of Minkowski's inequality (13.7) show that

$$\left\| \sum_{n=1}^{N} u_n \right\|_p \leqslant \sum_{n=1}^{N} \|u_n\|_p \leqslant \sum_{n=1}^{\infty} \|u_n\|_p,$$

and, since the right-hand side is independent of N, the inequality remains valid even if we pass to the $\sup_{N\in\mathbb{N}}$ on the left. By Beppo Levi's theorem, Theorem 9.6, we find

$$\begin{aligned} \sup_{N\in\mathbb{N}} \left\| \sum_{n=1}^{N} u_n \right\|_p^p &= \sup_{N\in\mathbb{N}} \int \left(\sum_{n=1}^{N} u_n \right)^p d\mu \\ &= \int \left(\sup_{N\in\mathbb{N}} \sum_{n=1}^{N} u_n \right)^p d\mu = \int \left(\sum_{n=1}^{\infty} u_n \right)^p d\mu. \end{aligned} \qquad \square$$

The completeness of $\mathcal{L}^p(\mu)$ was proved by E. Fischer (for $p = 2$) and F. Riesz (for $1 \leqslant p < \infty$).

Theorem 13.7 (Riesz–Fischer) *The spaces $\mathcal{L}^p(\mu)$, $p \in [1, \infty]$, are complete, i.e. every Cauchy sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$ converges to some limit $u \in \mathcal{L}^p(\mu)$.*

Proof Assume first that $p \in [1, \infty)$. The main difficulty here is to identify the limit u. By the definition of a Cauchy sequence we find numbers

$$1 < n(1) < n(2) < \cdots < n(k) < \ldots$$

such that

$$\|u_{n(k+1)} - u_{n(k)}\|_p < 2^{-k}, \quad k \in \mathbb{N}.$$

To find u, we turn the sequence into a series by

$$u_{n(k+1)} = \sum_{i=0}^{k} \big(u_{n(i+1)} - u_{n(i)}\big), \qquad u_{n(0)} := 0, \tag{13.9}$$

and the limit as $k \to \infty$ would formally be $u := \sum_{i=0}^{\infty} \big(u_{n(i+1)} - u_{n(i)}\big)$ – if we can make sense of this infinite sum. Since

$$\begin{aligned} \left\| \sum_{i=0}^{\infty} |u_{n(i+1)} - u_{n(i)}| \right\|_p &\overset{(13.8)}{\leqslant} \sum_{i=0}^{\infty} \|u_{n(i+1)} - u_{n(i)}\|_p \\ &\leqslant \|u_{n(1)}\|_p + \sum_{i=1}^{\infty} \frac{1}{2^i}, \end{aligned} \tag{13.10}$$

we conclude with Corollary 11.6 that $\big(\sum_{i=0}^{\infty} |u_{n(i+1)} - u_{n(i)}|\big)^p < \infty$ a.e., so that $u = \sum_{i=0}^{\infty} \big(u_{n(i+1)} - u_{n(i)}\big)$ is a.e. (absolutely) convergent.

Let us show that $u = \mathcal{L}^p\text{-}\lim_{k\to\infty} u_{n(k)}$. For this, observe that by the (ordinary) triangle inequality and (13.10),

$$\begin{aligned} \|u - u_{n(k)}\|_p = \left\| \sum_{i=k}^{\infty} \big(u_{n(i+1)} - u_{n(i)}\big) \right\|_p &\overset{\text{def}}{=} \left\| \left| \sum_{i=k}^{\infty} \big(u_{n(i+1)} - u_{n(i)}\big) \right| \right\|_p \\ &\leqslant \left\| \sum_{i=k}^{\infty} |u_{n(i+1)} - u_{n(i)}| \right\|_p \\ &\overset{(13.8)}{\leqslant} \sum_{i=k}^{\infty} \|u_{n(i+1)} - u_{n(i)}\|_p \xrightarrow[k\to\infty]{} 0. \end{aligned}$$

Finally, using that $(u_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, we get, for all $\epsilon > 0$ and suitable $N_\epsilon \in \mathbb{N}$,

$$\begin{aligned} \|u - u_n\|_p &\leqslant \|u - u_{n(k)}\|_p + \|u_{n(k)} - u_n\|_p \\ &\leqslant \|u - u_{n(k)}\|_p + \epsilon \qquad\qquad \forall n, n(k) \geqslant N_\epsilon. \end{aligned}$$

Letting $k \to \infty$ shows that $\|u - u_n\|_p \leqslant \epsilon$ if $n \geqslant N_\epsilon$.

Now we consider the case $p = \infty$. The completeness of $\mathcal{L}^\infty(\mu)$ is much easier to prove. If $(u_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{L}^\infty(\mu)$, we set

$$A_{k,\ell} := \{|u_k| > \|u_k\|_\infty\} \cup \{|u_k - u_\ell| > \|u_k - u_\ell\|_\infty\}, \quad A := \bigcup_{k,\ell\in\mathbb{N}} A_{k,\ell}.$$

By definition, $\mu(A_{k,\ell}) = 0$ and $\mu(A) = 0$, so that $\|u_n \mathbb{1}_A\|_\infty = 0$ for all $n \in \mathbb{N}$. On the set A^c, however, $(u_n)_{n\in\mathbb{N}}$ converges uniformly to a bounded function u, i.e. $u\mathbb{1}_{A^c} \in \mathcal{L}^\infty(\mu)$ as well as $\|(u_n - u)\mathbb{1}_{A^c}\|_\infty \to 0$. $\square$

The proof of Theorem 13.7 shows even a weak form of pointwise convergence:

Corollary 13.8 *Let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$, $p \in [1,\infty)$ with $\mathcal{L}^p$-$\lim_{n\to\infty} u_n = u$. Then there exists a subsequence $(u_{n(k)})_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} u_{n(k)}(x) = u(x)$ holds for almost every $x \in X$.*

Proof Since $(u_n)_{n\in\mathbb{N}}$ converges in $\mathcal{L}^p(\mu)$, it is also an $\mathcal{L}^p$-Cauchy sequence and the claim follows from (the remarks following) formula (13.10). □

As we have already remarked, pointwise convergence alone does not guarantee convergence in $\mathcal{L}^p$, not even of a subsequence, see Problem 13.8. Let us repeat the following sufficient criterion, which we have already proved on page 119.

Theorem 13.9 *Let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$, $p \in [1,\infty)$, be a sequence of functions such that $|u_n| \leqslant w$ for all $n \in \mathbb{N}$ and some $w \in \mathcal{L}^p(\mu)$. If $u(x) = \lim_{n\to\infty} u_n(x)$ exists for (almost) every $x \in X$, then*

$$u \in \mathcal{L}^p(\mu) \quad \textit{and} \quad \lim_{n\to\infty} \|u - u_n\|_p = 0.$$

Of a different flavour is the next result which is sometimes called Riesz's convergence theorem.

Theorem 13.10 (F. Riesz) *Let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$, $p \in [1,\infty)$, be a sequence such that $\lim_{n\to\infty} u_n(x) = u(x)$ for almost every $x \in X$ and some $u \in \mathcal{L}^p(\mu)$. Then*

$$\lim_{n\to\infty} \|u_n - u\|_p = 0 \iff \lim_{n\to\infty} \|u_n\|_p = \|u\|_p. \tag{13.11}$$

Proof The direction '$\Rightarrow$' in (13.11) follows from the lower triangle inequality[4] $|\|u_n\|_p - \|u\|_p| \leqslant \|u_n - u\|_p$ for $\|\cdot\|_p$.

For '$\Leftarrow$' we observe that

$$|u_n - u|^p \leqslant (|u_n| + |u|)^p \leqslant 2^p \max\{|u_n|^p, |u|^p\} \leqslant 2^p(|u_n|^p + |u|^p),$$

and we can apply Fatou's lemma (Theorem 9.11) to the sequence

$$2^p(|u_n|^p + |u|^p) - |u_n - u|^p \geqslant 0$$

[4] In the same way as $||a| - |b|| \leqslant |a - b|$ follows from $|a + b| \leqslant |a| + |b|$, $a, b \in \mathbb{R}$.

to get

$$
\begin{aligned}
2^{p+1} \int |u|^p \, d\mu &= \int \liminf_{n\to\infty} \left(2^p(|u_n|^p + |u|^p) - |u_n - u|^p\right) d\mu \\
&\leqslant \liminf_{n\to\infty} \left(2^p \int |u_n|^p \, d\mu + 2^p \int |u|^p \, d\mu - \int |u_n - u|^p \, d\mu\right) \\
&= 2^{p+1} \int |u|^p \, d\mu - \limsup_{n\to\infty} \int |u_n - u|^p \, d\mu,
\end{aligned}
$$

where we use that $\lim_{n\to\infty} \int |u_n|^p \, d\mu = \int |u|^p \, d\mu$. This shows that

$$
\limsup_{n\to\infty} \int |u_n - u|^p \, d\mu = 0, \quad \text{hence} \quad \lim_{n\to\infty} \int |u_n - u|^p \, d\mu = 0. \qquad \square
$$

Caution Theorem 13.10 does not hold for $p = \infty$. This can be seen on $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$ using $u_n(x) := e^{-|x/n|} \to \mathbb{1}_{\mathbb{R}}(x)$.

With a special choice of $(X, \mathscr{A}, \mu)$ we see that integrals generalize infinite series.

Example 13.11 Consider the counting measure $\mu = \sum_{n=1}^{\infty} \delta_n$, see Example 4.5(iii), on the measurable space $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$. From Examples 9.10(ii) and 10.6(ii) we know that a function $u : \mathbb{N} \to \mathbb{R}$ is μ-integrable if, and only if,

$$
\sum_{n=1}^{\infty} |u(n)| < \infty, \quad \text{in which case} \quad \int_{\mathbb{N}} u \, d\mu = \sum_{n=1}^{\infty} u(n).
$$

In a similar way one can show that $v \in \mathcal{L}^p(\mu)$ if, and only if, $\sum_{n=1}^{\infty} |v(n)|^p < \infty$. Functions $u : \mathbb{N} \to \mathbb{R}$ are determined by their values $(u(1), u(2), u(3), \dots)$, and every sequence $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ defines a function u by $u(n) := a_n$. This means that we can identify the function u with the sequence $(u(n))_{n\in\mathbb{N}}$ of real numbers. Thus

$$
\begin{aligned}
\mathcal{L}^p(\mu) &= \left\{ u : \mathbb{N} \to \mathbb{R} : \sum_{n=1}^{\infty} |u(n)|^p < \infty \right\} \\
&= \left\{ (a_n)_{n\in\mathbb{N}} \subset \mathbb{R} : \sum_{n=1}^{\infty} |a_n|^p < \infty \right\} =: \ell^p(\mathbb{N}),
\end{aligned}
$$

the latter being a so-called *sequence space*.

The space $\ell^\infty(\mathbb{N}) := \{(a_n)_{n\in\mathbb{N}} \subset \mathbb{R} : \sup_n |a_n| < \infty\}$ comprises all bounded sequences. Hölder's inequality now reads

$$\sum_{n=1}^\infty |a_n b_n| \leqslant \begin{cases} \left(\sum\limits_{n=1}^\infty |a_n|^p\right)^{1/p} \left(\sum\limits_{n=1}^\infty |b_n|^q\right)^{1/q} & \text{if } \dfrac{1}{p} + \dfrac{1}{q} = 1,\ p, q \in (1, \infty), \\ \sup\limits_{n\in\mathbb{N}} |b_n| \sum\limits_{n=1}^\infty |a_n| & \text{if } p = 1,\ q = \infty, \end{cases}$$

while Minkowski's inequality now reads

$$\begin{aligned} \left(\sum_{n=1}^\infty |a_n \pm b_n|^p\right)^{1/p} &\leqslant \left(\sum_{n=1}^\infty |a_n|^p\right)^{1/p} + \left(\sum_{n=1}^\infty |b_n|^p\right)^{1/p} && \text{if } p \in [1, \infty), \\ \sup_{n\in\mathbb{N}} |a_n \pm b_n| &\leqslant \sup_{n\in\mathbb{N}} |a_n| + \sup_{n\in\mathbb{N}} |b_n| && \text{if } p = \infty. \end{aligned}$$

Convexity and Jensen's Inequality

Recall that a function $V : [a, b] \to \mathbb{R}$ on an interval $[a, b] \subset \overline{\mathbb{R}}$ is *convex* if

$$V(tx + (1-t)y) \leqslant tV(x) + (1-t)V(y), \qquad 0 < t < 1, \tag{13.12}$$

holds for all $x, y \in [a, b]$. Geometrically this means that the graph of a convex function V between the points $(x, V(x))$ and $(y, V(y))$ lies below the chord linking $(x, V(x))$ and $(y, V(y))$, see Fig. 13.2. Convex functions have nice properties: they are continuous in (a, b) and, if V' exists, it is increasing. If V is twice differentiable, convexity is equivalent to $V'' \geqslant 0$ $[V'' \leqslant 0]$. Further details and proofs can be found in Boas [6].

A function $\Lambda : [a, b] \to \mathbb{R}$ is said to be *concave* if $V(x) := -\Lambda(x)$ is convex.

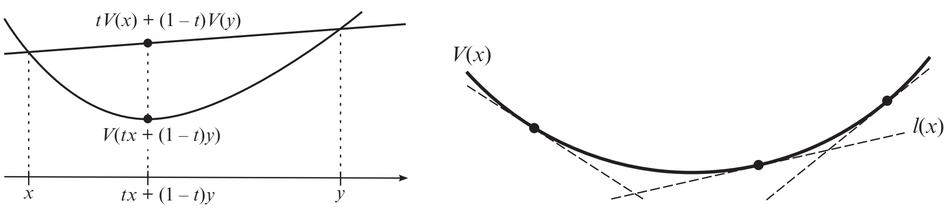


Fig. 13.2. *Left:* A convex function looks like a 'smile'. *Right:* A convex function is the upper envelope of all affine linear functions below its graph.

Lemma 13.12 *A convex function $V : [a, b] \to \mathbb{R}$ is continuous at each $y \in (a, b)$ and it admits finite right-hand derivatives $V_+{}'$ such that*

$$V_+{}'(x)(y - x) + V(x) \leqslant V(y) \quad \forall x, y \in (a, b). \tag{13.13}$$

In particular, $V|_{(a,b)}$ is the upper envelope of all affine linear functions below its graph

$$V(x) = \sup\left\{\ell(x) : \ell(y) = \alpha y + \beta \leqslant V(y) \quad \forall y \in (a,b)\right\}. \tag{13.14}$$

Proof Since the graph of a convex function looks like a smile, the last statement of the lemma is intuitively clear. A rigorous argument uses (13.13), which says that V admits at every point a tangent below its graph.

Pick numbers $a < x < \xi < y < b$ and write $\xi = tx + (1-t)y$ with $t \in (0,1)$, $t = (y-\xi)/(y-x)$. By convexity,

$$\begin{aligned}\frac{V(\xi) - V(x)}{\xi - x} &\leqslant \frac{tV(x) + (1-t)V(y) - V(x)}{\xi - x}\\ &= \frac{(1-t)(V(y) - V(x))}{\xi - x} = \frac{V(y) - V(x)}{y - x},\end{aligned}$$

This means that $\xi \mapsto (V(\xi) - V(x))/(\xi - x)$ is increasing, see Fig. 13.2. Letting $\xi \downarrow x$ proves right-continuity at x, letting $x \uparrow \xi$ gives left-continuity at ξ and, since x, ξ are arbitrary, we get continuity. Because of monotonicity, we find

$$V'_+(x) = \lim_{\xi \downarrow x} \frac{V(\xi) - V(x)}{\xi - x} \leqslant \frac{V(y) - V(x)}{y - x}$$

which entails (13.13). Since this inequality describes supporting affine linear functions below the graph of V, which touch the graph in x, it is clear that V is the upper envelope of affine linear functions. □

If $V : [0,\infty) \to \mathbb{R}$ is convex, then the extension $V : [0,\infty] \to (-\infty,\infty]$, where $V(\infty) := +\infty$ is again convex [✍].

Theorem 13.13 (Jensen's inequality) *Let $(X, \mathscr{A}, \mu)$ be a measure space and μ a probability measure.*

(i) *Let $V : [0,\infty) \to [0,\infty)$ be a convex function and extend it to $[0,\infty]$ as above:*

$$V\left(\int u\,d\mu\right) \leqslant \int V(u)d\mu \quad \forall u \in \mathcal{M}(\mathscr{A}),\ u \geqslant 0. \tag{13.15}$$

In particular, if $V(u) \in \mathcal{L}^1(\mu), u \geqslant 0$, then $u \in \mathcal{L}^1(\mu)$.

(ii) *Let $\Lambda : [0,\infty) \to [0,\infty)$ be a concave function:*

$$\int \Lambda(u)d\mu \leqslant \Lambda\left(\int u\,d\mu\right) \quad \forall u \in \mathcal{L}^1(\mu),\ u \geqslant 0. \tag{13.16}$$

In particular, $u \in \mathcal{L}^1(\mu), u \geqslant 0$, implies that $\Lambda(u) \in \mathcal{L}^1(\mu)$.

Proof (i) Observe that

$$V(u(x)) = \mathbb{1}_{\{u=0\}}(x)V(0) + \mathbb{1}_{\{\infty>u>0\}}(x)V(u(x)) + \infty\mathbb{1}_{\{u=\infty\}}(x).$$

Since $V|_{(0,\infty)}$ is continuous, see Lemma 13.12, the function V and its extension are measurable. Therefore, all integrals appearing in (13.15) are defined.

If $\int V(u)d\mu = \infty$, there is nothing to show. Let $\int V(u)d\mu < \infty$. Since μ is a probability measure, we find for any $\ell(x) := \alpha x + \beta \leqslant V(x)$ that

$$\ell\left(\int u\,d\mu\right) = \alpha\int u\,d\mu + \beta = \int(\alpha u + \beta)d\mu \leqslant \int V(u)d\mu.$$

We can now take the sup of all affine linear functions $\ell \leqslant V$, and (13.15) follows for positive convex functions.

(ii) Since $-\Lambda$ is convex, we see as in part (i) that $\Lambda(u)$ is measurable. For any $\ell(x) := \alpha x + \beta \geqslant \Lambda(x)$ we get

$$\int \Lambda(u)d\mu \leqslant \int(\alpha u + \beta)d\mu = \alpha\int u\,d\mu + \beta = \ell\left(\int u\,d\mu\right).$$

The inequality follows from Lemma 13.12 if we pass to the inf over all affine linear functions satisfying $\ell \geqslant \Lambda$. □

The following example shows some of the most common applications and extensions of Jensen's inequality.

Example 13.14 Let $(X, \mathscr{A}, \nu)$ be a measure space and $w \in \mathcal{L}^1(\nu)$, $w \geqslant 0$.

(i) Taking $\mu := w\nu / \int w\,d\nu$ in Theorem 13.13(i), we get for convex functions $V: [0,\infty) \to [0,\infty)$

$$V\left(\frac{\int uw\,d\nu}{\int w\,d\nu}\right) \leqslant \frac{\int V(u)w\,d\nu}{\int w\,d\nu}, \quad \forall u \in \mathcal{M}(\mathscr{A}),\ u \geqslant 0.$$

(ii) Taking $\mu := w\nu / \int w\,d\nu$ in Theorem 13.13(ii), we get for concave functions $\Lambda: [0,\infty) \to [0,\infty)$

$$\frac{\int \Lambda(u)w\,d\nu}{\int w\,d\nu} \leqslant \Lambda\left(\frac{\int uw\,d\nu}{\int w\,d\nu}\right) \quad \forall u \in \mathcal{L}^1(w\nu),\ u \geqslant 0.$$

(iii) Let μ be a probability measure and $p \in [1,\infty)$. Then $\left|\int u\,d\mu\right|^p \leqslant \int |u|^p\,d\mu$ for all $u \in \mathcal{M}(\mathscr{A})$, $u \geqslant 0$, or $u \in \mathcal{L}^1(\mu)$:

$$\left|\int u\,d\mu\right|^p \underset{\text{ineq.}}{\overset{\text{triangle}}{\leqslant}} \left(\int |u|d\mu\right)^p \underset{w=1}{\overset{\text{Jensen}}{\leqslant}} \int |u|^p\,d\mu.$$

In particular, $u \in \mathcal{L}^p(\mu)$, $p > 1 \implies u \in \mathcal{L}^1(\mu)$.

(iv) Let μ be a probability measure and $p \in (0,1)$. Then $\int |u|^p \, d\mu \leqslant \left(\int |u| \, d\mu\right)^p$ for all $u \in \mathcal{L}^1(\mu)$.

(v) If $V : \mathbb{R} \to \mathbb{R}$ is a convex function and u, $V(u) \in \mathcal{L}^1(\mu)$, then the same proof as for Theorem 13.13 shows that

$$V\left(\int u \, d\mu\right) \leqslant \int V(u) d\mu.$$

(vi) If $\Lambda : \mathbb{R} \to \mathbb{R}$ is a concave function and $u, \Lambda(u) \in \mathcal{L}^1(\mu)$, then the same proof as for Theorem 13.13 shows that

$$\int \Lambda(u) d\mu \leqslant \Lambda\left(\int u \, d\mu\right).$$

✻Convexity inequalities in $\mathbb{R}^2_+$

Denote $\mathbb{R}^2_+ = [0,\infty) \times [0,\infty)$. A function $F : \mathbb{R}^2_+ \to \mathbb{R}$ is said to be *convex* if

$$F[t(x,\xi) + (1-t)(y,\eta)] \leqslant tF[x,\xi] + (1-t)F[y,\eta], \quad 0 < t < 1,$$

for all $(x,\xi), (y,\eta) \in \mathbb{R}^2_+$. The function F is *concave* if $-F$ is convex, and F is *positive homogeneous* if

$$F[t(x,\xi)] = tF[x,\xi] \quad \forall (x,\xi) \in \mathbb{R}^2_+, \ t > 0.$$

Lemma 13.15 *Let $F : \mathbb{R}^2_+ \to \mathbb{R}$ be a continuous and positive homogeneous function. The following assertions are equivalent.*

(i) *F is convex.*
(ii) *$f(x) := F[x, 1-x]$ is convex on $[0,1]$.*
(iii) *$f(x) := F[x, 1]$ is convex on $[0,\infty)$.*
(iv) *There exists a set $C \subset \mathbb{R}^2$ such that $F[x,\xi] = \sup_{(\alpha,\beta)\in C} (\alpha x + \beta \xi)$.*

Proof (i) $\Rightarrow$ (ii) and (iii). Fix $x, y \in [0,1]$ and $t \in (0,1)$. By assumption

$$\begin{aligned} f(tx + (1-t)y) &= F[tx + (1-t)y, 1 - tx - (1-t)y] \\ &= F[t(x, 1-x) + (1-t)(y, 1-y)] \\ &\overset{\text{(i)}}{\leqslant} tF[x, 1-x] + (1-t)F[y, 1-y] \\ &= tf(x) + (1-t)f(y). \end{aligned}$$

Assertion (iii) follows in a similar way.

(ii) $\Rightarrow$ (iv). Since $f(x) := F[x, 1-x]$ is convex on $[0,1]$, Lemma 13.12 shows that f is the upper envelope of affine linear functions $\ell(x) = \alpha x + \beta$. Thus, there is some $\Gamma \subset \mathbb{R}^2$ such that

$$F[x, 1-x] = f(x) = \sup_{\ell \leqslant f} \ell(x) = \sup_{(\alpha,\beta)\in\Gamma} (\alpha x + \beta).$$

Since F is positively homogeneous, we get

$$\begin{aligned} F[x,\xi] = (x+\xi)F\left[\tfrac{x}{x+\xi}, \tfrac{\xi}{x+\xi}\right] &= (x+\xi) \sup_{(\alpha,\beta)\in\Gamma} \left(\alpha \tfrac{x}{x+\xi} + \beta\right) \\ &= \sup_{(\alpha,\beta)\in\Gamma} \big((\alpha+\beta)x + \beta\xi\big). \end{aligned}$$

(iii) $\Rightarrow$ (iv). Follows just as (ii) $\Rightarrow$ (iv).

(iv) $\Rightarrow$ (i). Let F_i, $i \in I$, be a family of convex functions on $\mathbb{R}^2_+$ with upper envelope $F := \sup_{i\in I} F_i$. For all $t \in (0,1)$ and $(x,\xi), (y,\eta) \in \mathbb{R}^2_+$ we have

$$\begin{aligned} F_i[t(x,\xi) + (1-t)(y,\eta)] &\leqslant tF_i[x,\xi] + (1-t)F_i[y,\eta] \\ &\leqslant tF[x,\xi] + (1-t)F[y,\eta] \end{aligned}$$

and take the sup on the left side, where upon we see that F is again convex. The claim follows since affine-linear functions are trivially convex. $\square$

We can use the criterion Lemma 13.15(iii) to construct many examples of convex functions in $\mathbb{R}^2_+$.

Example 13.16 The following functions on $\mathbb{R}^2_+$ are positive homogeneous and concave:

(i) $F[x,\xi] = x^{1/p}\xi^{1/q}$, where $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$,

(ii) $G[x,\xi] = (x^{1/p} + \xi^{1/p})^p$, where $p \geqslant 1$,

(iii) $H[x,\xi] = (x^{1/p} + \xi^{1/p})^p + |x^{1/p} - \xi^{1/p}|^p$, where $p \geqslant 2$,

resp. positive homogeneous and convex:

(iv) $F[x,\xi] = x^{1/p}\xi^{1/q}$, where $p \in (0,1)$ and $\frac{1}{p} + \frac{1}{q} = 1$,

(v) $G[x,\xi] = (x^{1/p} + \xi^{1/p})^p$, where $0 \leqslant p \leqslant 1$,

(vi) $H[x,\xi] = (x^{1/p} + \xi^{1/p})^p + |x^{1/p} - \xi^{1/p}|^p$, where $1 \leqslant p \leqslant 2$.

Indeed: apply Lemma 13.15(iii) to $f(x) = F[x,1]$, $g(x) = G[x,1]$, $h(x) = H[x,1]$ and calculate the second derivatives. This gives

$$\begin{aligned}
f''(x) &= \frac{1-p}{p^2} x^{\frac{1}{p}-2},\\
g''(x) &= \frac{1-p}{p} x^{\frac{1}{p}-2}\big(x^{\frac{1}{p}}+1\big)^{p-2},\\
h''(x) &= \frac{1-p}{p} x^{\frac{1}{p}-2}\Big[\big(x^{\frac{1}{p}}+1\big)^{p-2} - \big|x^{\frac{1}{p}}-1\big|^{p-2}\Big].
\end{aligned}$$

The claim follows since any twice differentiable function is convex [concave] if the second derivative is positive [negative].

Theorem 13.17 (Jensen's inequality in $\mathbb{R}^2_+$) *Let $(X, \mathscr{A}, \mu)$ be a measure space and $F : \mathbb{R}^2_+ \to \mathbb{R}$ be a continuous and positive homogeneous function. If F is convex, then*

$$F\left[\int u\,d\mu, \int v\,d\mu\right] \leqslant \int F[u,v]d\mu \quad \forall u,v \in \mathcal{L}^1(\mu),\ u,v \geqslant 0. \tag{13.17}$$

In particular, $F[u,v] \in \mathcal{L}^1(\mu)$. If F is concave, then

$$\int F[u,v]d\mu \leqslant F\left[\int u\,d\mu, \int v\,d\mu\right] \quad \forall u,v \in \mathcal{L}^1(\mu),\ u,v \geqslant 0. \tag{13.18}$$

Proof Since $(x,\xi) \mapsto F(x,\xi)$ is continuous, and $x \mapsto (u(x), v(x))$ is measurable,[5] the composition $x \mapsto F[u(x), v(x)]$ is measurable and all integrals appearing in the statement of the theorem are well-defined. Using the homogeneity of F we get for positive $u, v \geqslant 0$

$$|F[u,v]| = (u+v)\big|F\big[\tfrac{u}{u+v}, \tfrac{v}{u+v}\big]\big| \leqslant (u+v) \sup_{x\in[0,1]} |F[x, 1-x]| \in \mathcal{L}^1(\mu)$$

since $F[x, 1-x]$ is continuous, and hence bounded on $[0,1]$.

Now we can use Lemma 13.16(iv) to deduce

$$\begin{aligned}
F\left[\int u\,d\mu, \int v\,d\mu\right] &= \sup_{(\alpha,\beta)\in C}\left(\alpha\int u\,d\mu + \beta\int v\,d\mu\right)\\
&= \sup_{(\alpha,\beta)\in C}\int(\alpha u + \beta v)d\mu \leqslant \int F[u,v]d\mu.
\end{aligned}$$

The concavity estimate (13.18) follows similarly. □

[5] Note that $\{(u,v) \in [a,b) \times [c,d)\} = \{u \in [a,b)\} \times \{v \in [c,d)\}$ for all $a < b$ and $c < d$.

Combining Example 13.16 and Theorem 13.17 gives new proofs of the Hölder and Minkowski inequalities as well as the Hanner inequality. In analogy to (13.1) we write $\|u\|_r = \left(\int |u|^r \, d\mu\right)^{1/r}$ for *all* $r \in \mathbb{R} \setminus \{0\}$.

Corollary 13.18 *Let $(X, \mathscr{A}, \mu)$ be a measure space.*

(i) (*Hölder inequality*) *Let $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, $u \in \mathcal{L}^p(\mu)$ and $v \in \mathcal{L}^q(\mu)$. Then $uv \in \mathcal{L}^1(\mu)$ and*

$$\int |uv| d\mu \leqslant \|u\|_p \|v\|_q.$$

(ii) (*Reverse Hölder inequality*) *Let $0 < r < 1$, $r^{-1} + s^{-1} = 1$, $u \in \mathcal{L}^r(\mu)$ and $v \in \mathcal{L}^s(\mu)$. Then we have*

$$\|u\|_r \|v\|_s \leqslant \int |uv| d\mu.$$

(iii) (*Minkowski inequality*) *Let $1 \leqslant p < \infty$ and $u, v \in \mathcal{L}^p(\mu)$. Then $u + v \in \mathcal{L}^p(\mu)$ and*

$$\|u + v\|_p \leqslant \| |u| + |v| \|_p \leqslant \|u\|_p + \|v\|_p.$$

(iv) (*Reverse Minkowski inequality*) *Let $0 < r < 1$ and $u, v \in \mathcal{L}^r(\mu)$. Then we have*

$$\|u\|_r + \|v\|_r \leqslant \| |u| + |v| \|_r.$$

(v) (*Hanner's inequality*) *Let $1 \leqslant p \leqslant 2$ and $u, v \in \mathcal{L}^p(\mu)$. Then we have*

$$|\|u\|_p - \|v\|_p|^p + (\|u\|_p + \|v\|_p)^p \leqslant \|u + v\|_p^p + \|u - v\|_p^p.$$

(vi) (*Hanner's inequality*) *Let $2 \leqslant q < \infty$ and $u, v \in \mathcal{L}^q(\mu)$. Then we have*

$$\|u + v\|_q^q + \|u - v\|_q^q \leqslant |\|u\|_q - \|v\|_q|^q + (\|u\|_q + \|v\|_q)^q.$$

Proof (i) Use Theorem 13.17 for the concave function $F[x, \xi] = x^{1/p} \xi^{1/q}$ with $x = |u|^p$ and $\xi = |v|^q$:

$$\int |uv| d\mu = \int F[|u|^p, |v|^q] d\mu \leqslant F\left[\int |u|^p \, d\mu, \int |v|^q \, d\mu\right] = \|u\|_p \|v\|_q.$$

(ii), (iii), (iv) follow in a similar way using the functions $F[x, \xi] = x^{1/r} \xi^{1/s}$ (convex), $G[x, \xi] = (x^{1/p} + \xi^{1/p})^p$ (concave) and $G[x, \xi] = (x^{1/r} + \xi^{1/r})^r$ (convex), respectively.

(v) Here we take $H[x,\xi] = (x^{1/p} + \xi^{1/p})^p + |x^{1/p} - \xi^{1/p}|^p$, $x = |u|^p$, $\xi = |v|^p$, and observe that H is convex if $p \in [1,2]$. Thus,

$$\begin{aligned}(\|u\|_p + \|v\|_p)^p + |\|u\|_p - \|v\|_p|^p &= H\left(\int |u|^p\,d\mu, \int |v|^p\,d\mu\right)\\ &\leqslant \int H(|u|^p, |v|^p)d\mu\\ &= \int (|u| + |v|)^p\,d\mu + \int ||u| - |v||^p\,d\mu.\end{aligned}$$

Notice that for $x, y \in \mathbb{R}$

$$|x+y|^p + |x-y|^p = (|x| + |y|)^p + ||x| - |y||^p$$

(consider the four cases $x, y \geqslant 0$, $x, y \leqslant 0$, $x \leqslant 0 \leqslant y$ and $y \leqslant 0 \leqslant x$), and from this we conclude that

$$\int (|u| + |v|)^p\,d\mu + \int ||u| - |v||^p\,d\mu = \|u+v\|_p^p + \|u-v\|_p^p.$$

(vi) follows analogously to (v). □

Problems

13.1. Let $(X, \mathscr{A}, \mu)$ be a finite measure space and let $1 \leqslant q < p < \infty$.

(i) Show that $\|u\|_q \leqslant \mu(X)^{1/q - 1/p}\|u\|_p$.
[Hint: use Hölder's inequality for $u \cdot 1$.]

(ii) Conclude that $\mathcal{L}^p(\mu) \subset \mathcal{L}^q(\mu)$ for all $p \geqslant q \geqslant 1$ and that a Cauchy sequence in $\mathcal{L}^p$ is also a Cauchy sequence in $\mathcal{L}^q$.

(iii) Is this still true if the measure μ is not finite?

13.2. Let $(X, \mathscr{A}, \mu)$ be a general measure space and $1 \leqslant p \leqslant r \leqslant q \leqslant \infty$.
Prove that $\mathcal{L}^p(\mu) \cap \mathcal{L}^q(\mu) \subset \mathcal{L}^r(\mu)$ by establishing the inequality

$$\|u\|_r \leqslant \|u\|_p^{\lambda} \cdot \|u\|_q^{1-\lambda} \qquad \forall u \in \mathcal{L}^p(\mu) \cap \mathcal{L}^q(\mu),$$

with $\lambda = \left(\frac{1}{r} - \frac{1}{q}\right) \Big/ \left(\frac{1}{p} - \frac{1}{q}\right)$.
[Hint: use Hölder's inequality.]

13.3. Let $(X, \mathscr{A}, \mu)$ be a measure space and $u, v \in \mathcal{L}^p(\mu)$.

(i) Find conditions which guarantee that uv, $u + v$ and αu, $\alpha \in \mathbb{R}$ are in $\mathcal{L}^p(\mu)$.

(ii) Show that $\mathcal{L}^1(\mu)$ and $\mathcal{L}^2(\mu)$ are, in general, no algebras.

(iii) Show that the lower triangle inequality holds:

$$\big|\|u\|_p - \|v\|_p\big| \leqslant \|u - v\|_p.$$

13.4. Let Ω be a set and $B, B^c \subset \Omega$ such that B and B^c are not empty.

(i) Find all measurable functions $u : (\Omega, \{\emptyset, \Omega\}) \to (\mathbb{R}, \mathscr{A})$, if (a) $\mathscr{A} := \{\emptyset, \mathbb{R}\}$, (b) $\mathscr{A} := \mathscr{B}(\mathbb{R})$ and (c) $\mathscr{A} := \mathscr{P}(\mathbb{R})$.

(ii) What does $u \in \mathcal{L}^p(\Omega, \sigma(B), \mu)$, $p > 0$, look like?

13.5. **Generalized Hölder inequality.** Iterate Hölder's inequality to derive the following generalization:

$$\int |u_1 \cdot u_2 \cdot \dots \cdot u_N| d\mu \leqslant \|u_1\|_{p_1} \cdot \|u_2\|_{p_2} \cdot \dots \cdot \|u_N\|_{p_N} \tag{13.19}$$

for all $p_n \in (1, \infty)$ such that $\sum_{n=1}^N p_n^{-1} = 1$ and all measurable $u_n \in \mathcal{M}(\mathscr{A})$.

13.6. **Young functions.** Let $\phi : [0, \infty) \to [0, \infty)$ be a strictly increasing continuous function such that $\phi(0) = 0$ and $\lim_{\xi\to\infty} \phi(\xi) = \infty$. Denote by $\psi(\eta) := \phi^{-1}(\eta)$ the inverse function. The functions

$$\Phi(A) := \int_{[0,A)} \phi(\xi)\lambda^1(d\xi) \quad \text{and} \quad \Psi(B) := \int_{[0,B)} \psi(\eta)\lambda^1(d\eta) \tag{13.20}$$

are called *conjugate Young functions*. Adapt the proof of Lemma 13.1 to show the following general *Young inequality*:

$$AB \leqslant \Phi(A) + \Psi(B) \qquad \forall A, B \geqslant 0. \tag{13.21}$$

[Hint: interpret $\Phi(A)$ and $\Psi(B)$ as areas below the graph of $\phi(\xi)$ resp. $\psi(\eta)$.]

13.7. Let $1 \leqslant p < \infty$ and $u, u_n \in \mathcal{L}^p(\mu)$ such that $\sum_{n=1}^\infty \|u - u_n\|_p < \infty$. Show that almost everywhere $\lim_{n\to\infty} u_n(x) = u(x)$.
[Hint: mimic the proof of the Riesz–Fischer theorem using $\sum_n (u_{n+1} - u_n)$.]

13.8. Consider one-dimensional Lebesgue measure on $[0, 1]$. Verify that the sequence $u_n(x) := n\mathbb{1}_{(0,1/n)}(x)$, $n \in \mathbb{N}$, converges pointwise to the function $u \equiv 0$, but that no subsequence of u_n converges in $\mathcal{L}^p$-sense for any $p \geqslant 1$.

13.9. Let $p, q \in [1, \infty]$ be conjugate, i.e. $p^{-1} + q^{-1} = 1$, and assume that $(u_k)_{k\in\mathbb{N}} \subset \mathcal{L}^p$ and $(w_k)_{k\in\mathbb{N}} \subset \mathcal{L}^q$ are sequences with limits u and w in $\mathcal{L}^p$-sense, resp. $\mathcal{L}^q$-sense. Show that $u_k w_k$ converges in $\mathcal{L}^1$ to the function uw.

13.10. Prove that $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^2(\mu)$ converges in $\mathcal{L}^2$ if, and only if, $\lim_{n,m\to\infty} \int u_n u_m \, d\mu$ exists.
[Hint: verify and use the identity $\|u - w\|_2^2 = \|u\|_2^2 + \|w\|_2^2 - 2 \int uw \, d\mu$.]

13.11. Let $(X, \mathscr{A}, \mu)$ be a finite measure space. Show that every measurable $u \geqslant 0$ with $\int \exp(hu(x))\mu(dx) < \infty$ for some $h > 0$ is in $\mathcal{L}^p(\mu)$ for every $p \geqslant 1$.
[Hint: check that $|t|^N/N! \leqslant e^{|t|}$ implies $u \in \mathcal{L}^N$, $N \in \mathbb{N}$; then use Problem 13.1.]

13.12. Let λ be Lebesgue measure in $(0, \infty)$ and $p, q \geqslant 1$ arbitrary.

(i) Show that $u_n(x) := n^\alpha (x + n)^{-\beta}$ ($\alpha \in \mathbb{R}$, $\beta > 1$) holds for every $n \in \mathbb{N}$ in $\mathcal{L}^p(\lambda)$.
(ii) Show that $v_n(x) := n^\gamma e^{-nx}$ ($\gamma \in \mathbb{R}$) holds for every $n \in \mathbb{N}$ in $\mathcal{L}^q(\lambda)$.

13.13. Let $u(x) = (x^\alpha + x^\beta)^{-1}$, $x, \alpha, \beta > 0$. For which $p \geqslant 1$ is $u \in \mathcal{L}^p(\lambda^1, (0, \infty))$?

13.14. Consider the measure space $(\Omega = \{1, 2, \dots, n\}, \mathscr{P}(\Omega), \mu)$, $n \geqslant 2$, where μ is the counting measure. Show that $\left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ is a norm if $p \in [1, \infty)$, but not for $p \in (0, 1)$.
[Hint: you can identify $\mathcal{L}^p(\mu)$ with $\mathbb{R}^n$.]

13.15. Let $(X, \mathscr{A}, \mu)$ be a measure space. The space $\mathcal{L}^p(\mu)$ is called *separable* if there exists a countable dense subset $\mathscr{D}_p \subset \mathcal{L}^p(\mu)$. Show that $\mathcal{L}^p(\mu)$, $p \in (1, \infty)$, is separable if, and only if, $\mathcal{L}^1(\mu)$ is separable.
[Hint: use Riesz's convergence theorem, Theorem 13.10.]

13.16. Let $u_n \in \mathcal{L}^p$, $p \geqslant 1$, for all $n \in \mathbb{N}$. What can you say about u and w if you know that $\lim_{n\to\infty} \int |u_n - u|^p \, d\mu = 0$ and $\lim_{n\to\infty} u_n(x) = w(x)$ for almost every x?

13.17. Let $(X, \mathscr{A}, \mu)$ be a finite measure space and let $u \in \mathcal{L}^1(\mu)$ be strictly positive with $\int u \, d\mu = 1$. Show that

$$\int (\log u) d\mu \leqslant \mu(X) \log\left(\frac{1}{\mu(X)}\right).$$

13.18. Let u be a positive measurable function on $[0,1]$. Which of the following is larger:

$$\int_{(0,1)} u(x)\log u(x)\lambda(dx) \qquad \text{or} \qquad \int_{(0,1)} u(s)\lambda(ds)\cdot\int_{(0,1)} \log u(t)\lambda(dt)?$$

[Hint: show that $\log x \leqslant x\log x$, $x>0$, and assume first that $\int u\,d\lambda = 1$, then consider $u/\int u\,d\lambda$.]

13.19. Let $(X,\mathscr{A},\mu)$ be a measure space and $p\in(0,1)$. The conjugate index is given by $q := p/(p-1) < 0$. Prove for all measurable $u,v,w: X\to(0,\infty)$ with $\int u^p\,d\mu$, $\int v^p\,d\mu<\infty$ and $0<\int w^q\,d\mu<\infty$ the inequalities

$$\int uw\,d\mu \geqslant \left(\int u^p\,d\mu\right)^{1/p}\left(\int w^q\,d\mu\right)^{1/q}$$

and

$$\left(\int (u+v)^p\,d\mu\right)^{1/p} \geqslant \left(\int u^p\,d\mu\right)^{1/p} + \left(\int v^p\,d\mu\right)^{1/p}.$$

[Hint: consider Hölder's inequality for u and $1/w$.]

13.20. Let $(X,\mathscr{A},\mu)$ be a finite measure space and $u\in\mathcal{M}(\mathscr{A})$ be a bounded function with $\|u\|_\infty>0$. Prove that for all $n\in\mathbb{N}$

(i) $M_n := \int |u|^n\,d\mu \in (0,\infty)$;
(ii) $M_{n+1}M_{n-1}\geqslant M_n^2$;
(iii) $\mu(X)^{-1/n}\|u\|_n \leqslant M_{n+1}/M_n \leqslant \|u\|_\infty$;
(iv) $\lim_{n\to\infty} M_{n+1}/M_n = \|u\|_\infty$.

[Hints: for (ii) use Hölder's inequality; for (iii) use Jensen's inequality for the lower estimate and use Hölder's inequality for the upper estimate; and for (iv) observe that $\int u^n\,d\mu \geqslant \int_{\{u>\|u\|_\infty-\epsilon\}}(\|u\|_\infty-\epsilon)^n\,d\mu = \mu\{u>\|u\|_\infty-\epsilon\}(\|u\|_\infty-\epsilon)^n$, take the nth root and let $n\to\infty$.]

13.21. Let $(X,\mathscr{A},\mu)$ be a general measure space and let $u\in\bigcap_{p\geqslant 1}\mathcal{L}^p(\mu)$. Then

$$\lim_{p\to\infty}\|u\|_p = \|u\|_\infty,$$

where $\|u\|_\infty=\infty$ if u is unbounded.
[Hint: start with $\|u\|_\infty<\infty$. Show that for any sequence $q_n\to\infty$ one has $\|u\|_{p+q_n}\leqslant \|u\|_\infty^{q_n/(p+q_n)}\cdot\|u\|_p^{p/(p+q_n)}$ and conclude that $\limsup_{p\to\infty}\|u\|_p\leqslant\|u\|_\infty$. The other estimate follows from $\|u\|_p\geqslant \mu(\{|u|>(1-\epsilon)\|u\|_\infty\})^{1/p}(1-\epsilon)\|u\|_\infty$ and $p\to\infty,\epsilon\to 0$, see also the hint to Problem 13.20, where $\mu(\dots)$ is finite in view of the Markov inequality. If $\|u\|_\infty=\infty$, use the First part of the hint and observe that

$$\begin{aligned}\liminf_{p\to\infty}\sup_{k\in\mathbb{N}}\||u|\wedge k\|_p &\geqslant \sup_{k\in\mathbb{N}}\lim_{p\to\infty}\||u|\wedge k\|_p = \sup_{k\in\mathbb{N}}\||u|\wedge k\|_\infty\\ &= \sup_{k\in\mathbb{N}}\sup_x(|u(x)|\wedge k) = \sup_x\sup_{k\in\mathbb{N}}(|u(x)|\wedge k) = \|u\|_\infty=\infty.]\end{aligned}$$

13.22. Let $(X,\mathscr{A},\mu)$ be a probability space and assume that $\|u\|_q<\infty$ for some $q>0$. Show that $\lim_{p\to 0}\|u\|_p = \exp\left(\int\log|u|d\mu\right)$ (we set $e^{-\infty}:=0$).

13.23. Let $(X,\mathscr{A},\mu)$ be a measure space and $1\leqslant p<\infty$. Show that $f\in\mathcal{E}(\mathscr{A})\cap\mathcal{L}^p(\mu)$ if, and only if, $f\in\mathcal{E}(\mathscr{A})$ and $\mu\{f\neq 0\}<\infty$.
In particular, $\mathcal{E}(\mathscr{A})\cap\mathcal{L}^p(\mu) = \mathcal{E}(\mathscr{A})\cap\mathcal{L}^1(\mu)$.

13.24. Variants of Jensen's inequality. Let $(X, \mathscr{A}, \mu)$ be a probability space.

(i) Show Jensen's inequality for convex $V : \mathbb{R} \to \mathbb{R}$, see Example 13.14(v).

(ii) Show Jensen's inequality for concave $\Lambda : \mathbb{R} \to \mathbb{R}$, see Example 13.14(vi).

(iii) Let $-\infty \leqslant a < b \leqslant \infty$ and $V : (a, b) \to \mathbb{R}$ be a convex function. For $u : X \to (a, b)$ such that $u, V(u) \in \mathcal{L}^1(\mu)$ we have $\int u\, d\mu \in (a, b)$ and $V\left(\int u\, d\mu\right) \leqslant \int V(u) d\mu$.

(iv) Let $-\infty \leqslant a < b \leqslant \infty$ and $\Lambda : (a, b) \to \mathbb{R}$ be a concave function. For $u : X \to (a, b)$ such that $u, \Lambda(u) \in \mathcal{L}^1(\mu)$ we have $\int u\, d\mu \in (a, b)$ and $\int \Lambda(u) d\mu \leqslant \Lambda\left(\int u\, d\mu\right)$.

13.25. Use Jensen's inequality (Example 13.14(i), (ii)) to derive Hölder's inequality and Minkowski's inequality. Use

$$\Lambda(x) = x^{1/q},\ x \geqslant 0, \quad w = |f|^p \quad \text{and} \quad u = |g|^q |f|^{-p} \mathbb{1}_{\{f \neq 0\}}$$

for Hölder's inequality and

$$\Lambda(x) = (x^{1/p} + 1)^p,\ x \geqslant 0, \quad w = |f|^p \mathbb{1}_{\{f \neq 0\}} \quad \text{and} \quad u = |f|^{-p} |g|^p \mathbb{1}_{\{f \neq 0\}}$$

for Minkowski's inequality.

13.26. Let $(X, \mathscr{A}, \mu)$ be a finite measure space, $1 \leqslant p < \infty$ and $f \in \mathcal{L}^p$. Show that for $m := \int f\, d\mu$ and all $a \in \mathbb{R}$ the following inequality holds:

$$2^{-p} \int |f(x) - m|^p\, \mu(dx) \leqslant \int |f(x) - a|^p\, \mu(dx).$$

14
Product Measures and Fubini's Theorem

Lebesgue measure on $\mathbb{R}^n$ has, inherent in its definition, the following interesting additional property: if $n > d \geqslant 1$

$$\begin{aligned}\lambda^n\big([a_1, b_1) \times \cdots \times [a_n, b_n)\big)\\ &= (b_1 - a_1) \cdot \cdots \cdot (b_d - a_d) \cdot (b_{d+1} - a_{d+1}) \cdot \cdots \cdot (b_n - a_n)\\ &= \lambda^d\big([a_1, b_1) \times \cdots \times [a_d, b_d)\big) \cdot \lambda^{n-d}\big([a_{d+1}, b_{d+1}) \times \cdots \times [a_n, b_n)\big),\end{aligned} \tag{14.1}$$

i.e. it is – at least for rectangles – the product of Lebesgue measures in lower-dimensional spaces. In this chapter we will see that (14.1) remains true for any product $A \times B$ of sets $A \in \mathscr{B}(\mathbb{R}^d)$ and $B \in \mathscr{B}(\mathbb{R}^{n-d})$. More importantly, we will prove the following version of Cavalieri's principle (Fig. 14.1), which just says that we can carve up the set $E \subset \mathbb{R}^n$ horizontally or vertically, measure the volume of the slices and 'sum' them up along the other direction to get the volume of the whole set E:

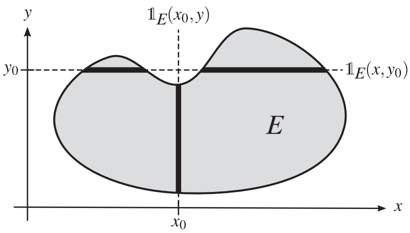


Fig. 14.1. Cavalieri's principle.

$$\begin{aligned}\lambda^n(E) = \int \mathbb{1}_E(x, y)\lambda^n(d(x, y)) &= \int \left(\int \mathbb{1}_E(x, y_0)\lambda^d(dx)\right) \lambda^{n-d}(dy_0)\\ &= \int \left(\int \mathbb{1}_E(x_0, y)\lambda^{n-d}(dy)\right) \lambda^d(dx_0).\end{aligned}$$

Clearly, we must be careful about the measurability of products of sets. Recall the following simple rules for Cartesian products of sets $A, A', A_i \subset X$, $i \in I$

(I is an arbitrary index set), and $B, B' \subset Y$:

$$\begin{aligned}(\textstyle\bigcup_i A_i) \times B &= \textstyle\bigcup_i (A_i \times B),\\ (\textstyle\bigcap_i A_i) \times B &= \textstyle\bigcap_i (A_i \times B),\\ (A \times B) \cap (A' \times B') &= (A \cap A') \times (B \cap B'),\\ A^c \times B &= (X \times B) \setminus (A \times B),\\ A \times B \subset A' \times B' &\iff A \subset A' \text{ und } B \subset B',\end{aligned} \tag{14.2}$$

which are easily derived [✍] from the formula

$$A \times B = (A \times Y) \cap (X \times B) = \pi_1^{-1}(A) \cap \pi_2^{-1}(B),$$

where $\pi_1 : X \times Y \to X$ and $\pi_2 : X \times Y \to Y$ are the coordinate projections, and the compatibility of inverse mappings and set operations. To treat measurability, we assume throughout this chapter that **$(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ are σ-finite measure spaces**.

Following (14.1) we want to define a measure ρ on rectangles of the form $A \times B$ such that $\rho(A \times B) = \mu(A)\nu(B)$ for $A \in \mathscr{A}$ and $B \in \mathscr{B}$. The first problem which we encounter is that the family

$$\mathscr{A} \times \mathscr{B} := \{A \times B : A \in \mathscr{A},\ B \in \mathscr{B}\} \tag{14.3}$$

is, in general, no σ-algebra – it contains only 'rectangles'. [✍]

Lemma 14.1 *Let $\mathscr{A}$ and $\mathscr{B}$ be two σ-algebras (or semi-rings), then $\mathscr{A} \times \mathscr{B}$ is a semi-ring.*[1]

Proof The proof is literally the same as the induction step in the proof of Proposition 6.5. We have to check (S_1)–(S_3) for $\mathscr{A} \times \mathscr{B}$. Let $A, A' \in \mathscr{A}$ and $B, B' \in \mathscr{B}$.

(S_1) follows from $\emptyset = \emptyset \times \emptyset \in \mathscr{A} \times \mathscr{B}$.

(S_2) follows from $(A \times B) \cap (A' \times B') = (A \cap A') \times (B \cap B') \in \mathscr{A} \times \mathscr{B}$.

(S_3) We have

$$\begin{aligned}(A \times B)^c &= \{(x, y) : x \notin A,\ y \in B \text{ or } x \in A,\ y \notin B \text{ or } x \notin A,\ y \notin B\}\\ &= (A^c \times B) \mathbin{\dot\cup} (A \times B^c) \mathbin{\dot\cup} (A^c \times B^c)\end{aligned}$$

and the formula used for the proof of (S_2) gives

$$\begin{aligned}&(A \times B) \setminus (A' \times B')\\ &\quad = (A \times B) \cap (A' \times B')^c\\ &\quad = \left[(A \setminus A') \times (B \cap B')\right] \mathbin{\dot\cup} \left[(A \cap A') \times (B \setminus B')\right] \mathbin{\dot\cup} \left[(A \setminus A') \times (B \setminus B')\right].\end{aligned}$$

[1] See (S_1)–(S_3) on p. 39 for the definition of a semi-ring.

By assumption, $A \setminus A'$ and $B \setminus B'$ can be written as finite disjoint unions of sets from $\mathscr{A}$ and $\mathscr{B}$, respectively. This proves (S_3). □

Definition 14.2 Let $(X, \mathscr{A})$ and $(Y, \mathscr{B})$ be measurable spaces. Then the σ-algebra $\mathscr{A} \otimes \mathscr{B} := \sigma(\mathscr{A} \times \mathscr{B})$ is called a *product σ-algebra*, and $(X \times Y, \mathscr{A} \otimes \mathscr{B})$ is the *product of measurable spaces*.

The following lemma is quite useful since it allows us to reduce considerations for $\mathscr{A} \otimes \mathscr{B}$ to generators $\mathscr{F}$ and $\mathscr{G}$ of $\mathscr{A}$ and $\mathscr{B}$ – just as we did in (14.1).

Lemma 14.3 *If $\mathscr{A} = \sigma(\mathscr{F})$ and $\mathscr{B} = \sigma(\mathscr{G})$ and if $\mathscr{F}$ and $\mathscr{G}$ contain exhausting sequences $(F_n)_{n \in \mathbb{N}} \subset \mathscr{F}$, $F_n \uparrow X$ and $(G_n)_{n \in \mathbb{N}} \subset \mathscr{G}$, $G_n \uparrow Y$, then*

$$\sigma(\mathscr{F} \times \mathscr{G}) = \sigma(\mathscr{A} \times \mathscr{B}) \stackrel{\text{def}}{=} \mathscr{A} \otimes \mathscr{B}.$$

Proof Since $\mathscr{F} \times \mathscr{G} \subset \mathscr{A} \times \mathscr{B}$, we have $\sigma(\mathscr{F} \times \mathscr{G}) \subset \mathscr{A} \otimes \mathscr{B}$. On the other hand, the system

$$\Sigma := \left\{ A \in \mathscr{A} : A \times G \in \sigma(\mathscr{F} \times \mathscr{G}) \quad \forall G \in \mathscr{G} \right\}$$

is a σ-algebra. Let $A, A_n \in \Sigma$, $n \in \mathbb{N}$, and $G \in \mathscr{G}$; (Σ_1) follows from

$$X \times G = \bigcup_{n \in \mathbb{N}} \underbrace{(F_n \times G)}_{\in \mathscr{F} \times \mathscr{G}} \in \sigma(\mathscr{F} \times \mathscr{G}),$$

(Σ_2) from $A^c \times G = (X \times G) \setminus (A \times G) \in \sigma(\mathscr{F} \times \mathscr{G})$ and (Σ_3) from

$$\Big(\bigcup_{n \in \mathbb{N}} A_n \Big) \times G = \bigcup_{n \in \mathbb{N}} \underbrace{(A_n \times G)}_{\in \sigma(\mathscr{F} \times \mathscr{G})} \in \sigma(\mathscr{F} \times \mathscr{G}).$$

Obviously, $\mathscr{F} \subset \Sigma \subset \mathscr{A}$, and therefore $\Sigma = \mathscr{A}$; by the very definition of Σ we conclude that $\mathscr{A} \times \mathscr{G} \subset \sigma(\mathscr{F} \times \mathscr{G})$. A similar consideration shows that we have $\mathscr{F} \times \mathscr{B} \subset \sigma(\mathscr{F} \times \mathscr{G})$. This means that for all $A \in \mathscr{A}$ and $B \in \mathscr{B}$

$$A \times B = (A \times Y) \cap (X \times B) = \bigcup_{k,n \in \mathbb{N}} \underbrace{(A \times G_k)}_{\in \sigma(\mathscr{F} \times \mathscr{G})} \cap \underbrace{(F_n \times B)}_{\in \sigma(\mathscr{F} \times \mathscr{G})},$$

so that $\mathscr{A} \times \mathscr{B} \subset \sigma(\mathscr{F} \times \mathscr{G})$ and thus $\mathscr{A} \otimes \mathscr{B} \subset \sigma(\mathscr{F} \times \mathscr{G})$. □

If the generators $\mathscr{F}$, $\mathscr{G}$ are rich enough, we have not too many choices of measures ρ with $\rho(F \times G) = \mu(F)\nu(G)$. In fact, we can write the following theorem.

Theorem 14.4 (uniqueness of product measures) *Let $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ be two measure spaces and assume that $\mathscr{A} = \sigma(\mathscr{F})$ and $\mathscr{B} = \sigma(\mathscr{G})$. If*

- *$\mathscr{F}, \mathscr{G}$ are $\cap$-stable,*
- *$\mathscr{F}, \mathscr{G}$ contain exhausting sequences $F_k \uparrow X$ and $G_k \uparrow Y$ with $\mu(F_k) < \infty$ and $\nu(G_n) < \infty$ for all $k, n \in \mathbb{N}$,*

then there is at most one measure ρ on $(X \times Y, \mathscr{A} \otimes \mathscr{B})$ satisfying

$$\rho(F \times G) = \mu(F)\nu(G) \qquad \forall F \in \mathscr{F},\ G \in \mathscr{G}.$$

Proof By Lemma 14.3 $\mathscr{F} \times \mathscr{G}$ generates $\mathscr{A} \otimes \mathscr{B}$. Moreover, $\mathscr{F} \times \mathscr{G}$ inherits the $\cap$-stability of $\mathscr{F}$ and $\mathscr{G}$ [✍], the sequence $F_n \times G_n$ increases towards $X \times Y$ and $\rho(F_n \times G_n) = \mu(F_n)\nu(G_n) < \infty$. These are the assumptions of the uniqueness theorem for measures (Theorem 5.7), showing that there is at most one such product measure ρ. □

As so often, it is the existence which is more difficult to prove than uniqueness.

Theorem 14.5 (existence of product measures) *Let $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ be σ-finite measure spaces. The set function*

$$\rho : \mathscr{A} \times \mathscr{B} \to [0, \infty], \quad \rho(A \times B) := \mu(A)\nu(B),$$

extends uniquely to a σ-finite measure on $(X \times Y, \mathscr{A} \otimes \mathscr{B})$ such that

$$\rho(E) = \iint \mathbb{1}_E(x,y)\mu(dx)\nu(dy) = \iint \mathbb{1}_E(x,y)\nu(dy)\mu(dx) \tag{14.4}$$

holds for all $E \in \mathscr{A} \otimes \mathscr{B}$.[2] *In particular, the functions*

$$x \mapsto \mathbb{1}_E(x,y),\ y \mapsto \mathbb{1}_E(x,y),\ x \mapsto \int \mathbb{1}_E(x,y)\nu(dy),\ y \mapsto \int \mathbb{1}_E(x,y)\mu(dx)$$

are $\mathscr{A}$, resp. $\mathscr{B}$-measurable for every fixed $y \in Y$, resp. $x \in X$.

Proof *Uniqueness* of ρ follows from Theorem 14.4.

Existence. Let $(A_n)_{n\in\mathbb{N}}$, $(B_n)_{n\in\mathbb{N}}$ be sequences in $\mathscr{A}$ resp. $\mathscr{B}$ with $A_n \uparrow X$, $B_n \uparrow Y$ and $\mu(A_n), \nu(B_n) < \infty$. Clearly, $E_n := A_n \times B_n \uparrow X \times Y$.

[2] We use the symbols $\int \ldots d\mu$ like brackets, i.e. $\iint \ldots d\mu\, d\nu = \int \left(\int \ldots d\mu \right) d\nu$.

For every $n\in\mathbb{N}$ we consider the family $\mathscr{D}_n$ of all subsets $D\subset X\times Y$ satisfying the following conditions:

- $x\mapsto \mathbb{1}_{D\cap E_n}(x,y)$ and $y\mapsto \mathbb{1}_{D\cap E_n}(x,y)$ are measurable,
- $x\mapsto \int \mathbb{1}_{D\cap E_n}(x,y)\nu(dy)$ and $y\mapsto \int \mathbb{1}_{D\cap E_n}(x,y)\mu(dx)$ are measurable,
- $\iint \mathbb{1}_{D\cap E_n}(x,y)\mu(dx)\nu(dy) = \iint \mathbb{1}_{D\cap E_n}(x,y)\nu(dy)\mu(dx).$

That $\mathscr{A}\times\mathscr{B}\subset\mathscr{D}_n$ follows from

$$\begin{aligned}\iint \mathbb{1}_{(A\times B)\cap E_n}(x,y)\mu(dx)\nu(dy) &= \iint \mathbb{1}_{A\cap A_n}(x)\mathbb{1}_{B\cap B_n}(y)\mu(dx)\nu(dy)\\ &=\mu(A\cap A_n)\int \mathbb{1}_{B\cap B_n}(y)\nu(dy)\\ &=\mu(A\cap A_n)\nu(B\cap B_n)\\ &=\cdots=\iint \mathbb{1}_{(A\times B)\cap E_n}(x,y)\nu(dy)\mu(dx),\end{aligned}$$

where the ellipsis ... stands for the same calculations run through backwards. In each step the measurability conditions needed to perform the integrations are fulfilled because of the product structure. [✍] In particular, $X\times Y, \emptyset, E_k\in\mathscr{D}_n$. If $D\in\mathscr{D}_n$, then $\mathbb{1}_{D^c\cap E_n}=\mathbb{1}_{E_n}-\mathbb{1}_{E_n\cap D}$ and

$$\begin{aligned}&\iint \mathbb{1}_{D^c\cap E_n}(x,y)\mu(dx)\nu(dy)\\ &\quad=\int\left(\int \mathbb{1}_{E_n}(x,y)\mu(dx)-\int \mathbb{1}_{E_n\cap D}(x,y)\mu(dx)\right)\nu(dy)\\ &\quad=\iint \mathbb{1}_{E_n}(x,y)\mu(dx)\nu(dy)-\iint \mathbb{1}_{E_n\cap D}(x,y)\mu(dx)\,\nu(dy)\\ &\quad=\iint \mathbb{1}_{E_n}(x,y)\nu(dy)\mu(dx)-\iint \mathbb{1}_{E_n\cap D}(x,y)\nu(dy)\mu(dx)\\ &\qquad\text{(by definition, since } E_n, D\in\mathscr{D}_n)\\ &\quad=\cdots=\iint \mathbb{1}_{D^c\cap E_n}(x,y)\nu(dy)\mu(dx).\end{aligned}$$

Again, in each step the measurability conditions hold since measurable functions form a vector space. If $(D_k)_{k\in\mathbb{N}}\subset\mathscr{D}_n$ are mutually disjoint sets, $D:=\bigcupdot_{k\in\mathbb{N}}D_k$, the linearity of the integral and Beppo Levi's theorem in the form of Corollary 9.9 show that

$$\begin{aligned}\iint \mathbb{1}_{D\cap E_n}(x,y)\mu(dx)\nu(dy) &= \int \left(\sum_{k=1}^{\infty}\int \mathbb{1}_{D_k\cap E_n}(x,y)\mu(dx)\right)\nu(dy)\\ &= \sum_{k=1}^{\infty}\iint \mathbb{1}_{D_k\cap E_n}(x,y)\mu(dx)\nu(dy)\\ &= \sum_{k=1}^{\infty}\iint \mathbb{1}_{D_k\cap E_n}(x,y)\nu(dy)\mu(dx)\\ &\qquad \text{(by definition, since } D_k \in \mathscr{D}_n)\\ &= \cdots = \iint \mathbb{1}_{D\cap E_n}(x,y)\nu(dy)\mu(dx)\end{aligned}$$

and the measurability conditions hold since measurability is preserved under sums and increasing limits.

The last three calculations show that $\mathscr{D}_n$ is a Dynkin system containing the $\cap$-stable family $\mathscr{A}\times\mathscr{B}$. By Theorem 5.5, $\mathscr{A}\otimes\mathscr{B}\subset\mathscr{D}_n$ for every $n\in\mathbb{N}$. Since $E_n\uparrow X\times Y$, Beppo Levi's theorem proves (14.4) along with the measurability of the functions $\mathbb{1}_E(\cdot\,,y)$, $\mathbb{1}_E(x,\,\cdot)$, $\int\mathbb{1}_E(\cdot\,,y)\nu(dy)$ and $\int\mathbb{1}_E(x,\,\cdot)\mu(dx)$ since $\mathcal{M}$ is stable under pointwise limits.

Replacing E_n by $X\times Y$ in the above calculations finally proves that

$$E\mapsto\rho(E):=\iint\mathbb{1}_E(x,y)\mu(dx)\nu(dy)$$

is indeed a measure on $(X\times Y,\mathscr{A}\otimes\mathscr{B})$ with $\rho(A\times B)=\mu(A)\nu(B)$. □

Definition 14.6 Let $(X,\mathscr{A},\mu)$ and $(Y,\mathscr{B},\nu)$ be two σ-finite measure spaces. The unique measure ρ constructed in Theorem 14.5 is called the *product* of the measures μ and ν, denoted by $\mu\times\nu$. $(X\times Y,\mathscr{A}\otimes\mathscr{B},\mu\times\nu)$ is called the *product measure space*.

Returning to the example considered at the beginning we finally can finish the construction of n-dimensional Lebesgue measure.

Corollary 14.7 *If* $n>d\geqslant 1$,

$$(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n),\lambda^n)=\left(\mathbb{R}^d\times\mathbb{R}^{n-d},\mathscr{B}(\mathbb{R}^d)\otimes\mathscr{B}(\mathbb{R}^{n-d}),\lambda^d\times\lambda^{n-d}\right).$$

The next step is to see how we can integrate w.r.t. $\mu\times\nu$. The following two results are often stated together as the Fubini or Fubini–Tonelli theorem. We prefer to distinguish between them since the first result, Theorem 14.8, says that we can *always swap iterated integrals of positive functions* (even if we get $+\infty$), whereas Corollary 14.9 applies to more general functions but requires the (iterated) integrals to be finite.

Theorem 14.8 (Tonelli) *Let $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ be σ-finite measure spaces and let $u : X \times Y \to [0, \infty]$ be $\mathscr{A} \otimes \mathscr{B}$-measurable. Then*

(i) *$x \mapsto u(x, y)$, $y \mapsto u(x, y)$ are $\mathscr{A}$- resp. $\mathscr{B}$-measurable for fixed y resp. x;*

(ii) *$x \mapsto \int_Y u(x, y)\nu(dy)$, $y \mapsto \int_X u(x, y)\mu(dx)$ are $\mathscr{A}$- resp. $\mathscr{B}$-measurable;*

(iii) $$\int_{X\times Y} u \, d(\mu \times \nu) = \int_Y \int_X u(x, y)\mu(dx)\nu(dy) = \int_X \int_Y u(x, y)\nu(dy)\mu(dx) \in [0, \infty].$$

Proof Since u is positive and $\mathscr{A} \otimes \mathscr{B}$-measurable, we find an increasing sequence of simple functions $f_n \in \mathcal{E}^+(\mathscr{A} \otimes \mathscr{B})$ with $\sup_{n\in\mathbb{N}} f_n = u$. Each f_n is of the form $f_n(x, y) = \sum_{k=0}^{N(n)} \alpha_k \mathbb{1}_{E_k}(x, y)$, where $\alpha_k \geqslant 0$ and the $E_k \in \mathscr{A} \otimes \mathscr{B}$, $0 \leqslant k \leqslant N(n)$, are disjoint. By Theorem 14.5, the fact that $\mathcal{M}(\mathscr{A} \otimes \mathscr{B})$ is a vector space and the linearity of the integral we conclude that

$$x \mapsto f_n(x, y), \quad y \mapsto f_n(x, y), \quad x \mapsto \int_Y f_n(x, y)\nu(dy), \quad y \mapsto \int_X f_n(x, y)\mu(dx)$$

are measurable functions and (i), (ii) follow from the usual Beppo Levi argument since $\mathcal{M}(\mathscr{A})$ and $\mathcal{M}(\mathscr{B})$ are stable under increasing limits, see Corollary 8.10. The linearity of the integral and Theorem 14.5 also show that

$$\int_{X\times Y} f_n \, d(\mu \times \nu) = \int_Y \int_X f_n \, d\mu \, d\nu = \int_X \int_Y f_n \, d\nu \, d\mu \qquad \forall n \in \mathbb{N},$$

and (iii) follows from several applications of Beppo Levi's theorem. □

Corollary 14.9 (Fubini's theorem) *Let $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ be σ-finite measure spaces and let $u : X \times Y \to \overline{\mathbb{R}}$ be $\mathscr{A} \otimes \mathscr{B}$-measurable. If at least one of the three integrals*

$$\int_{X\times Y} |u| d(\mu \times \nu), \quad \int_Y \int_X |u(x, y)|\mu(dx)\nu(dy), \quad \int_X \int_Y |u(x, y)|\nu(dy)\mu(dx)$$

is finite then all three integrals are finite, $u \in \mathcal{L}^1(\mu \times \nu)$, and

(i) *$x \mapsto u(x, y)$ is in $\mathcal{L}^1(\mu)$ for ν-a.e. $y \in Y$;*

(ii) *$y \mapsto u(x, y)$ is in $\mathcal{L}^1(\nu)$ for μ-a.e. $x \in X$;*

(iii) $y \mapsto \int_X u(x,y)\mu(dx)$ *is in* $\mathcal{L}^1(\nu)$;

(iv) $x \mapsto \int_Y u(x,y)\nu(dy)$ *is in* $\mathcal{L}^1(\mu)$;

(v) $\displaystyle\int_{X\times Y} u\, d(\mu\times\nu) = \int_Y \int_X u(x,y)\mu(dx)\nu(dy) = \int_X \int_Y u(x,y)\nu(dy)\mu(dx).$

Proof Tonelli's theorem, Theorem 14.8, shows that in $[0,\infty]$

$$\int_{X\times Y} |u| d(\mu\times\nu) = \int_Y \int_X |u| d\mu d\nu = \int_X \int_Y |u| d\nu d\mu. \tag{14.5}$$

If one of the integrals is finite, then all of them are finite and $u \in \mathcal{L}^1(\mu\times\nu)$ follows. Again by Tonelli's theorem, we have that $x \mapsto u^{\pm}(x,y)$ is $\mathscr{A}$-measurable and $y \mapsto \int u^{\pm}(x,y)\mu(dx)$ is $\mathscr{B}$-measurable. Since $u^{\pm} \leqslant |u|$, (14.5) and Corollary 11.6 show that

$$\int_X u^{\pm}(x,y)\mu(dx) \leqslant \int_X |u(x,y)|\mu(dx) < \infty \quad \text{for } \nu\text{-a.e. } y \in Y$$

and

$$\int_Y \int_X u^{\pm}(x,y)\mu(dx)\nu(dy) \leqslant \int_Y \int_X |u(x,y)|\mu(dx)\nu(dy) < \infty.$$

This proves (i) and (iii); (ii) and (iv) are shown in a similar way. Finally, (v) follows for u^+ and u^- from Theorem 14.8 and for $u = u^+ - u^-$ by linearity, since (i)–(iv) exclude the possibility of '$\infty - \infty$'. □

Integration by Parts and Two Interesting Integrals

Fubini's theorem can be used to show the integration-by-parts formula for Lebesgue integrals. Let $f, g : \mathbb{R} \to \mathbb{R}$ be Borel measurable functions which are *locally integrable*, i.e. such that $\int_I |f| d\lambda$ and $\int_I |g| d\lambda$ exist and are finite for every bounded interval $I \subset \mathbb{R}$. We adopt the convention

$$\int_a^b f(t)dt := \begin{cases} \int f(t)\mathbb{1}_{[a,b]}(t)dt & \text{if } a \leqslant b, \\ -\int f(t)\mathbb{1}_{[b,a]}(t)dt & \text{if } a \geqslant b, \end{cases}$$

and define the *primitives* $F(x) := \int_0^x f(t)dt$ and $G(x) := \int_0^x g(t)dt$.

Theorem 14.10 (integration by parts) *Let $f, g : \mathbb{R} \to \mathbb{R}$ be measurable and locally integrable functions and denote by F, G their primitives. The following formula holds for all $-\infty < a < b < \infty$:*

$$F(b)G(b) - F(a)G(a) = \int_a^b f(t)G(t)dt + \int_a^b F(t)g(t)dt. \tag{14.6}$$

Proof Observe that

$$\begin{aligned}
\int_a^b f(t)(G(t) - G(a))dt &= \int_a^b f(t) \left\{ \int_a^t g(s)\, ds \right\} dt \\
&= \int_a^b \int_a^b f(t)g(s)\mathbb{1}_{[a,t]}(s)ds\, dt \\
&\overset{\text{Fubini}}{=} \int_a^b \int_a^b f(t)g(s)\mathbb{1}_{[s,b]}(t)dt\, ds \\
&= \int_a^b \left(\int_s^b f(t)\, dt \right) g(s)ds \\
&= \int_a^b g(s)(F(b) - F(s))ds.
\end{aligned}$$

Rearranging this identity yields (14.6). In the step marked 'Fubini' we use the identity $\mathbb{1}_{[a,t]}(s) = \mathbb{1}_{[s,b]}(t)$ and the fact that

$$\iint_{[a,b]\times[a,b]} |f(t)g(s)|dt\, ds = \int_{[a,b]} |f(t)|dt \cdot \int_{[a,b]} |g(s)|ds < \infty. \qquad \square$$

Fubini's theorem helps us to evaluate some important integrals. Here are two typical examples.

Example 14.11 (error integral, normal distribution)

$$\int_{\mathbb{R}} e^{-x^2/2}\, dx = \sqrt{2\pi}.$$

Proof For $I := \int_{\mathbb{R}} e^{-x^2/2}\, dx$ we find

$$\begin{aligned}
I^2 = \int_{\mathbb{R}} e^{-x^2/2}\, dx \int_{\mathbb{R}} e^{-y^2/2}\, dy &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dy\, dx \\
&= 4 \int_0^{\infty} \int_0^{\infty} e^{-(x^2+y^2)/2}\, dy\, dx
\end{aligned}$$

and we may interpret these integrals as improper Riemann integrals. Changing variables in the inner (Riemann) integral according to $y = tx$, $dy = x\,dt$ yields

$$I^2 = 4\int_0^\infty \int_0^\infty xe^{-x^2(1+t^2)/2}\,dt\,dx = 4\int_0^\infty \int_0^\infty xe^{-x^2(1+t^2)/2}\,dx\,dt.$$

The inner integral has a primitive

$$I^2 = 4\int_0^\infty \left[-\frac{1}{1+t^2}e^{-x^2(1+t^2)/2}\right]_{x=0}^\infty dt = 4\int_0^\infty \frac{dt}{1+t^2} = 4[\arctan t]_0^\infty = 2\pi.$$

□

Example 14.12 (sine integral)

$$\lim_{T\to\infty}\int_0^T \frac{\sin\xi}{\xi}\,d\xi = \frac{\pi}{2}.$$

Proof Note that $1/\xi = \int_0^\infty e^{-t\xi}\,dt$ and $\operatorname{Im} e^{i\xi} = \sin\xi$. Fubini's theorem shows that

$$\begin{aligned}\int_0^T \frac{\sin\xi}{\xi}\,d\xi = \int_0^T\int_0^\infty e^{-t\xi}\sin\xi\,dt\,d\xi &= \int_0^\infty\int_0^T e^{-t\xi}\operatorname{Im} e^{i\xi}\,d\xi\,dt\\ &= \int_0^\infty \operatorname{Im}\int_0^T e^{-(t-i)\xi}\,d\xi\,dt.\end{aligned}$$

The inner integral yields

$$\begin{aligned}\operatorname{Im}\left[\int_0^T e^{-(t-i)\xi}\,d\xi\right] = \operatorname{Im}\left[\frac{e^{-(t-i)\xi}}{i-t}\right]_0^T &= \operatorname{Im}\left[\frac{e^{(i-t)T}-1}{i-t}\right]\\ &= \operatorname{Im}\left[\frac{(e^{(i-t)T}-1)(-i-t)}{1+t^2}\right].\end{aligned}$$

Using dominated convergence (Theorem 12.2) we finally get

$$\begin{aligned}\int_0^T \frac{\sin\xi}{\xi}\,d\xi \;&=\; \int_0^\infty \frac{-te^{-tT}\sin T - e^{-tT}\cos T + 1}{1+t^2}\,dt\\ &\overset{s=tT}{=} \int_0^\infty \frac{-se^{-s}\sin T}{T^2+s^2}\,ds + \int_0^\infty \frac{-e^{-tT}\cos T+1}{1+t^2}\,dt\\ &\xrightarrow[T\to\infty]{\text{dom. conv.}} \int_0^\infty \frac{1}{1+t^2}\,dt = \arctan t\Big|_0^\infty = \frac{\pi}{2}.\end{aligned}$$

□

Distribution Functions

Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space. For $u \in \mathcal{M}(\mathscr{A})$ the decreasing (and, if finite, left-continuous [✍]) function

$$\mathbb{R} \ni t \mapsto \mu\{u \geqslant t\}$$

is called the *distribution function* of u (under μ).

The next theorem shows that Lebesgue integrals still represent the area between the graph of a function and the abscissa.

Theorem 14.13 *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and $u: X \to [0, \infty)$ be $\mathscr{A}$-measurable. Then*

$$\int u \, d\mu = \int_{(0,\infty)} \mu\{u \geqslant t\} \lambda^1(dt) \in [0, \infty]. \tag{14.7}$$

Proof Consider the function $U(x, t) := (u(x), t)$ on $X \times [0, \infty)$. Measurability follows from

$$U^{-1}(A \times I) = u^{-1}(A) \times I \in \mathscr{A} \times \mathscr{B}[0, \infty) \quad \forall A \in \mathscr{A},\ I \in \mathscr{B}[0, \infty)$$

(see also Theorem 14.17(ii)), and, in particular,

$$E := \{(x, t) : u(x) \geqslant t\} \in \mathscr{A} \otimes \mathscr{B}[0, \infty).$$

An application of Tonelli's theorem (Theorem 14.8) shows that

$$\begin{aligned}
\int u(x)\mu(dx) &= \iint \mathbb{1}_{(0,u(x)]}(t)\lambda^1(dt)\mu(dx) \\
&= \iint_{X \times (0,\infty)} \mathbb{1}_E(x, t)\lambda^1(dt)\mu(dx) \\
&= \iint_{(0,\infty) \times X} \mathbb{1}_E(x, t)\mu(dx)\lambda^1(dt) \\
&= \int_{(0,\infty)} \mu\{u \geqslant t\}\lambda^1(dt).
\end{aligned}$$

□

If $\phi:[0,\infty)\to[0,\infty)$ is continuously differentiable, increasing and $\phi(0)=0$, we even have, in the setting of Theorem 14.13,

$$\begin{aligned}
\int \phi\circ u\,d\mu &= \int_{(0,\infty)} \mu\{\phi\circ u\geqslant t\}\lambda^1(dt)\\
&\overset{(*)}{=} \int_0^\infty \mu\{\phi\circ u\geqslant t\}dt\\
&\overset{t=\phi(s)}{=} \int_0^\infty \phi'(s)\mu\{\phi\circ u\geqslant \phi(s)\}ds\\
&= \int_0^\infty \phi'(s)\mu\{u\geqslant s\}ds.
\end{aligned}$$

The problem with this calculation is the step marked $(*)$ where we equate a Lebesgue integral with a Riemann integral. By Theorem 12.9(ii) we can do this if $t\mapsto\mu\{\phi\circ u\geqslant t\}$ is Lebesgue a.e. continuous and bounded. Boundedness is not a problem since we may consider $\kappa\wedge\mu\{\phi\circ u\geqslant t\}\mathbb{1}_{[\kappa^{-1},\kappa]}(t)$, $\kappa\in\mathbb{N}$, and let $\kappa\to\infty$ using Theorem 9.6. For the a.e. continuity we need the following lemma.

Lemma 14.14 *A monotone function $\phi:\mathbb{R}\to\mathbb{R}$ has at most countably many discontinuities and is, in particular, Lebesgue a.e. continuous.*

Proof Without loss of generality we may assume that ϕ increases. Therefore, the one-sided limits $\lim_{s\uparrow t}\phi(s)=:\phi(t-)\leqslant\phi(t+):=\lim_{s\downarrow t}\phi(s)$ exist in $\mathbb{R}$, so that ϕ can only have jump discontinuities where $\phi(t-)<\phi(t+)$. Define for all $\epsilon>0$

$$J^\epsilon:=\{t\in\mathbb{R}:\Delta\phi(t):=\phi(t+)-\phi(t-)\geqslant\epsilon\}.$$

Since on every compact interval $[a,b]$ and for every $\epsilon>0$

$$0\leqslant\phi(b)-\phi(a)=\frac{\phi(b)-\phi(a)}{\epsilon}\epsilon<\infty,$$

we can have at most $(\phi(b)-\phi(a))/\epsilon$ jumps of size ϵ or larger in the interval $[a,b]$, that is $\#([a,b]\cap J^\epsilon)<\infty$. Therefore, the set of all discontinuities of ϕ

$$J:=\{t\in\mathbb{R}:\Delta\phi(t)>0\}=\bigcup_{k,n\in\mathbb{N}}[-n,n]\cap J^{1/k}$$

is a countable set, and hence a Lebesgue null set. $\square$

Since $t\mapsto\mu\{\phi(u)\geqslant t\}$ is decreasing, we finally have the following corollary.

Corollary 14.15 *Let $(X,\mathscr{A},\mu)$ be σ-finite and let $\phi:[0,\infty)\to[0,\infty)$ with $\phi(0)=0$ be increasing and continuously differentiable. Then*

$$\int\phi\circ u\,d\mu=\int_0^\infty\phi'(s)\mu\{u\geqslant s\}ds \tag{14.8}$$

holds for all $u \in \mathcal{M}^+(\mathscr{A})$*; the right-hand side is an improper Riemann integral. Moreover,* $\phi \circ u \in \mathcal{L}^1(\mu)$ *if, and only if, this Riemann integral is finite.*

In the important special case where $\phi(t) = t^p$, $p \geqslant 1$, (14.8) reads

$$\|u\|_p^p = \int |u|^p \, d\mu = \int_0^\infty ps^{p-1}\mu(|u| \geqslant s)ds. \tag{14.9}$$

*Minkowski's Inequality for Integrals

The next inequality is a generalization of Minkowski's inequality (Corollary 13.4) to double integrals. In some sense it is also a theorem on the change of the order of iterated integrals, but equality is obtained only if $p = 1$.

Theorem 14.16 (Minkowski's integral inequality) *Let* $(X, \mathscr{A}, \mu)$ *and* $(Y, \mathscr{B}, \mu)$ *be* σ*-finite measure spaces and* $u : X \times Y \to \overline{\mathbb{R}}$ *be* $\mathscr{A} \otimes \mathscr{B}$*-measurable. Then*

$$\left(\int_X \left(\int_Y |u(x,y)|\nu(dy)\right)^p \mu(dx)\right)^{1/p} \leqslant \int_Y \left(\int_X |u(x,y)|^p \mu(dx)\right)^{1/p} \nu(dy)$$

holds for all $p \in [1, \infty)$*, with equality for* $p = 1$*.*

Proof If $p = 1$, the assertion follows directly from Theorem 14.8 (Tonelli). If $p > 1$ we set

$$U_k(x) := \left(\int_Y |u(x,y)|\nu(dy) \wedge k\right) \mathbb{1}_{A_k}(x),$$

where $A_k \in \mathscr{A}$ is a sequence with $A_k \uparrow X$ and $\mu(A_k) < \infty$. Without loss of generality we may assume that $U_k(x) > 0$ on a set of positive μ-measure, otherwise the left-hand side of the above inequality would be 0 (using Beppo Levi's theorem) and there would be nothing to prove. By Tonelli's theorem and Hölder's inequality with $\frac{1}{p} + \frac{1}{q} = 1$ or $q = \frac{p}{p-1}$, we find

$$\begin{aligned}
\int_X U_k^p(x)\mu(dx) &\leqslant \int_X U_k^{p-1}(x) \left(\int_Y |u(x,y)|\nu(dy)\right) \mu(dx) \\
&= \int_Y \int_X U_k^{p-1}(x)|u(x,y)|\mu(dx)\nu(dy) \\
&\leqslant \int_Y \left(\int_X U_k^p(x)\mu(dx)\right)^{1-1/p} \left(\int_X |u(x,y)|^p \mu(dx)\right)^{1/p} \nu(dy).
\end{aligned}$$

The claim follows upon dividing both sides by $\left(\int_X U_k^p(x)\mu(dx)\right)^{1-1/p}$ and letting $k \to \infty$ with Beppo Levi's theorem. $\square$

More on Measurable Functions

There is an alternative way to introduce the product σ-algebra $\mathscr{A} \otimes \mathscr{B}$. Recall that the coordinate projections

$$\pi_n : X_1 \times X_2 \to X_n, \quad (x_1, x_2) \mapsto x_n, \quad n = 1, 2,$$

induce the σ-algebra $\sigma(\pi_1, \pi_2)$ on $X_1 \times X_2$ which is by Definition 7.5 the smallest σ-algebra such that both π_1 and π_2 are measurable maps.

Theorem 14.17 *Let $(X_n, \mathscr{A}_n)$, $n = 1, 2$, and $(Z, \mathscr{C})$ be measurable spaces.*

(i) *$\mathscr{A}_1 \otimes \mathscr{A}_2 = \sigma(\pi_1, \pi_2)$;*
(ii) *$T : (Z, \mathscr{C}) \to (X_1 \times X_2, \mathscr{A}_1 \otimes \mathscr{A}_2)$ is measurable if, and only if, the maps $\pi_n \circ T : (Z, \mathscr{C}) \to (X_n, \mathscr{A}_n)$ are measurable ($n = 1, 2$);*
(iii) *if $S : (X_1 \times X_2, \mathscr{A}_1 \otimes \mathscr{A}_2) \to (Z, \mathscr{C})$ is measurable, then $S(x_1, \cdot)$ and $S(\cdot, x_2)$ are $\mathscr{A}_2/\mathscr{C}$-, resp. $\mathscr{A}_1/\mathscr{C}$-measurable for every $x_1 \in X_1$, resp. $x_2 \in X_2$.*

Proof (i) Since it holds here that $\pi_1^{-1}(x) = \{x\} \times X_2$, $\pi_2^{-1}(y) = X_1 \times \{y\}$ and $A_1 \times A_2 = (A_1 \times X_2) \cap (X_1 \times A_2)$, we have

$$\sigma(\pi_1, \pi_2) \stackrel{7.5}{=} \sigma(\pi_1^{-1}(\mathscr{A}_1), \pi_2^{-1}(\mathscr{A}_2)) = \sigma\left(\{A_1 \times X_2, X_1 \times A_2 : A_n \in \mathscr{A}_n\}\right),$$

which shows that $\mathscr{A}_1 \times \mathscr{A}_2 \subset \sigma(\pi_1, \pi_2) \subset \mathscr{A}_1 \otimes \mathscr{A}_2$, hence $\sigma(\pi_1, \pi_2) = \mathscr{A}_1 \otimes \mathscr{A}_2$.

(ii) If $T : Z \to X_1 \times X_2$ is measurable, then so is $\pi_n \circ T$ by part (i) and Theorem 7.4. Conversely, if $\pi_n \circ T$, $n = 1, 2$, are measurable we find

$$\begin{aligned} T^{-1}(A_1 \times A_2) &= T^{-1}\left(\pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2)\right) \\ &= T^{-1}\left(\pi_1^{-1}(A_1)\right) \cap T^{-1}\left(\pi_2^{-1}(A_2)\right) \\ &= (\pi_1 \circ T)^{-1}(A_1) \cap (\pi_2 \circ T)^{-1}(A_2) \in \mathscr{C}. \end{aligned}$$

Since $\mathscr{A}_1 \times \mathscr{A}_2$ generates $\mathscr{A}_1 \otimes \mathscr{A}_2$, T is measurable by Lemma 7.2.

(iii) Fix $x_1 \in X_1$ and consider $y \mapsto S(x_1, y)$. Then $S(x_1, \cdot) = S \circ i_{x_1}(\cdot)$, where $i_{x_1} : X_2 \to X_1 \times X_2$, $y \mapsto (x_1, y)$. By part (ii), i_{x_1} is $\mathscr{A}_2/\mathscr{A}_1 \otimes \mathscr{A}_2$-measurable since the maps $\pi_n \circ i_{x_1}(x_2) = x_n$ are $\mathscr{A}_2/\mathscr{A}_n$-measurable ($n = 1, 2$). The claim follows now from Theorem 7.4. □

Problems

14.1. Prove the rules (14.2) for Cartesian products.

14.2. Let $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ be two σ-finite measure spaces. Show that $A \times N$, where $A \in \mathscr{A}$ and $N \in \mathscr{B}$, $\nu(N) = 0$, is a $\mu \times \nu$-null set.

14.3. Let $(X_i, \mathscr{A}_i, \mu_i)$, $i = 1, 2$, be σ-finite measure spaces and $f\colon X_1 \times X_2 \to \mathbb{C}$ a measurable function. A function is *negligible* (*w.r.t. the measure* μ) if $\int |f|\,d\mu = 0$. Show that the following assertions are equivalent.

(a) f is $\mu_1 \times \mu_2$-negligible.
(b) For μ_1-almost all x_1 the function $f(x_1, \cdot)$ is μ_2-negligible.
(c) For μ_2-almost all x_2 the function $f(\cdot, x_2)$ is μ_1-negligible.

14.4. Denote by λ Lebesgue measure on $(0,\infty)$. Prove that the following iterated integrals exist and that

$$\int_{(0,\infty)} \int_{(0,\infty)} e^{-xy} \sin x \sin y \, \lambda(dx)\lambda(dy) = \int_{(0,\infty)} \int_{(0,\infty)} e^{-xy} \sin x \sin y \, \lambda(dy)\lambda(dx).$$

Does this imply that the double integral exists?

14.5. Denote by λ Lebesgue measure on $(0,1)$. Show that the following iterated integrals exist, but yield different values:

$$\int_{(0,1)} \int_{(0,1)} \frac{x^2 - y^2}{(x^2 + y^2)^2} \, \lambda(dx)\lambda(dy) \neq \int_{(0,1)} \int_{(0,1)} \frac{x^2 - y^2}{(x^2 + y^2)^2} \, \lambda(dy)\lambda(dx).$$

What does this tell us about the double integral?

14.6. Denote by λ Lebesgue measure on $(-1,1)$. Show that the iterated integrals exist and coincide,

$$\int_{(-1,1)} \int_{(-1,1)} \frac{xy}{(x^2 + y^2)^2} \, \lambda(dx)\lambda(dy) = \int_{(-1,1)} \int_{(-1,1)} \frac{xy}{(x^2 + y^2)^2} \, \lambda(dy)\lambda(dx),$$

but that the double integral does not exist.

14.7. Evaluate $\int_0^1 \int_0^1 f(x,y)\,dx\,dy$, $\int_0^1 \int_0^1 f(x,y)\,dy\,dx$ and $\int_{[0,1]^2} |f(x,y)|\,d(x,y)$ if

$$\text{(a)} \quad \left(x - \tfrac{1}{2}\right)^{-3} \mathbb{1}_{\{0 < y < |x - \frac{1}{2}|\}}; \quad \text{(b)} \quad \frac{x - y}{(x^2 + y^2)^{3/2}}; \quad \text{(c)} \quad \frac{1}{(1 - xy)^p}, \; p > 0.$$

14.8. Consider the measure space $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and denote by $\zeta_M : \mathscr{B}(\mathbb{R}) \to [0,\infty]$, $M \subset \mathbb{R}$, $\zeta_M(A) := \#(A \cap M)$ the counting measure.

(i) Show that Lebesgue's measure λ and the counting measure $\zeta_{\mathbb{Q}}$ are σ-finite, the counting measure $\zeta_{\mathbb{R}}$ is not σ-finite.
(ii) Show that $\int_{(0,1)} \int_{(0,1)} \mathbb{1}_{\mathbb{Q}}(x \cdot y)\,d\lambda(x)\,d\zeta_{\mathbb{R}}(y) = 0$.
(iii) Show that $\int_{(0,1)} \int_{(0,1)} \mathbb{1}_{\mathbb{Q}}(x \cdot y)\,d\zeta_{\mathbb{R}}(y)\,d\lambda(x) = \infty$.
(iv) Show that $\int_{(0,1)} \int_{(0,1)} \mathbb{1}_{\mathbb{Q}}(x \cdot y)\,d\zeta_{\mathbb{Q}}(y)\,d\lambda(x) = 0$.
(v) Do (ii) and (iii) contradict Fubini's theorem?

14.9. (i) Evaluate

$$\int_{[0,\infty)^2} \frac{dx\,dy}{(1 + y)(1 + x^2 y)}.$$

(ii) Use (i), in order to evaluate

$$\int_0^\infty \frac{\ln x}{x^2 - 1}\,dx.$$

(iii) Use a series representation in part (ii) to show that

$$\sum_{n=0}^\infty \frac{1}{(2n + 1)^2} = \frac{\pi^2}{8}.$$

14.10. Let μ, ν be σ-finite measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Show that

(i) The set $D := \{x \in \mathbb{R} : \mu\{x\} > 0\}$ is at most countable.
(ii) The diagonal $\Delta \subset \mathbb{R}^2$ has measure $\mu \times \nu(\Delta) = \sum_{x \in D} \mu\{x\}\nu\{x\}$.

14.11. Let $\mu(A) := \#A$ be the counting measure and λ be Lebesgue measure on the measurable space $([0,1], \mathscr{B}[0,1])$. Denote by $\Delta := \{(x,y) \in [0,1]^2 : x = y\}$ the diagonal in $[0,1]^2$. Check that

$$\int_{[0,1]} \int_{[0,1]} \mathbb{1}_\Delta(x,y)\lambda(dx)\mu(dy) \neq \int_{[0,1]} \int_{[0,1]} \mathbb{1}_\Delta(x,y)\mu(dy)\lambda(dx).$$

Does this contradict Tonelli's theorem?

14.12. (i) State Tonelli's and Fubini's theorems for spaces of sequences, i.e. for the measure space $(\mathbb{N}, \mathscr{P}(\mathbb{N}), \mu)$, where $\mu := \sum_{n \in \mathbb{N}} \delta_n$, and obtain criteria specifying when one can interchange two infinite summations.

(ii) Using similar considerations to those in part (i) deduce the following.

Lemma. *Let $(A_n)_n$ be countably many mutually disjoint sets whose union is $\mathbb{N}$, and let $(x_k)_{k \in \mathbb{N}} \subset \mathbb{R}$ be a sequence. Then*

$$\sum_{k \in \mathbb{N}} x_k = \sum_n \sum_{k \in A_n} x_k$$

in the sense that, if either side converges absolutely, so does the other, in which case the two sides are equal.

14.13. Let $u : \mathbb{R}^2 \to [0, \infty]$ be a measurable function. $S[u] := \{(x,y) : 0 \leqslant y \leqslant u(x)\}$ is the set above the abscissa and below the graph $\Gamma[u] := \{(x, u(x)) : x \in \mathbb{R}\}$.

(i) Show that $S[u] \in \mathscr{B}(\mathbb{R}^2)$.
(ii) Is it true that $\lambda^2(S[u]) = \int u \, d\lambda^1$?
(iii) Show that $\Gamma[u] \in \mathscr{B}(\mathbb{R}^2)$ and that $\lambda^2(\Gamma[u]) = 0$.

[Hint: (i) use Theorem 8.8 to approximate u by simple functions $f_n \uparrow u$. Thus we have $S[u] = \bigcup_n S[f_n]$ and it is easy to see that $S[f_n] \in \mathscr{B}(\mathbb{R}^2)$; alternatively, use Theorem 14.17, set $U(x,y) := (u(x), y)$ and observe that $S[u] = U^{-1}(C)$ for the closed set $C := \{(x,y) : x \geqslant y\}$; (ii), (iii) use Tonelli's theorem.]

14.14. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and let $u \in \mathcal{M}^+(\mathscr{A})$ be a $[0,\infty]$-valued measurable function. Show that the set

$$Y := \big\{y \in \mathbb{R} : \mu\{x : u(x) = y\} \neq 0\big\} \subset \mathbb{R}$$

is countable.

[Hint: assume that $u \in \mathcal{L}^1_+(\mu)$. Set $Y_{\epsilon,\eta} := \{y > \eta : \mu\{u = y\} > \epsilon\}$ and observe that for $t_1, \ldots, t_N \in Y_{\epsilon,\eta}$ we have $N\epsilon\eta \leqslant \sum_{n=1}^N t_n \mu\{u = t_n\} \leqslant \int u \, d\mu$. Thus $Y_{\epsilon,\eta}$ is a finite set, and $Y = \bigcup_{k,n \in \mathbb{N}} Y_{\frac{1}{n}, \frac{1}{k}}$ is countable. If u is not integrable, consider $(u \wedge m)\mathbb{1}_{A_m}$, $m \in \mathbb{N}$, where $A_m \uparrow X$ is an exhaustion.]

14.15. **Completion (5).** Let $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ be any two measure spaces such that $\mathscr{A} \neq \mathscr{P}(X)$ and such that $\mathscr{B}$ contains non-empty null sets.

(i) Show that $\mu \times \nu$ on $(X \times Y, \mathscr{A} \otimes \mathscr{B})$ is not complete, even if both μ and ν are complete.
(ii) Conclude from (i) that neither $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}), \lambda \times \lambda)$ nor the product of the completed spaces $(\mathbb{R}^2, \mathscr{B}^*(\mathbb{R}) \otimes \mathscr{B}^*(\mathbb{R}), \bar{\lambda} \times \bar{\lambda})$ is complete.

[Hint: you may assume in (ii) that $\mathscr{B}(\mathbb{R}), \mathscr{B}^*(\mathbb{R}) \neq \mathscr{P}(\mathbb{R})$, see Appendix G.]

14.16. Let μ be a bounded measure on the measure space $([0,\infty), \mathscr{B}[0,\infty))$.

(i) Show that $A \in \mathscr{B}[0,\infty) \otimes \mathscr{P}(\mathbb{N})$ if, and only if, $A = \bigcup_{n\in\mathbb{N}} B_n \times \{n\}$, where $(B_n)_{n\in\mathbb{N}} \subset \mathscr{B}[0,\infty)$.

(ii) Show that there exists a unique measure π on $\mathscr{B}[0,\infty) \otimes \mathscr{P}(\mathbb{N})$ satisfying

$$\pi(B \times \{n\}) = \int_B e^{-t} \frac{t^n}{n!}\, \mu(dt).$$

14.17. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space, i.e. a measure space such that $\mathbb{P}(\Omega) = 1$. Show that for a measurable function $T: \Omega \to [0,\infty)$ and every $\lambda > 0$ the following formula holds:

$$\int e^{-\lambda T}\, d\mathbb{P} = 1 - \lambda \int_0^\infty e^{-\lambda s} \mathbb{P}(T \geqslant s) ds.$$

What happens if we also allow negative values of λ?

14.18. **Stieltjes measure (2). Stieltjes integrals.** This continues Problem 6.1. Let μ and ν be two measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\mu((-n,n]), \nu((-n,n]) < \infty$ for all $n \in \mathbb{N}$, and denote by

$$F(x) := \begin{cases} \mu(0,x], & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -\mu(x,0], & \text{if } x < 0, \end{cases} \quad \text{and} \quad G(x) := \begin{cases} \nu(0,x], & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -\nu(x,0], & \text{if } x < 0, \end{cases}$$

the associated *right-continuous* distribution functions (in Problem 6.1 we considered *left-continuous* distribution functions). Moreover, set $\Delta F(x) = F(x) - F(x-)$ and $\Delta G(x) = G(x) - G(x-)$.

(i) Show that F, G are increasing and right-continuous, and that $\Delta F(x) = 0$ if, and only if, $\mu\{x\} = 0$. Moreover, F and μ are in one-to-one correspondence.

(ii) Since measures and distribution functions are in one-to-one correspondence, it is customary to write $\int u\, d\mu = \int u\, dF$, etc.
If $a < b$ we set $B := \{(x,y) : a < x \leqslant b,\ x \leqslant y \leqslant b\}$. Show that B is measurable and that

$$\mu \times \nu(B) = \int_{(a,b]} F(s) dG(s) - F(a)(G(b) - G(a)).$$

(iii) **Integration by parts.** Show that

$$\begin{aligned} &F(b)G(b) - F(a)G(a) \\ &\quad = \int_{(a,b]} F(s) dG(s) + \int_{(a,b]} G(s-) dF(s) \\ &\quad = \int_{(a,b]} F(s-) dG(s) + \int_{(a,b]} G(s-) dF(s) + \sum_{a<s\leqslant b} \Delta F(s) \Delta G(s). \end{aligned}$$

[Hint: expand $\mu \times \nu((a,b]^2)$ in two different ways, using (ii). Note that the series appearing in the formula is at most countable because of Lemma 14.14.]

(iv) **Change of variable formula.** Let ϕ be a C^1-function. Then

$$\begin{aligned} &\phi(F(b)) - \phi(F(a)) \\ &= \int_{(a,b]} \phi'(F(s-)) dF(s) + \sum_{a<s\leqslant b} \Big[\phi(F(s)) - \phi(F(s-)) - \phi'(F(s-)) \Delta F(s)\Big]. \end{aligned}$$

[Hint: use (iii) to show the change-of-variable formula for polynomials and then use the fact that continuous functions can be uniformly approximated by a sequence of polynomials – see Weierstraß' approximation theorem, Theorem 28.6.]

14.19. Rearrangements. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and let $f \in \mathcal{L}^p(\mu)$ for some $p \in [1, \infty)$. The distribution function of f is given by $\mu_f(t) = \mu\{|f| \geqslant t\}$ and the *decreasing rearrangement* of f is the generalized inverse of μ_f,

$$f^*(\xi) := \inf\{t : \mu_f(t) \leqslant \xi\}, \quad \xi \geqslant 0, \qquad (\inf \emptyset = +\infty).$$

(i) Let $f = 2\mathbb{1}_{[1,3]} + 4\mathbb{1}_{[4,5]} + 3\mathbb{1}_{[6,9]}$. Draw the graphs of $f(x)$, $\mu_f(t)$ and $f^*(\xi)$.
(ii) Show that for $f \in \mathcal{L}^p(\mu)$

$$\int_{\mathbb{R}} |f|^p \, d\mu = p \int_0^\infty t^{p-1} \mu_f(t) dt = \int_{(0,\infty)} (f^*)^p \, d\lambda.$$

In other words, $\|f\|_p = \|f^*\|_p$. Because of this the space $\mathcal{L}^p$ is said to be *rearrangement invariant*.

14.20. The differentiability lemma revisited. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and $\phi : \mathbb{R} \times X \to \mathbb{R}$ a mapping with the following properties.

(i) $\int_X |\phi(t,x)| \mu(dx) < \infty$ for all $t \in \mathbb{R}$,
(ii) $t \mapsto \phi(t,x)$ is differentiable for all $x \in X$, $\int_X |\partial_t \phi(t,x)| \mu(dx) < \infty$ and

$$\int_K \int_X |\partial_t \phi(t,x)| \mu(dx) dt < \infty$$

for any compact set $K \subset \mathbb{R}$
(iii) $t \mapsto \int_X \partial_t \phi(t,x) \mu(dx)$ is continuous.

Then show that $F(t) := \int_X \phi(t,x) \mu(dx)$ is continuously differentiable for all $t \in \mathbb{R}$ and

$$F'(t) = \int_X \partial_t \phi(t,x) \mu(dx), \quad t \in \mathbb{R}.$$

Remark. I owe this variant of Theorem 12.5 to Franziska Kühn. The paper by Talvila [55] contains further versions of this theorem.

15

Integrals with Respect to Image Measures

Let $(X, \mathscr{A}, \mu)$ be a measure space and $(X', \mathscr{A}')$ be a measurable space. We have seen in Theorem 7.6 that any measurable map $T:(X, \mathscr{A}) \to (X', \mathscr{A}')$ can be used to transport the measure μ, defined on $(X, \mathscr{A})$, to a measure on $(X', \mathscr{A}')$:

$$T(\mu)(A') := \mu(T^{-1}(A')) \quad \forall A' \in \mathscr{A}'. \tag{15.1}$$

Using the general method to extend measures to integrals, see Fig. 10.1 on page 82, we can define the integral $\int \dots dT(\mu)$ w.r.t. the image measure $T(\mu)$. This is the idea behind the proof of the following theorem.

Theorem 15.1 (transformation theorem) *Let $T: X \to X'$ be a measurable map between the measure space $(X, \mathscr{A}, \mu)$ and the measurable space $(X', \mathscr{A}')$. The $T(\mu)$-integral of a measurable function $u: (X', \mathscr{A}') \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ is given by*

$$\int_{X'} u \, dT(\mu) = \int_X u \circ T \, d\mu. \tag{15.2}$$

Formula (15.2) *is understood in the following way: either $u \geqslant 0$ and* (15.2) *holds in $[0, \infty]$, or $u \in \mathcal{L}^1(T(\mu))$ if, and only if, $u \circ T \in \mathcal{L}^1(\mu)$.*

Proof Since u and T are measurable, so is $u \circ T$, see Theorem 7.4, and the integrals in (15.2) are well-defined. In order to prove (15.2), we mimic the construction of the integral starting from a measure, see Fig. 10.1 on page 82.

Step 1. Assume that $u = \mathbb{1}_{A'}$ for some $A' \in \mathscr{A}'$. Since $\mathbb{1}_{T^{-1}(A')} = \mathbb{1}_{A'}(T)$, we get

$$\underbrace{\int_{X'} \mathbb{1}_{A'} \, dT(\mu)}_{= \int_{X'} u \, dT(\mu)} = T(\mu)(A') = \mu(T^{-1}(A')) = \int_X \mathbb{1}_{T^{-1}(A')} d\mu = \underbrace{\int_X \mathbb{1}_{A'}(T) d\mu}_{= \int_X u(T) d\mu}.$$

This proves (15.2) for $u = \mathbb{1}_{A'}$.

Step 2. For $u = \sum_{n=0}^{N} \alpha_n \mathbb{1}_{A'_n} \in \mathcal{E}(\mathscr{A}')$, $\alpha_n \geqslant 0$, we get (15.2) using Step 1 and the linearity of the integral.

Step 3. If $u \in \mathcal{M}(\mathscr{A}')$, $u \geqslant 0$, then we use the sombrero lemma (Theorem 8.8) to construct a sequence $u_n \in \mathcal{E}(\mathscr{A}')$, $u_n \geqslant 0$, with $u_n \uparrow u$. Applying Beppo Levi's theorem twice yields

$$\int_{X'} u \, dT(\mu) = \sup_{n\in\mathbb{N}} \int_{X'} u_n \, dT(\mu) \overset{\text{Step 2}}{=} \sup_{n\in\mathbb{N}} \int_X u_n(T) d\mu = \int_X u(T) d\mu,$$

i.e. (15.2) holds for positive measurable functions. Moreover, our arguments show that $\int_{X'} u \, dT(\mu) < \infty$ if, and only if, $\int_X u(T) d\mu < \infty$.

Step 4: Finally, let $u \in \mathcal{M}(\mathscr{A}')$. By definition

$$\begin{aligned} u \in \mathcal{L}^1(T(\mu)) &\overset{\text{def}}{\iff} \int u^{\pm} \, dT(\mu) < \infty \\ &\overset{\text{Step 3}}{\iff} \int u^{\pm}(T) d\mu < \infty \overset{\text{def}}{\iff} u(T) \in \mathcal{L}^1(\mu). \end{aligned}$$

□

In probability theory, Theorem 15.1 is of paramount importance since it gives the link between random variables and their distributions. This is illustrated in the next example.

Example 15.2 Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and $\xi : \Omega \to \mathbb{R}$ be a random variable (i.e. a measurable function). The image measure $\mathbb{P}_\xi(B) := \mathbb{P} \circ \xi^{-1}(B) = \mathbb{P}(\xi \in B)$, $B \in \mathscr{B}(\mathbb{R})$, is the distribution of the random variable, see Example 7.8(ii). If $\xi \in \mathcal{L}^1(\mathbb{P})$ then we call

$$\mathbb{E}\xi := \int \xi \, d\mathbb{P} = \int_\Omega \xi(\omega) \mathbb{P}(d\omega)$$

the *expected value* of the random variable ξ.

(i) $\xi \in \mathcal{L}^1(\mathbb{P})$ if, and only if, $\mathbb{E}|\xi| = \int_{\mathbb{R}} |x| \mathbb{P}(\xi \in dx) < \infty$.
(ii) If $\xi \in \mathcal{L}^1(\mathbb{P})$, then $\mathbb{E}\xi = \int_{\mathbb{R}} x \mathbb{P}(\xi \in dx)$.
(iii) If $\xi \in \mathcal{L}^2(\mathbb{P})$, then $\mathbb{V}\xi := \mathbb{E}\left[(\xi - \mathbb{E}\xi)^2\right] = \int_{\mathbb{R}} (x - \mathbb{E}\xi)^2 \mathbb{P}(\xi \in dx)$. $\mathbb{V}\xi$ is called the *variance* of the random variable ξ.
(iv) If ξ has a density $f(x)$, i.e. if $\mathbb{P}_\xi(dx) = f(x)dx$ for some measurable $f \geqslant 0$, then

$$\mathbb{E}u(\xi) = \int_{\mathbb{R}} u(x) \mathbb{P}(\xi \in dx) = \int_{\mathbb{R}} u(x) f(x) dx \quad \forall u \in \mathcal{L}^1(f\lambda^1).$$

(v) If ξ is a discrete random variable, i.e. $\mathbb{P}_\xi = \sum_{n=1}^{N} p_n \delta_{x_n}$, then

$$\mathbb{E}u(\xi) = \int_{\mathbb{R}} u(y) \mathbb{P}(\xi \in dy) = \sum_{n=1}^{N} p_n \, u(x_n).$$

The following examples are important for the following chapters.

Example 15.3 Let $(X, \mathscr{A}, \mu)$ be $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \lambda)$, where $\lambda = \lambda^n$ is n-dimensional Lebesgue measure and let $(X', \mathscr{A}') = (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. The maps

$$\sigma : \mathbb{R}^n \to \mathbb{R}^n,\ y \mapsto -y \quad \text{and} \quad \tau_x : \mathbb{R}^n \to \mathbb{R}^n,\ y \mapsto y - x$$

are continuous, hence measurable, and so are their inverses, namely $\sigma^{-1} = \sigma$ and $\tau_x^{-1} = \tau_{-x}$. By Corollary 7.11, Lebesgue measure λ is invariant under reflections and translations, so that $\lambda = \sigma(\lambda)$ and $\lambda = \tau_x(\lambda)$ for all $x \in \mathbb{R}^n$.

(i) These invariance properties have the following analogues for integrals:

$$\int u(-y)\lambda(dy) = \int u(\sigma y)\lambda(dy) = \int u(y)\sigma(\lambda)(dy) = \int u(y)\lambda(dy), \tag{15.3}$$

and, for all $x \in \mathbb{R}^n$,

$$\int u(y \pm x)\lambda(dy) = \int u(\tau_{\mp x} y)\lambda(dy) = \int u(y)\tau_{\mp x}(\lambda)(dy) = \int u(y)\lambda(dy), \tag{15.4}$$

where $u : \mathbb{R}^n \to \mathbb{R}$ is a measurable function, and if one side of (15.3), (15.4) is finite/integrable so is the other.

(ii) If we consider Lebesgue measure with a density $f \geqslant 0, f\lambda$, see Lemma 10.8, we find

$$\begin{aligned} \int u(y)\tau_x(f\lambda)(dy) &= \int u(\tau_x y) f(y)\lambda(dy) \\ &= \int u(\tau_x y) f(\tau_x \tau_{-x} y)\lambda(dy) \\ &= \int u(y) f(y + x)\lambda(dy), \end{aligned} \tag{15.5}$$

which also proves that $\tau_x(f\lambda) = (f \circ \tau_{-x})\lambda$.

(iii) Let $A \in \mathbb{R}^{n \times n}$ be an invertible matrix. The linear map $x \mapsto Ax$ is continuous, hence measurable, and we have seen in Theorem 7.10 that $A(\lambda) = |\det A|^{-1}\lambda$. Consequently, we find for all $b \in \mathbb{R}^n$

$$\int u(Ax + b)\lambda(dx) = |\det A|^{-1} \int u(y)\lambda(dy) \quad \forall u \in \mathcal{L}^1(\lambda). \tag{15.6}$$

Convolutions

The *convolution* or *Faltung* of functions and measures on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ appears naturally in functional analysis, Fourier analysis, probability theory and other branches of mathematics. One can understand it as an averaging process that respects translations and results in a gain of smoothness.

Definition 15.4 Let μ and ν be measures on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and $u, v: \mathbb{R}^n \to \overline{\mathbb{R}}$ measurable functions. If the expressions

(i) $u \star v(x) := \int_{\mathbb{R}^n} u(x-y)v(y)\lambda^n(dy)$,

(ii) $u \star \mu(x) := \int_{\mathbb{R}^n} u(x-y)\mu(dy)$,

(iii) $\mu \star \nu(B) := \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \mathbb{1}_B(x+y)\mu(dx)\nu(dy)$, $B \in \mathscr{B}(\mathbb{R}^n)$

make sense, they are called *convolutions* (of two functions, of a measure with a function or of two measures, respectively).

Caution The key issue in Definition 15.4 is the existence of the integrals. Note that the convolution of two measures is always defined and it is a measure. The other two convolutions are functions; they are always defined (in $[0,\infty]$) if $u, v \geqslant 0$. Good finiteness conditions will be given below.

Properties 15.5 (of the convolution) Let μ and ν be measures on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, $u, v: \mathbb{R}^n \to \overline{\mathbb{R}}$ measurable functions.

(i) $\mu \star \nu = \alpha(\mu \times \nu)$, where $\alpha: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, $(x,y) \mapsto \alpha(x,y) := x+y$.
Indeed: α is continuous, hence measurable, and we get for $B \in \mathscr{B}(\mathbb{R}^n)$

$$\begin{aligned}\mu \star \nu(B) &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \mathbb{1}_B(x+y)\mu(dx)\nu(dy) \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbb{1}_B(\alpha(x,y))(\mu \times \nu)(d(x,y)) \\ &= \int_{\mathbb{R}^n} \mathbb{1}_B(z)\alpha(\mu \times \nu)(dz).\end{aligned}$$

(ii) $\mu \star \nu = \nu \star \mu$. This follows from (i) and the symmetry of $\alpha(x,y) = \alpha(y,x)$.

(iii) $\mu \star \nu(B) = \int_{\mathbb{R}^n} \mu(B-y)\nu(dy) = \int_{\mathbb{R}^n} \nu(B-x)\mu(dx)$ for all $B \in \mathscr{B}(\mathbb{R}^n)$.
Indeed: Note that $\mathbb{1}_B(x+y) = \mathbb{1}_{B-y}(x)$ as $x+y \in B \Leftrightarrow x \in B-y$. Therefore,

$$\begin{aligned}\mu \star \nu(B) &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \mathbb{1}_B(x+y)\mu(dx)\nu(dy)\\ &= \int_{\mathbb{R}^n}\left\{\int_{\mathbb{R}^n} \mathbb{1}_{B-y}(x)\mu(dx)\right\}\nu(dy)\\ &= \int_{\mathbb{R}^n} \mu(B-y)\nu(dy).\end{aligned}$$

The second formula follows in a similar way since $\mu \star \nu = \nu \star \mu$.

(iv) $\int u(z)\mu \star \nu(dz) = \iint u(x+y)\mu(dy)\nu(dx)$ and $u \in \mathcal{L}^1(\mu \star \nu)$ if, and only if $u(x+y)$ is integrable for $\mu \times \nu(dx, dy)$. This is an immediate consequence of Theorem 15.1 and the first part (i).

(v) If $u, v \geqslant 0$ are measurable functions, then $(u\lambda^n) \star (v\lambda^n) = (u \star v)\lambda^n$.
Indeed:

$$\begin{aligned}(u\lambda^n) \star (v\lambda^n)(B) &= \iint \mathbb{1}_B(x+y)u(x)v(y)\lambda^n(dx)\lambda^n(dy)\\ &= \int_B \left(\int u(x-y)v(y)\lambda^n(dy)\right)\lambda^n(dx),\end{aligned}$$

where we use Tonelli's theorem (Theorem 14.8) and (15.5). In a similar way one shows that $(u\lambda^n) \star \mu = (u \star \mu)\lambda^n$.

(vi) The convolution of two functions (or of a function with a measure or of two measures) is linear in each argument, e.g.

$$(\alpha u + \beta v) \star w = \alpha(u \star w) + \beta(v \star w), \quad \alpha, \beta \in \mathbb{R}.$$

Similar formulae hold for the second argument and the other cases.

(vii) $u \star v = v \star u$ for all measurable $u, v : \mathbb{R}^n \to \mathbb{R}$ such any one of the convolutions exist. This follows from (v), (ii) and by linearity:

$$u \star v = (u^+ - u^-) \star (v^+ - v^-) = u^+ \star v^+ + u^- \star v^- - u^- \star v^+ - u^+ \star v^-.$$

Let us now tend to the problem when the convolution $u \star v$ exists. Here is a very useful (sufficient) criterion due to W. H. Young.

Theorem 15.6 (Young's inequality) *Let $u \in \mathcal{L}^1(\lambda^n)$ and $v \in \mathcal{L}^p(\lambda^n)$, $p \in [1, \infty)$. The convolution $u \star v$ defines a function in $\mathcal{L}^p(\lambda^n)$ and satisfies $u \star v = v \star u$, and*

$$\|u \star v\|_p \leqslant \|u\|_1 \cdot \|v\|_p. \tag{15.7}$$

Proof We may safely assume that $u, v \geqslant 0$, since we have

$$|u \star v(x)| = \left| \int u(x-y)v(y)\lambda^n(dy) \right| \leqslant \int |u(x-y)||v(y)|\lambda^n(dy) = |u| \star |v|(x)$$

and $\|u\|_p = \||u|\|_p$. Using that λ^n is invariant under translations, we see with Jensen's inequality, Tonelli's theorem and (15.3)–(15.5) that

$$\begin{aligned}
\|u \star v\|_p^p &= \int \left(\int u(y)v(x-y)\lambda^n(dy) \right)^p \lambda^n(dx) \\
&= \|u\|_1^p \int \left(\int v(x-y) \frac{u(y)}{\|u\|_1} \lambda^n(dy) \right)^p \lambda^n(dx) \\
&\underset{\text{Jensen}}{\overset{13.13}{\leqslant}} \|u\|_1^p \iint v(x-y)^p \frac{u(y)}{\|u\|_1} \lambda^n(dy)\lambda^n(dx) \\
&\underset{\text{Tonelli}}{\overset{14.8}{=}} \|u\|_1^p \int \underbrace{\left(\int v(x-y)^p \lambda^n(dx) \right)}_{=\|v\|_p^p \text{ by (15.4)}} \frac{u(y)}{\|u\|_1} \lambda^n(dy) \\
&= \|u\|_1^p \|v\|_p^p,
\end{aligned}$$

which gives (15.7) and implies that $u \star v \in \mathcal{L}^p(\lambda^n)$. □

The convolution $u \star v$ is a hybrid of u and v which inherits those properties which are preserved under translations and averages, see Problems 15.10–15.12. In general, $u \star v$ is smoother than u and v. To see this, we need the following result which exploits the interplay of measure and topology. Its proof is postponed to Chapter 17, Theorem 17.8.

Lemma 15.7 *The continuous functions with compact support $C_c(\mathbb{R}^n)$ are a dense subset of $\mathcal{L}^p(\lambda^n)$, $p \in [1, \infty)$.*

Theorem 15.8 *Let $u \in \mathcal{L}^p(\lambda^n)$, $p \in [1, \infty)$.*

(i) *The map $x \mapsto \displaystyle\int |u(x+y) - u(y)|^p \lambda^n(dy)$ is continuous.*

(ii) *If $u \in \mathcal{L}^1(\lambda^n)$, $v \in \mathcal{L}^\infty(\lambda^n)$, then $u \star v$ is bounded and continuous.*

Proof (i) Let $\phi \in C_c(\mathbb{R}^n)$ and assume that $\operatorname{supp}\phi = \overline{\{\phi \neq 0\}} \subset B_R(0)$. From the continuity lemma (Theorem 12.4) we infer that

$$x \mapsto I(\phi; x) := \int |\phi(x+y) - \phi(y)|^p \lambda^n(dy) = \|\phi(x + \cdot\,) - \phi(\cdot)\|_p^p$$

is continuous. (Use $|\phi(x + \cdot\,) - \phi|^p \leqslant 2^p \|\phi\|_\infty^p \mathbb{1}_{B_{R+1}(0)}$ as a majorant if $|x| < 1$.)

Let $u \in \mathcal{L}^p(\lambda^n)$. Using the density of $C_c(\mathbb{R}^n)$ in $\mathcal{L}^p(\lambda^n)$, we find a sequence $(\phi_n)_{n\in\mathbb{N}} \subset C_c(\mathbb{R}^n)$ such that $\lim_{n\to\infty} \|u - \phi_n\|_p = 0$. Since we have any $x \in \mathbb{R}^n$,

$$\begin{aligned} &\|(\phi_n(x+\,\cdot\,) - \phi_n(\cdot)) - (u(x+\,\cdot\,) - u(\cdot))\|_p \\ &\qquad \leqslant \|\phi_n(x+\,\cdot\,) - u(x+\,\cdot\,)\|_p + \|\phi_n - u\|_p \overset{(15.4)}{=} 2\|\phi_n - u\|_p, \end{aligned}$$

we see that $I(\phi_n; x)$ converges to $I(u; x)$ uniformly. In particular, $I(u; x)$ inherits the continuity of the $I(\phi_n; x)$.

(ii) For all $x, x' \in \mathbb{R}^n$ we see that

$$\begin{aligned} \left|u \star v(x) - u \star v(x')\right| &= \int \left|v(y)u(x-y) - v(y)u(x'-y)\right| \lambda^n(dy) \\ &= \|v\|_\infty \|u(x-\,\cdot\,) - u(x'-\,\cdot\,)\|_1 \\ &\overset{(15.4)}{=} \|v\|_\infty \|u(x - x' +\,\cdot\,) - u\|_1 \end{aligned}$$

and continuity follows from part (i). The boundedness of $u \star v$ is proved with a similar calculation. □

✻Regularization

We will now show how we can use the convolution to smooth out and approximate functions in $\mathcal{L}^p(\mu)$. We begin with a result from real analysis which is not as obvious as one might think. As usual, we write $C^\infty(\mathbb{R}^n)$ for the class of arbitrarily often differentiable functions on $\mathbb{R}^n$.

Lemma 15.9 *There is a function $\phi \in C^\infty(\mathbb{R}^n)$ such that $\phi \geqslant 0$,* $\operatorname{supp}\phi \subset \overline{B_1(0)}$ *and $\int \phi \, d\lambda^n = 1$.*

Proof Set $\phi(x) := \kappa\psi(|x|^2)$, where $\psi(r) = \exp\left[1/(r^2-1)\right] \mathbb{1}_{(-1,1)}(r)$. Since $\psi(|x|^2)$ is bounded by $\mathbb{1}_{B_1(0)}(x)$, the integral $\kappa^{-1} := \int_{|x|<1} \psi(|x|^2)dx$ is finite, and we need only show that $\psi(r)$ is smooth. On $[-1,1]^c$, there is nothing to show, for symmetry reasons it suffices to consider the point $r = 1$.

Since $\lim_{\epsilon\to 0} \psi(1-\epsilon) = 0$, the function ψ is continuous. Differentiability is shown using induction:

$$\psi'(r) = \frac{-2r}{(r^2-1)^2} \exp\left(\frac{1}{r^2-1}\right) \mathbb{1}_{(-1,1)}(r)$$

and if $\psi^{(n)}(r) = \rho_n(r)\exp\left[1/(r^2-1)\right] \mathbb{1}_{(-1,1)}(r)$, then

$$\psi^{(n+1)}(r) = \left(\rho_n'(r) - \rho_n(r)\frac{2r}{(r^2-1)^2}\right) \exp\left(\frac{1}{r^2-1}\right) \mathbb{1}_{(-1,1)}(r).$$

This shows that $\rho_n(r)$ is a rational function, and it is clear that $\rho_n(r)e^{1/(r^2-1)} \to 0$ as $r \uparrow 1$. This proves that $\psi(r)$ is arbitrarily often differentiable at the points $r = \pm 1$ (with zero derivative). □

Let ϕ be a function as in Lemma 15.9 and set $\phi_\epsilon(x) := \epsilon^{-n}\phi(x/\epsilon)$ for $\epsilon > 0$. The function $J_\epsilon(u) := \phi_\epsilon \star u$ is the *Friedrichs mollifier* of $u \in \mathcal{L}^p(\mu)$, $1 \leqslant p < \infty$.

Lemma 15.10 *Let $J_\epsilon(u)$ be the Friedrichs mollifier of $u \in \mathcal{L}^p(\mu)$, $1 \leqslant p < \infty$.*

(i) $\phi_\epsilon \in C_c^\infty(\mathbb{R}^n)$, $\operatorname{supp}\phi_\epsilon \subset \overline{B_\epsilon(0)}$ *and* $\|\phi_\epsilon\|_1 = 1$.
(ii) $\operatorname{supp} u \star \phi_\epsilon \subset \operatorname{supp} u + \operatorname{supp}\phi_\epsilon \subset \{y : \exists x \in \operatorname{supp} u : |x - y| \leqslant \epsilon\}$.
(iii) $\|\phi_\epsilon \star u\|_p \leqslant \|u\|_p$ *for all* $\epsilon > 0$.

Proof The first two properties in (i) follow directly from the definition of ϕ_ϵ. Since the linear map $x \mapsto \epsilon^{-1}x$ is represented by a diagonal matrix A with entries ϵ^{-1} and determinant ϵ^{-n}, we can use (15.6) and find

$$\int \phi_\epsilon(x)dx = \int \epsilon^{-n}\phi(\epsilon^{-1}x)dx = \int \phi(y)dy = 1.$$

(ii) Write $K_\epsilon := \operatorname{supp}\phi_\epsilon$. By definition,

$$\int u(x-y)\phi_\epsilon(y)dy = \int_{K_\epsilon} u(x-y)\phi_\epsilon(y)dy.$$

Since $x - y \notin \operatorname{supp} u \Leftrightarrow x \notin y + \operatorname{supp} u$, we conclude that for all $x \notin \operatorname{supp} u + K_\epsilon$ the convolution $u \star \phi_\epsilon(x) = 0$. Thus, $\{u \star \phi_\epsilon \neq 0\} \subset \operatorname{supp} u + K_\epsilon$. Since the set on the right is closed [✍], the assertion follows.

(iii) is an immediate consequence of Young's inequality (15.7). □

Theorem 15.11 *Let λ^n be Lebesgue measure on $\mathbb{R}^n$ and $u \in \mathcal{L}^p(\lambda^n)$ for some $p \in [1, \infty)$. The Friedrichs mollifier $\phi_\epsilon \star u$ is a C^∞-function which converges in $\mathcal{L}^p(\mu)$ to u:*

$$\lim_{\epsilon \downarrow 0} \|u - \phi_\epsilon \star u\|_p = 0.$$

Proof Let $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$ be a multi-index and write $\partial^\alpha = \prod_{i=1}^n \partial^{\alpha_i}/\partial x_i^{\alpha_i}$. Observe that $\partial^\alpha \phi_\epsilon \in C_c{}^\infty(\mathbb{R}^n)$ and for all $|x| < R$

$$|\partial_x^\alpha \phi_\epsilon(x-y)u(y)| \leqslant c_\alpha \mathbb{1}_{K_\epsilon}(y-x)|u(y)| \leqslant c_\alpha \mathbb{1}_{K_{\epsilon+R}}(y)|u(y)|,$$

where $K_\epsilon = \overline{B_\epsilon(0)}$ and $c_\alpha = \sup_x |\partial^\alpha \phi_\epsilon(x)|$. By Hölder's inequality, the right-hand side is integrable. Therefore, we can (repeatedly) use the differentiability lemma (Theorem 12.5) and find that $\partial^\alpha(\phi_\epsilon \star u) = (\partial^\alpha \phi_\epsilon) \star u$; in particular, we have that $\phi_\epsilon \star u \in C^\infty(\mathbb{R}^n)$.

Since $\int \phi_\epsilon(y)dy = 1$, we get from Jensen's inequality and Tonelli's theorem,

$$\begin{aligned}\|u - u \star \phi_\epsilon\|_p^p &= \int \left| \int \big(u(x) - u(x-y)\big)\phi_\epsilon(y)dy \right|^p dx \\ &\leqslant \int \|u(\cdot) - u(\cdot\, - y)\|_p^p \phi_\epsilon(y)dy \\ &\overset{(15.6)}{=} \int \|u(\cdot) - u(\cdot\, - \epsilon z)\|_p^p \phi(z)dz.\end{aligned}$$

Since the integrand $z \mapsto \|u(\cdot) - u(\cdot\, - z)\|_p^p$ is continuous, see Theorem 15.8, and bounded

$$\|u(\cdot) - u(\cdot\, - y)\|_p^p \leqslant (\|u\|_p + \|u(\cdot\, - y)\|_p)^p \overset{(15.4)}{=} 2^p \|u\|_p^p,$$

we can use dominated convergence to see that

$$\limsup_{\epsilon \to 0} \|u - u \star \phi_\epsilon\|_p^p \leqslant \int \lim_{\epsilon \to 0} \|u(\cdot) - u(\cdot\, - \epsilon z)\|_p^p \phi(z)dz = 0. \qquad \square$$

Problems

15.1. Let $(X, \mathscr{A}, \mu)$ be a measure space and let $T: X \to X$ be a bijective measurable map whose inverse $T^{-1}: X \to X$ is again measurable. Show that for every $f \in \mathcal{M}^+(\mathscr{A})$ one has

$$\int u\, d(T(f\mu)) = \int u \circ T f\, d\mu = \int u f \circ T^{-1}\, dT(\mu) = \int u\, d\big(f \circ T^{-1}\, T(\mu)\big).$$

15.2. Let $u \in \mathcal{L}^1(\mathbb{R}^n, \lambda^n)$ and $\epsilon > 0$. Show that $\int u(\epsilon x)dx = \epsilon^{-n} \int u(y)dy$.

15.3. Find $\mathbb{1}_{[0,1]} * \mathbb{1}_{[0,1]}$ and $\mathbb{1}_{[0,1]} * \mathbb{1}_{[0,1]} * \mathbb{1}_{[0,1]}$.

15.4. Show that $\operatorname{supp}(u * w) \subset \overline{\operatorname{supp} u + \operatorname{supp} w}$, $(A + B := \{a + b : a \in A,\ b \in B\})$.

15.5. **(Mellin convolution in the group $((0,\infty), \cdot)$)** Define on $((0,\infty), \mathscr{B}(0,\infty))$ the measure $\mu(dx) = x^{-1}\, dx$. The *Mellin convolution* of measurable $u, w : (0,\infty) \to \mathbb{R}$ is given by

$$u \circledast w(x) := \int_{(0,\infty)} u(xy^{-1}) w(y) \mu(dy).$$

(i) If $u, w \geqslant 0$, then $u \circledast w = w \circledast u$ is measurable and

$$\int_0^\infty u \circledast w\, d\mu = \int_0^\infty u\, d\mu \int_0^\infty w\, d\mu.$$

(ii) If $u \in \mathcal{L}^p(\mu)$, $p \in [1,\infty]$, and $w \in \mathcal{L}^1(\mu)$, then $u \circledast w \in \mathcal{L}^p(\mu)$ and

$$\|u \circledast w\|_p \leqslant \|u\|_p \cdot \|w\|_1.$$

15.6. Let $(X, \mathscr{A}, \mu)$ be a measure space and $(Y, \mathscr{B})$ be a measurable space. Assume that $T: A \to B$, $A \in \mathscr{A}$, $B \in \mathscr{B}$, is an invertible measurable map. Show that

$$T(\mu)|_B = T(\mu|_A)$$

with the restrictions $\mu|_A(\cdot) := \mu(A \cap \cdot\,)$ and $T(\mu)|_B := T(\mu)(B \cap \cdot\,)$.

15.7. Let μ be a measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and $x, y, z \in \mathbb{R}^n$. Find $\delta_x \star \delta_y$ and $\delta_z \star \mu$.

15.8. Let μ, ν be two σ-finite measures on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. Show that $\mu \star \nu$ has no atoms (see Problem 6.8) if μ has no atoms.

15.9. Let $\mathbb{P}$ be a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and denote by $\mathbb{P}^{\star n}$ the n-fold convolution product $\mathbb{P} \star \mathbb{P} \star \cdots \star \mathbb{P}$. If $\int |\omega| \mathbb{P}(d\omega) < \infty$, then

$$\int |\omega| \mathbb{P}^{\star n}(d\omega) \leqslant n \int |\omega| \mathbb{P}(d\omega) \quad \text{and} \quad \int \omega \mathbb{P}^{\star n}(d\omega) = n \int \omega \mathbb{P}(d\omega).$$

15.10. Let $p: \mathbb{R} \to \mathbb{R}$ be a polynomial and $u \in C_c(\mathbb{R})$. Show that $u \star p$ exists and is again a polynomial.

15.11. Let $w: \mathbb{R} \to \mathbb{R}$ be an increasing (and hence measurable, by Problem 8.21) and bounded function. Show that for every positive $u \in \mathcal{L}^1(\lambda^1)$ the convolution $u \star w$ is again increasing, bounded and continuous.

15.12. Assume that $u \in C_c(\mathbb{R}^n)$ and $w \in C^\infty(\mathbb{R}^n)$. Show that $u \star w$ exists, is of class C^∞ and satisfies

$$\frac{\partial}{\partial x_i}(u \star w) = u \star \left(\frac{\partial}{\partial x_i} w \right).$$

15.13. Modify the proof of Theorem 15.11 and show that $C_c^\infty(\mathbb{R}^n)$ is uniformly dense in $C_c(\mathbb{R}^n)$.

15.14. **Young's inequality.** Adapt the proof of Theorem 15.6 and show that

$$\|u \star w\|_r \leqslant \|u\|_p \cdot \|w\|_q$$

for all $p, q, r \in [1, \infty)$, $u \in \mathcal{L}^p(\lambda^n)$, $w \in \mathcal{L}^q(\lambda^n)$ and $r^{-1} + 1 = p^{-1} + q^{-1}$.

15.15. **A general Young inequality.** Generalize Young's inequality given in Problem 15.14 and show that

$$\|f_1 \star f_2 \star \cdots \star f_N\|_r \leqslant \prod_{j=1}^N \|f_j\|_p, \qquad p = \frac{Nr}{(N-1)r + 1},$$

for all $N \in \mathbb{N}$, $r \in [1, \infty)$ and $f_j \in \mathcal{L}^p(\lambda^n)$.

15.16. Define $\phi: \mathbb{R} \to \mathbb{R}$ by $\phi(x) := (1 - \cos x)\mathbb{1}_{[0, 2\pi)}(x)$, let $u(x) := 1$, $v(x) := \phi'(x)$ and $w(x) := \int_{(-\infty, x)} \phi(t)dt$. Then

(i) $u \star v(x) = 0$ for all $x \in \mathbb{R}$;
(ii) $v \star w(x) = \phi \star \phi(x) > 0$ for all $x \in (0, 4\pi)$;
(iii) $(u \star v) \star w \equiv 0 \neq u \star (v \star w)$.

Does this contradict the associativity of the convolution which is implicit in Theorem 15.6?

16

Jacobi's Transformation Theorem

In this chapter we will discuss the general change-of-variables formula for Lebesgue integrals and its most prominent application: coordinate changes. For Riemann integrals we know the following substitution rule:

$$\int_{\phi(a)}^{\phi(b)} u(y)dy = \int_a^b u(\phi(x))\phi'(x)dx, \tag{16.1}$$

where $\phi : [a,b] \to [\phi(a), \phi(b)] = \phi([a,b])$ is continuously differentiable and strictly increasing. The integral on the left-hand side of (16.1) ranges over the direct image $\phi([a,b])$ of $[a,b]$ and not the pre-image, as in the previous chapter, and we must be careful with its measurability.

If ϕ is continuously invertible, there is no problem: $\phi([a,b]) = \psi^{-1}([a,b])$ where $\psi := \phi^{-1}$ is continuous, and hence measurable. The analogue of (16.1) for *affine linear maps* $\Phi(x) = Ax + b$, where A is an invertible $n \times n$ matrix and $b \in \mathbb{R}^n$ can be treated with Theorem 15.1; see also Example 15.3(iii).

Lemma 16.1 *Let* $\Phi(x) = Ax + b$ *with* $A \in \mathbb{R}^{n\times n}$ *such that* $\det A \neq 0$ *and* $b \in \mathbb{R}^n$.

$$\int_{\Phi(U)} u(y)dy = \int_U u(Ax+b)|\det A|dx \quad \forall u \in \mathcal{L}^1(\lambda^n),\ U \subset \mathbb{R}^n \text{ open}. \tag{16.2}$$

Proof If we apply Theorem 15.1 to the right-hand side of (16.2), we get

$$\begin{aligned}\int_U u(Ax+b)|\det A|dx &= \int_{\mathbb{R}^n} u(\Phi(x))\mathbb{1}_{\Phi(U)}(\Phi(x))|\det A|dx \\ &= \int_{\mathbb{R}^n} u(y)\mathbb{1}_{\Phi(U)}(y)\Phi(|\det A|\lambda^d)(dy).\end{aligned}$$

We are done, if we show $\Phi(|\det A|\lambda^d) = \lambda^d$. Since $\Phi^{-1}(y) = A^{-1}y - A^{-1}b$, we can use Theorem 7.10 and Corollary 7.11 and find for all $B \in \mathscr{B}(\mathbb{R}^n)$

$$\begin{aligned}\Phi(|\det A|\lambda^n)(B) &= |\det A|\lambda^n(A^{-1}(B) - A^{-1}b)\\ &\overset{7.11}{=} |\det A|\lambda^n(A^{-1}(B)) \overset{7.10}{=} |\det A| \cdot |\det A|^{-1}\lambda^n(B). \qquad \square\end{aligned}$$

On comparing the formulae (16.1) and (16.2) we see that $|\det A|$ corresponds to $|\phi'(x)|$. Indeed, A is the Jacobian, i.e. the derivative

$$D\Phi(x) = \left(\frac{\partial}{\partial x_i}\Phi_k(x)\right)_{i,k=1}^{n},$$

of the affine linear map $\Phi(x) = Ax + b$.

Definition 16.2 Let $U, V \subset \mathbb{R}^n$ be open sets. A C^1-*diffeomorphism* is a bijection $\Phi : U \to V$ such that both Φ and Φ^{-1} are continuously differentiable. The derivative at the point $x \in U$,

$$D\Phi(x) = \left(\frac{\partial}{\partial x_i}\Phi_k(x)\right)_{i,k=1}^{n}$$

is called the *Jacobian*.

Caution If $\Psi := \Phi^{-1}$ is continuous, then it is measurable and $\Phi(B) = \Psi^{-1}(B)$, $B \in \mathscr{B}(U)$, is a Borel set. Recall that $\mathscr{B}(U) = U \cap \mathscr{B}(\mathbb{R}^n)$, see Example 3.3(vi) and Problem 3.13. [✍][1]

We will need the (rather deep) inverse function theorem from real analysis, see e.g. Rudin [42, Theorem 9.24, pp. 221–228] for a proof.

Theorem 16.3 *Let $U \subset \mathbb{R}^n$ be an open subset. A map $\Phi : U \to \mathbb{R}^n$ is a C^1-diffeomorphism $\Phi : U \to V = \Phi(U)$ if, and only if,*

(a) *$\Phi : U \to \mathbb{R}^n$ is injective,*
(b) *Φ is continuously differentiable,*
(c) *$D\Phi(x)$ is for each $x \in U$ an invertible matrix.*

If (a)–(c) *hold, then $D(\Phi^{-1})(y) = (D\Phi)^{-1}(\Phi^{-1}(y))$ for every $y \in V$.*

Let us first give an intuitive argument for the general transformation formula. The starting point is the observation that a C^1-function $\Phi : U \to V$ has a Taylor expansion of the form

$$\Phi(x) = \Phi(x_0) + D\Phi(x_0)(x - x_0) + o(|x - x_0|) \quad \forall x, x_0 \in U. \tag{16.3}$$

[1] The Landau symbol $o(r)$ denotes an expression $f(r)$ which satisfies $\lim_{r\to 0} f(r)/r = 0$.

Exhaust the open set U with disjoint n-dimensional, half-open squares of side-length $< \delta$: $U = \dot{\bigcup}_{n=1}^{\infty} I_n$, $I_n \in \mathscr{J}(\mathbb{R}^n)$. Denoting by $x_n \in I_n$ the centre of the square I_n, we get for small δ

$$\begin{aligned}
\int_V u(y)dy &= \sum_{n=1}^{\infty} \int_{\Phi(I_n)} u(y)dy \\
&\underset{y\in\Phi(I_n)}{\overset{y\approx\Phi(x_n)}{\approx}} \sum_{n=1}^{\infty} \int_{\Phi(I_n)} u(\Phi(x_n))dy \\
&= \sum_{n=1}^{\infty} u(\Phi(x_n))\lambda^d(\Phi(I_n)) \\
&\overset{(16.3)}{\approx} \sum_{n=1}^{\infty} u(\Phi(x_n))\lambda^d\left(\Phi(x_n) + D\Phi(x_n)\cdot(I_n - x_n)\right) \\
&\overset{16.1}{=} \sum_{n=1}^{\infty} u(\Phi(x_n))|\det D\Phi(x_n)|\lambda^d(I_n) \\
&\underset{x\in I_n}{\overset{x\approx x_n}{\approx}} \sum_{n=1}^{\infty} \int_{I_n} u(\Phi(x))|\det D\Phi(x)|dx \\
&= \int_U u(\Phi(x))|\det D\Phi(x)|dx.
\end{aligned}$$

Let us now justify this heuristic argument.

Theorem 16.4 (Jacobi's transformation theorem) *Let $U, V \subset \mathbb{R}^n$ be open subsets, $\Phi : U \to V$ a C^1-diffeomorphism and $\lambda_W := \lambda^n(\cdot \cap W)$ the n-dimensional Lebesgue measure on $(W, \mathscr{B}(W))$. Then*

$$\lambda_V = \Phi(|\det D\Phi(\cdot)|\lambda_U) \tag{16.4}$$

or, equivalently,

$$\int_V u(y)dy = \int_U u(\Phi(x))|\det D\Phi(x)|dx \quad \forall u \in \mathcal{M}(\mathscr{B}(V)),\ \ u \geqslant 0. \tag{16.5}$$

In particular, $u \in \mathcal{L}^1(\lambda_V)$ if, and only if, $|\det D\Phi| u \circ \Phi \in \mathcal{L}^1(\lambda_U)$.

The proof of Theorem 16.4 is based on several auxiliary results. We will use $|x|_{\ell^\infty(n)} := \max_{1\leqslant i\leqslant n} |x_i|$ for the maximum norm of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and denote by $I \in \mathscr{J}^n$ a half-open rectangle $I = \bigtimes_{i=1}^{n} [a_i, b_i) \subset \mathbb{R}^n$.

Lemma 16.5 *Let $\Phi = (\Phi_1, \dots, \Phi_d) : \mathbb{R}^n \to \mathbb{R}^d$ be a Hölder continuous map of index $\alpha \in (0, 1]$, i.e.*

$$|\Phi(x) - \Phi(y)|_{\ell^\infty(d)} \leqslant L|x - y|^{\alpha}_{\ell^\infty(n)} \quad \forall x, y \in \mathbb{R}^n.$$

For all $c \in \mathbb{R}^n$ and $s > 0$ one has

$$\Phi\left(\bigtimes_{i=1}^{n}[c_i - s, c_i + s]\right) \subset \bigtimes_{i=1}^{d}\left[\Phi_i(c) - Ls^\alpha, \Phi_i(c) + Ls^\alpha\right].$$

In particular,

$$\lambda^d(\Phi(I)) \leqslant (2^{1-\alpha}L)^d(\lambda^n(I))^{\alpha d/n}$$

for all 'squares' $I \in \mathcal{J}^n$, i.e. n-dimensional rectangles with all sides of equal length.

Proof Set $I := \bigtimes_{i=1}^{n}[c_i - s, c_i + s] = \left\{y \in \mathbb{R}^n : |c - y|_{\ell^\infty(n)} \leqslant s\right\}$. The Hölder condition implies that

$$\Phi\left\{y : |c - y|_{\ell^\infty(n)} \leqslant s\right\} \subset \left\{\Phi(y) : |\Phi(c) - \Phi(y)|_{\ell^\infty(d)} \leqslant Ls^\alpha\right\} =: J.$$

Since the squares I and J have side-length $2s$ and $2Ls^\alpha$, respectively, we get

$$\lambda^d(\Phi(I)) \leqslant \lambda^d(J) = (2^{1-\alpha}L)^d[(2s)^n]^{\alpha d/n} = (2^{1-\alpha}L)^d(\lambda^n(I))^{\alpha d/n}. \qquad \square$$

Lemma 16.6 *Let μ and ν be two measures on the measurable space $(X, \mathscr{A})$ and let $\mathcal{S}$ be a semi-ring such that $\sigma(\mathcal{S}) = \mathscr{A}$. If*

$$\mu(S) \leqslant \nu(S) < \infty \quad \forall S \in \mathcal{S},$$

and if there is a sequence $(S_i)_{i\in\mathbb{N}} \subset \mathcal{S}$ with $S_i \uparrow X$, then $\mu \leqslant \nu$.

Proof It is clear from the properties of μ and ν that $\rho := \nu - \mu : \mathcal{S} \to [0, \infty)$ is a pre-measure. By Theorem 6.1, ρ has a unique extension to a measure $\tilde{\rho}$ on $\mathscr{A}$ and

$$\nu(S) = (\rho + \mu)(S) = \widetilde{\rho + \mu}(S) \quad \forall S \in \mathcal{S},$$

where $\widetilde{\rho + \mu}$ is the unique extension of the pre-measure $(\rho + \mu)\big|_{\mathcal{S}}$ to a measure on $\mathscr{A}$. But the measures $\tilde{\rho} + \mu$ and ν satisfy

$$\nu(S) = \tilde{\rho}(S) + \mu(S) = \rho(S) + \mu(S) = \widetilde{\rho + \mu}(S) \quad \forall S \in \mathcal{S},$$

and we conclude from the uniqueness of the extensions that $\nu = \tilde{\rho} + \mu$ on $\mathscr{A}$, i.e. $\nu(A) - \mu(A) = \tilde{\rho}(A) \geqslant 0$ for all $A \in \mathscr{A}$. $\square$

Caution Lemma 16.6 fails if $\mathcal{S}$ is not a semi-ring; see Problem 16.5.

Lemma 16.7 *For any C^1-diffeomorphism $\Phi: U \to V$ ($U, V \subset \mathbb{R}^n$ are open sets) we have*

$$\lambda_V(\Phi(I)) \leqslant \int_I |\det D\Phi(x)| \lambda_U(dx) \quad \forall I \in \mathcal{J}, \ \bar{I} \subset U. \tag{16.6}$$

Proof Let $I = \times_{i=1}^n [a_i, b_i) \in \mathcal{J}^n$ such that the closure $\bar{I} = \times_{i=1}^n [a_i, b_i] \subset U$ and denote by $\|A\| := \sup_{|x|_{\ell^\infty(n)} \leqslant 1} |Ax|_{\ell^\infty(n)}$, $A \in \mathbb{R}^{n \times n}$, the subordinate matrix norm for the vector norm $|\cdot|_{\ell^\infty(n)}$.[2] Using the inverse function theorem (Theorem 16.3) we find

$$L := \sup_{x \in \bar{I}} \|(D\Phi)^{-1}(x)\| \leqslant \sup_{y \in \Phi(\bar{I})} \|D(\Phi^{-1})(y)\|.$$

Since $D\Phi$ is uniformly continuous on the compact set $\bar{I} \subset U$, we find for a given $\epsilon > 0$ some $\delta > 0$ such that

$$\sup_{\substack{|x - x'|_{\ell^\infty(n)} \leqslant \delta \\ x, x' \in \bar{I}}} \|D\Phi(x) - D\Phi(x')\| \leqslant \frac{\epsilon}{L}. \tag{16.7}$$

Now we partition $I = \dot{\bigcup}_{i=1}^N I_i$ into disjoint half-open squares $I_i \in \mathcal{J}^n$ of the same side-length $< \delta$. Since $D\Phi$ and $\det D\Phi$ are continuous functions [✍], we can find for each $i = 1, \ldots, N$ a point

$$x_i \in \bar{I}_i \quad \text{such that} \quad |\det D\Phi(x_i)| = \inf_{x \in I_i} |\det D\Phi(x)|. \tag{16.8}$$

Since $A_i := D\Phi(x_i)$ is an invertible matrix, the linear map $x \mapsto A_i x$ satisfies

$$\begin{aligned} D\Big(A_i^{-1} \circ \Phi\Big)(x) &= A_i^{-1} \circ (D\Phi)(x) \\ &= \mathrm{id}_n + A_i^{-1} \circ (D\Phi(x) - D\Phi(x_i)) \end{aligned}$$

(id_n denotes the $n \times n$ identity matrix). The estimates (16.7), (16.8) show that

$$\sup_{x \in \bar{I}_i} \Big\| D\Big(A_i^{-1} \circ \Phi\Big)(x) \Big\| \leqslant 1 + L\frac{\epsilon}{L} = 1 + \epsilon \quad \forall i = 1, \ldots, N.$$

By the mean value theorem there are intermediate points $\xi \in \bar{I}$ such that

$$\begin{aligned} \Big| A_i^{-1} \circ \Phi(x) - A_i^{-1} \circ \Phi(x') \Big|_{\ell^\infty(n)} &\leqslant \Big| \Big(D(A_i^{-1} \circ \Phi)(\xi)\Big)(x - x') \Big|_{\ell^\infty(n)} \\ &\leqslant \Big\| D(A_i^{-1} \circ \Phi)(\xi) \Big\| \cdot |x - x'|_{\ell^\infty(n)}. \end{aligned}$$

[2] It is not difficult to see that $\|A\| = \max_{1 \leqslant i \leqslant n} \sum_{k=1}^n |a_{ik}|$ is the row-sum norm, see Stoer and Bulirsch [51, p. 178].

This shows that $A_n^{-1} \circ \Phi$ is Lipschitz (1-Hölder) continuous with Lipschitz constant $L = 1 + \epsilon$. Lemma 16.1 and Lemma 16.5 (with $n = d$ and $\alpha = 1$) now show

$$\begin{aligned}\lambda_V(\Phi(I_i)) &= \lambda_V(A_i \circ A_i^{-1} \circ \Phi(I_i)) \\ &= |\det A_i| \lambda_{A_i^{-1}V}(A_i^{-1} \circ \Phi(I_i)) \\ &\leqslant |\det A_i| (1+\epsilon)^n \lambda_U(I_i).\end{aligned}$$

Since $I = \dot{\bigcup}_{i=1}^N I_i$ and $|\det A_i| \leqslant |\det D\Phi(x)|$ for $x \in I_i$, we have

$$\begin{aligned}\lambda_V(\Phi(I)) \leqslant \sum_{i=1}^N \lambda_V(\Phi(I_i)) &\leqslant (1+\epsilon)^n \sum_{i=1}^N |\det A_i| \lambda_U(I_i) \\ &\leqslant (1+\epsilon)^n \sum_{i=1}^N \int_{I_i} |\det D\Phi(x)| \lambda_U(dx) \\ &= (1+\epsilon)^n \int_I |\det D\Phi(x)| \lambda_U(dx),\end{aligned}$$

and the proof is finished by letting $\epsilon \to 0$. $\square$

We can finally proceed to the proof of Theorem 16.4.

Proof of Theorem 16.4 Step 1. (16.4) $\Leftrightarrow$ (16.5) If $u = \mathbb{1}_B$, then (16.5) becomes (16.4); on the other hand, (16.5) follows from (16.4) with the usual 'from measures to integrals' construction, see Fig. 10.1. It suffices, therefore, to prove (16.4).

Step 2. Using $\Psi := \Phi^{-1}$ we can write $\mu := \lambda_V \circ \Phi = \lambda_V \circ \Psi^{-1} = \Psi(\lambda_V)$ as an image measure. On the other hand, $\nu(B) := \int_{B \cap U} |\det D\Phi(x)| \lambda_U(dx)$ is a measure with a continuous density function $|\det D\Phi(x)|$.

From Lemma 16.7 we know that $\mu \leqslant \nu$ on $\mathscr{J}^n$, and Lemma 16.6 shows that $\mu \leqslant \nu$ on $\mathscr{B}(\mathbb{R}^n)$. This proves '$\leqslant$' in (16.4). Because of Step 1, this means that

$$\int u(y) \lambda_V(dy) \leqslant \int u(\Phi(x)) |\det D\Phi(x)| \lambda_U(dx) \quad \forall u \in \mathcal{M}^+(\mathscr{B}(V)). \tag{16.9}$$

Step 3. For the converse inequality we use the C^1-diffeomorphism $\Psi = \Phi^{-1}$, switch in (16.9) the rôles of $U \leftrightarrow V$ and $x \leftrightarrow y$, and consider for $A \in \mathscr{B}(U)$ the function $u(x) = \mathbb{1}_{\Phi(A)} \circ \Phi(x)|\det D\Phi(x)|$. This gives

$$\begin{aligned}
\int u(x)\lambda_U(dx) &= \int \mathbb{1}_{\Phi(A)} \circ \Phi(x)|\det D\Phi(x)|\lambda_U(dx) \\
&\overset{(16.9)}{\leqslant} \int \left(\mathbb{1}_{\Phi(A)} \circ \Phi|\det D\Phi|\right) \circ \Psi(y)|\det D\Psi(y)|\lambda_V(dy) \\
&= \int \mathbb{1}_{\Phi(A)}(y)|\det(D\Phi) \circ \Psi(y)| \cdot |\det D\Psi(y)|\lambda_V(dy) \\
&= \int \mathbb{1}_{\Phi(A)}(y)\Big|\det\big[\underbrace{(D\Phi) \circ \Psi(y) \cdot D\Psi(y)}_{\mathrm{id}_n = D(\mathrm{id}_n) = D(\Phi\circ\Psi) = (D\Phi)\circ\Psi\cdot D\Psi}\big]\Big|\lambda_V(dy) \\
&= \int \mathbb{1}_{\Phi(A)}(y)\lambda_V(dy) = \lambda_V(\Phi(A)).
\end{aligned}$$

This gives the direction '$\geqslant$' in (16.4), finishing the proof. □

Remark 16.8 The formula (16.4) has the following interesting interpretation in connection with the Radon–Nikodým theorem (Theorem 25.2) and Lebesgue's differentiation theorem for measures (Theorem 25.20, in particular Corollary 25.21):

$$\frac{d\lambda^n \circ \Phi}{d\lambda^n}(x) = |\det D\Phi(x)| = \lim_{r\to 0} \frac{\lambda^n(\Phi(B_r(x)))}{\lambda^n(B_r(x))}.$$

A Useful Generalization of the Transformation Theorem

It is easy to see that the one-dimensional change-of-variable formula (16.1) still holds for strictly monotone functions ϕ which are only piecewise C^1. The argument is simple: split the domain of integration at the 'bad' points and observe that these are a Lebesgue null set. This is also true in $\mathbb{R}^n$: Jacobi's transformation formula works for maps $\Phi: U \to V$ which are only *almost everywhere diffeomorphic*. Typical examples are coordinate changes, e.g. polar coordinates in $\mathbb{R}^2$,

$$(0,\infty) \times (0, 2\pi] \to \mathbb{R}^2 \setminus \{(0,0)\}, \quad (r,\theta) \mapsto (r\cos\theta, r\sin\theta).$$

But now we need to understand what happens to images of Lebesgue null sets and Borel sets. Recall that $\mathscr{N}(\lambda^n) := \{N \in \mathscr{B}(\mathbb{R}^n) : \lambda^n(N) = 0\}$ are the (Borel measurable) null sets of the measure λ^n.

Lemma 16.9 *Let $\Phi: \mathbb{R}^n \to \mathbb{R}^d$ be a Hölder continuous map with index $\alpha \in (0,1]$. If $\alpha d \geqslant n$, then*

$$N^* \subset N \in \mathscr{N}(\lambda^n) \implies \exists M \in \mathscr{N}(\lambda^d) : \Phi(N^*) \subset M.$$

Proof From Carathéodory's extension of Lebesgue pre-measure λ^n we know, see (6.1) with $\mathscr{S} = \mathscr{J}^n$, that N is a λ^n null set if for every $\epsilon > 0$ there is a sequence of rectangles $(I_i^\epsilon)_{i\in\mathbb{N}} \subset \mathscr{J}^n$ such that

$$N \subset \bigcup_{i=1}^\infty I_i^\epsilon \quad \text{and} \quad \sum_{i=1}^\infty \lambda^n(I_i^\epsilon) < \epsilon.$$

Without loss of generality we may assume that the rectangles I_i^ϵ are squares (i.e. rectangles with all sides equally long) with side-length $\leqslant 1$, otherwise we would subdivide each I_i^ϵ into non-overlapping squares. Now we can use Lemma 16.5 and see that $\Phi(I_i^\epsilon) \subset J_i{}^\epsilon \in \mathscr{J}^n$, where $\lambda^d(J_i{}^\epsilon) \leqslant (2^{1-\alpha}L)^d(\lambda^n(I_i^\epsilon))^{\alpha d/n}$. We have

$$\Phi(N^*) \subset \Phi(N) \subset \bigcup_{i=1}^\infty \Phi(I_i^\epsilon) \subset \bigcup_{i=1}^\infty J_i{}^\epsilon$$

and, since $\alpha d \geqslant n$, we get $(\lambda^n(I_i^\epsilon))^{\alpha d/n} \leqslant \lambda^n(I_i^\epsilon)$, i.e.

$$\sum_{i=1}^\infty \lambda^d(J_i{}^\epsilon) \leqslant (2^{1-\alpha}L)^d \sum_{i=1}^\infty \underbrace{(\lambda^n(I_i^\epsilon))^{\alpha d/n}}_{\leqslant \lambda^n(I_i^\epsilon)} \leqslant (2^{1-\alpha}L)^d\epsilon. \qquad \square$$

We need the completion of the measure space $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \lambda^n)$. It is defined as

$$\overline{\mathscr{B}}(\mathbb{R}^n) = \{B^* = B \cup N^* : B \in \mathscr{B}(\mathbb{R}^n),\ N^* \subset N \in \mathscr{B}(\mathbb{R}^n),\ \lambda^n(N) = 0\},$$

$$\bar\lambda^n(B^*) := \lambda^n(B) \ \text{ and } \ \mathscr{N}(\bar\lambda^n) = \{N^* : \exists N \in \mathscr{N}(\lambda^n),\ N^* \subset N\}$$

We need that $(\mathbb{R}^n, \overline{\mathscr{B}}(\mathbb{R}^n), \bar\lambda^n)$ is a measure space – and this is easy to check [✍].[3]

Corollary 16.10 *Let $B \in \overline{\mathscr{B}}(\mathbb{R}^n)$, and let $U := B^\circ$ be the open interior*[4] *and $U' \supset B$ an open neighbourhood. Assume that $\Phi : U' \to \mathbb{R}^n$ is Lipschitz* (1 – *Hölder*) continuous. If $B \setminus U \in \mathscr{N}(\bar\lambda^n)$ and $\Phi : U \to \Phi(U)$ is a C^1-diffeomorphism, then

$$\int_{\Phi(B)} u(y)\bar\lambda^n(dy) = \int_B u\circ\Phi(x)|\det D\Phi(x)|\bar\lambda^n(dx) \quad \forall u \in \mathcal{M}^+(\overline{\mathscr{B}}(\Phi(B))).$$

In particular, $u \in \mathcal{L}^1(\bar\lambda^n, \Phi(B))$ if, and only if, $|\det D\Phi| \cdot u\circ\Phi \in \mathcal{L}^1(\bar\lambda^n, B)$.

Proof We show that $\Phi(B)$ is measurable. Clearly, $\Phi(B) = \Phi(U) \cup \Phi(B \setminus U)$, and from Lemma 16.9 we know that $\Phi(B \setminus U) \in \mathscr{N}(\bar\lambda^n)$. Since $\Phi|_U$ is continuously invertible, $\Phi(U)$ is a Borel set, so $\Phi(B) \in \overline{\mathscr{B}}(\mathbb{R}^n)$.

[3] More details on the completion of measure spaces are contained in Problems 4.15, 6.4, 11.5 and 11.6. The advantage of $\overline{\mathscr{B}}(\mathbb{R}^d)$ over $\mathscr{B}(\mathbb{R}^d)$ is that α-Hölder continuous maps $\Phi : \mathbb{R}^n \to \mathbb{R}^d$ map $\overline{\mathscr{B}}(\mathbb{R}^n)$-measurable sets into $\overline{\mathscr{B}}(\mathbb{R}^d)$-sets if $\alpha d \geqslant n$, see the following section on images of Borel sets.

[4] This is the largest open set contained in B.

The argument used for the proof of Theorem 16.4 remains literally valid for $\bar{\lambda}^n$, i.e. the difficulty in Corollary 16.10 is not the completion of the measure space but the fact that Φ is only almost everywhere a diffeomorphism.

Since $\bar{\lambda}^n(B \setminus U) = 0$, we get $\Phi(B) \setminus \Phi(U) \subset \Phi(B \setminus U) \in \mathcal{N}(\bar{\lambda}^n)$. In view of Corollary 11.3 we can alter $\mathcal{L}^1$-functions on null sets, which means that

$$\int_{\Phi(B)} u\, d\bar{\lambda}^n = \int_{\Phi(U)} u\, d\bar{\lambda}^n = \int_{\Phi(U)} u\, d\lambda^n$$

and

$$\begin{aligned}\int \left(u \mathbb{1}_{\Phi(B)}\right) \circ \Phi |\det D\Phi|\, d\bar{\lambda}^n &= \int \left(u \mathbb{1}_{\Phi(U)}\right) \circ \Phi |\det D\Phi|\, d\bar{\lambda}^n \\ &= \int \left(u \mathbb{1}_{\Phi(U)}\right) \circ \Phi |\det D\Phi|\, d\lambda^n\end{aligned}$$

hold. Therefore, the integral formula follows from Theorem 16.4. □

Images of Borel Sets

Often we need that the direct image $\Phi(B)$ of a Borel set $B \in \mathscr{B}(\mathbb{R}^n)$ is measurable. If Φ has a continuous (or at least measurable) inverse $\Psi = \Phi^{-1}$, this is no issue at all. In the proof of Corollary 16.10 we had the situation that Φ is only a.e. measurably invertible and – to a certain extent – we lost control of null sets, which prompted us to consider the completed Borel σ-algebra $\overline{\mathscr{B}}(\mathbb{R}^n)$.

In this section we want to discuss the structure of direct images of Borel sets. For this, we need a few notions from topology. A set $F \subset \mathbb{R}^n$ [$G \subset \mathbb{R}^n$] is called an F_σ-set [G_δ-set] if it is the *countable union of closed sets* [*countable intersection of open sets*], i.e. if

$$F = \bigcup_{i \in \mathbb{N}} C_i \qquad \left[G = \bigcap_{i \in \mathbb{N}} U_i\right] \tag{16.10}$$

for closed sets C_i [open sets U_i]. Obviously, both F_σ- and G_δ-sets are Borel sets; but, in general, it holds neither that F_σ-sets are closed nor that G_δ-sets are open.

Lemma 16.11 *Let $\Phi : \mathbb{R}^n \to \mathbb{R}^d$ be a continuous map and $F \subset \mathbb{R}^n$ be an F_σ-set. The image $\Phi(F)$ is an F_σ-set in $\mathbb{R}^d$.*

Proof Since F is an F_σ-set, it has a representation $F = \bigcup_{i \in \mathbb{N}} C_i$ with closed sets $C_i \subset \mathbb{R}^n$. Moreover,

$$F = \bigcup_{k \in \mathbb{N}} F \cap \overline{B_k(0)} = \bigcup_{i \in \mathbb{N}} \bigcup_{k \in \mathbb{N}} C_i \cap \overline{B_k(0)} =: \bigcup_{l \in \mathbb{N}} K_l,$$

where $(K_l)_{l\in\mathbb{N}}$ is an enumeration of the family $(C_i \cap \overline{B_k(0)})_{i,k\in\mathbb{N}}$ of closed and bounded, and hence compact, sets. Since images of compact sets under continuous maps are compact, $\Phi(K_l) \subset \mathbb{R}^d$ is compact and, in particular, closed. So,

$$\Phi(F) \stackrel{2.5}{=} \bigcup_{l\in\mathbb{N}} \Phi(K_l)$$

is an F_σ-set in $\mathbb{R}^d$. □

Lemma 16.12 *Let $B \in \mathscr{B}(\mathbb{R}^n)$ be a Borel set. Then there exists an F_σ-set F and a G_δ-set G such that*

$$F \subset B \subset G \quad \textit{and} \quad \lambda^n(F) = \lambda^n(B) = \lambda^n(G).$$

Proof The proof consists of three stages.

Step 1. Construction of the set G. If $\lambda^n(B) = \infty$, then we take $G = \mathbb{R}^n$. If $\lambda^n(B) < \infty$, we use Carathéodory's extension theorem (we take in (6.1) of Theorem 6.1 $\mu = \lambda^n$ and $\mathscr{S} = \mathscr{J}^n$, see also Lemma 16.9), to construct for every $k \in \mathbb{N}$ a sequence of half-open rectangles $(I_i^k)_{i\in\mathbb{N}} \subset \mathscr{J}^n$ such that

$$B \subset \bigcup_{i\in\mathbb{N}} I_i^k \quad \text{and} \quad \sum_{i\in\mathbb{N}} \lambda^n(I_i^k) \leqslant \lambda^n(B) + \frac{1}{k}.$$

Without loss of generality we may assume that the I_i^k, $k \in \mathbb{N}$, are 'squares', i.e. all sides have the same length s_i; otherwise we would partition I_i^k in such squares. We can now enlarge I_i^k by moving the lower left corner by $\epsilon_i := (s_i^n + 2^{-i}/k)^{1/n} - s_i$ units 'to the left' in each coordinate direction (Fig. 16.1). The new *open* square $\tilde{I}_i^k$ has volume

$$\lambda^n(\tilde{I}_i^k) = \lambda^n(I_i^k) + \frac{1}{k}2^{-i},$$

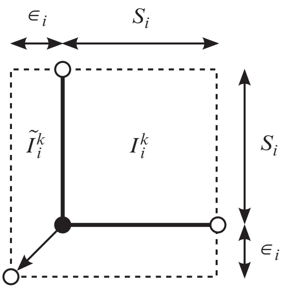


Fig. 16.1. For the proof of Lemma 16.12.

and we see that the open sets $G^k := \bigcup_{i\in\mathbb{N}} \tilde{I}_i^k \supset B$ satisfy the following estimate

$$\lambda^n(G^k) \stackrel{4.3(\text{viii})}{\leqslant} \sum_{i\in\mathbb{N}} \lambda^n(\tilde{I}_i^k) = \sum_{i\in\mathbb{N}} \lambda^n(I_i^k) + \frac{1}{k} \leqslant \left(\lambda^n(B) + \frac{1}{k}\right) + \frac{1}{k}.$$

Thus $G := \bigcap_{k\in\mathbb{N}} G^k$ is a G_δ-set with $G \supset B$, and

$$\lambda^n(B) \leqslant \lambda^n(G) \stackrel{4.3(\text{vii})}{=} \lim_{k\to\infty} \lambda^n(G^k) \leqslant \lim_{k\to\infty} \left(\lambda^n(B) + \frac{2}{k}\right) = \lambda^n(B).$$

Step 2. Construction of the set F if $\lambda^n(B) < \infty$. Denote by $\overline{B}$ the closure[5] of B. Since $\overline{B} \setminus B$ is a Borel set, we find as in Step 1 open sets U^k with

$$\overline{B} \setminus B \subset U^k \quad \text{and} \quad \lambda^n(U^k) \leqslant \lambda^n(\overline{B} \setminus B) + \frac{1}{k}. \tag{16.11}$$

Observe that

$$B \subset \big(B \setminus U^k\big) \cup \big(U^k \cap B\big) \subset \big(\overline{B} \setminus U^k\big) \cup \big(U^k \setminus \big(\overline{B} \setminus B\big)\big),$$

so that by the subadditivity of measures

$$\begin{aligned}
\lambda^n(B) &\leqslant \lambda^n\big(\overline{B} \setminus U^k\big) + \lambda^n\big(U^k \setminus (\overline{B} \setminus B)\big) \\
&= \lambda^n(\overline{B} \setminus U^k) + \lambda^n(U^k) - \lambda^n(\overline{B} \setminus B) \\
&\overset{(16.11)}{\leqslant} \lambda^n(\overline{B} \setminus U^k) + \frac{1}{k}.
\end{aligned}$$

By construction, $C_k := \overline{B} \setminus U^k \subset \overline{B} \setminus (\overline{B} \setminus B) = B$ is a closed set and $F := \bigcup_{k \in \mathbb{N}} C_k \subset B$ is an F_σ-set satisfying

$$\lambda^n(B) - \frac{1}{k} \leqslant \lambda^n(C_k) \leqslant \lambda^n\Big(\bigcup_{i \in \mathbb{N}} C_i\Big) = \lambda^n(F) \leqslant \lambda^n(B).$$

The claim follows as $k \to \infty$.

Step 3. Construction of the set F if $\lambda^n(B) = \infty$. Setting

$$B_i := B \cap \big(B_i(0) \setminus B_{i-1}(0)\big), \quad i \in \mathbb{N},$$

we get a disjoint partitioning of $B = \dot{\bigcup}_{i \in \mathbb{N}} B_i$, where each set B_i is a Borel set with finite volume. Applying Step 2 to each B_i, we find F_σ-sets $F_i \subset B_i$ with $\lambda^n(F_i) = \lambda^n(B_i)$, $i \in \mathbb{N}$. Since the B_i are mutually disjoint, so are the F_i, and since $F := \bigcup_{i \in \mathbb{N}} F_i$ is again an F_σ-set (see Problem 16.1), we end up with $F \subset B$ and

$$\lambda^n(F) = \sum_{i \in \mathbb{N}} \lambda^n(F_i) = \sum_{i \in \mathbb{N}} \lambda^n(B_i) = \lambda^n(B). \qquad \square$$

The following result and the condition (N) are due to Lusin.

Theorem 16.13 (Lusin) *Let* $(\mathbb{R}^m, \overline{\mathscr{B}}(\mathbb{R}^m), \bar{\lambda}^m)$, $m = d, n$, *be the completed Lebesgue measure spaces and* $\Phi : \mathbb{R}^n \to \mathbb{R}^d$ *be a continuous function. The following assertions are equivalent.*

(i) Φ *has the property* (N), *i.e.* $\Phi(N) \in \mathscr{N}(\bar{\lambda}^d)$ *for all* $N \in \mathscr{N}(\bar{\lambda}^n)$.

(ii) $\Phi(B^*) \in \overline{\mathscr{B}}(\mathbb{R}^d)$ *for all* $B^* \in \overline{\mathscr{B}}(\mathbb{R}^n)$.

[5] That is, the smallest closed set containing B.

Proof (i) $\Rightarrow$ (ii): Every $B^* \in \overline{\mathscr{B}}(\mathbb{R}^n)$ is of the form $B \cup N^*$, where $B \in \mathscr{B}(\mathbb{R}^n)$ and $N^* \in \mathscr{N}(\bar{\lambda}^n)$. Using Lemma 16.12 we see that $B^* = F \cup M^*$, where F is an F_σ-set and $M^* := N^* \cup (B \setminus F) \in \mathscr{N}(\bar{\lambda}^n)$. Observe that

$$\Phi(B^*) = \Phi(F) \cup \Phi(M^*);$$

by Lemma 16.11, $\Phi(F)$ is an F_σ-set, and hence a Borel set, and, because of property (N), $\Phi(M^*)$ is a Lebesgue null set. This shows that $\Phi(B^*) \in \overline{\mathscr{B}}(\mathbb{R}^d)$.

(ii) $\Rightarrow$ (i): Assume that Φ does *not* enjoy the property (N). Thus, there exists a set $N \in \mathscr{N}(\bar{\lambda}^n)$ such that the outer Lebesgue measure $(\lambda^d)^*(\Phi(N)) > 0$. In particular, $\Phi(N)$ contains a non-Lebesgue measurable set $B \notin \overline{\mathscr{B}}(\mathbb{R}^d)$.[6]

For any $b \in B$ we pick some $x \in N$ such that $\Phi(x) = b$, and let A be the set of all these x. By definition, A is the subset of a null set, and hence itself a null set, and as such measurable in $\overline{\mathscr{B}}(\mathbb{R}^d)$. Because of (ii), $B = \Phi(A)$ is measurable, and we have reached a contradiction. $\square$

Corollary 16.14 *Let $\Phi : \mathbb{R}^n \to \mathbb{R}^d$ be an α-Hölder continuous map with exponent $\alpha \in (0, 1]$. If $\alpha d \geqslant n$, then Φ maps the completed Borel σ-algebra $\overline{\mathscr{B}}(\mathbb{R}^n)$ into $\overline{\mathscr{B}}(\mathbb{R}^d)$, and the inequality*

$$\bar{\lambda}^d(\Phi(B^*)) \leqslant (2^{1-\alpha} L)^d \bar{\lambda}^n(B^*) \tag{16.12}$$

holds for all $B^* \in \overline{\mathscr{B}}(\mathbb{R}^n)$ for the completed Lebesgue measures $\bar{\lambda}^n$ and $\bar{\lambda}^d$.

Proof By Lemma 16.9 the map Φ enjoys Lusin's property (N), and the first part of the assertion follows from (the 'easy' direction (i) $\Rightarrow$ (ii) of) Theorem 16.13.

Assume that $B^* \in \overline{\mathscr{B}}(\mathbb{R}^n)$. If $\bar{\lambda}^n(B^*) = \infty$, (16.12) is trivial, and we will consider only the case $\bar{\lambda}^n(B^*) < \infty$. As in Step 1 of the proof of Lemma 16.12, there is for every $\epsilon > 0$ a sequence $(I_i^\epsilon)_{i \in \mathbb{N}} \subset \mathscr{J}^n$ such that all I_i^ϵ are *squares*, i.e. have sides of equal length $s_i^\epsilon < s < 1$, and such that

$$B^* \subset \bigcup_{i \in \mathbb{N}} I_i^\epsilon \quad \text{and} \quad \sum_{i \in \mathbb{N}} \lambda^n(I_i^\epsilon) \leqslant \bar{\lambda}^n(B^*) + \epsilon. \tag{16.13}$$

Moreover,

$$\bar{\lambda}^d(\Phi(B^*)) \leqslant \lambda^d\Big(\bigcup_{i \in \mathbb{N}} \Phi(I_i^\epsilon)\Big) \overset{4.3(\text{viii})}{\leqslant} \sum_{i \in \mathbb{N}} \lambda^d\big(\Phi(I_i^\epsilon)\big), \tag{16.14}$$

which means that it suffices to check (16.12) for a square $B^* = I$ of side-length $s < 1$ and centre $c \in \mathbb{R}^n$. This is just Lemma 16.5.

[6] This is non-trivial, see Corollary G.5 in Appendix G.

From (16.13), (16.14) we conclude

$$\bar{\lambda}^d(\Phi(B^*)) \leqslant \sum_{i\in\mathbb{N}} (2^{1-\alpha}L)^d (\lambda^n(I_i^\epsilon))^{\alpha d/n} \leqslant \sum_{i\in\mathbb{N}} (2^{1-\alpha}L)^d \lambda^n(I_i^\epsilon) \leqslant L^d\big(\bar{\lambda}^n(B^*) + \epsilon\big),$$

and the claim follows upon letting $\epsilon \to 0$. □

Polar Coordinates and the Volume of the Unit Ball

Some of the most interesting applications of Theorem 16.4 and Corollary 16.10 are coordinate changes.

Example 16.15 (planar polar coordinates) Consider the map

$$\Psi : (0,\infty) \times (-\pi,\pi) \to \mathbb{R}^2 \setminus ((-\infty,0] \times \{0\}), \quad \Psi(r,\theta) := (r\cos\theta, r\sin\theta),$$

which defines *polar coordinates* (r,θ) in $\mathbb{R}^2$. It is not hard to see that Ψ is a C^1-diffeomorphism. The determinant of the Jacobian is given by

$$\det\left(\frac{\partial\Psi(r,\theta)}{\partial(r,\theta)}\right) = \left|\begin{pmatrix}\cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta\end{pmatrix}\right| = r\cos^2\theta + r\sin^2\theta = r.$$

Since $(-\infty,0] \times \{0\}$ is a λ^2-null set, we can apply Corollary 16.10 and find for every $u : \mathbb{R}^2 \to \mathbb{R}$, $u \in \mathcal{L}^1(\mathbb{R}^2, \lambda^2)$

$$\begin{aligned}\int_{\mathbb{R}^2} u(x,y)d(x,y) &= \int_{(0,\infty)\times(0,2\pi)} ru(r\cos\theta, r\sin\theta)d(r,\theta) \\ &= \int_{(0,\infty)} \int_{(0,2\pi)} ru(r\cos\theta, r\sin\theta)d\theta\, dr,\end{aligned}$$

where we use Fubini's theorem for the last equality. This shows, in particular, that

$$u \in \mathcal{L}^1(\mathbb{R}^2) \iff (r,\theta) \mapsto ru(r\cos\theta, r\sin\theta) \in \mathcal{L}^1\big((0,\infty) \times (-\pi,\pi)\big).$$

A simple but quite interesting application of planar polar coordinates is the following formula which plays a central rôle in probability theory.

Example 16.16 We have

$$\int_{\mathbb{R}} e^{-x^2} d\lambda^1(x) = \sqrt{\pi}.$$

Proof We use the following trick: by Tonelli's theorem (Theorem 14.8)

$$\begin{aligned}\left(\int_{\mathbb{R}} e^{-x^2}\,dx\right)^2 &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-x^2}e^{-y^2}\,dx\,dy\\ &= \int_{\mathbb{R}^2} e^{-(x^2+y^2)}\,d(x,y)\\ &= \int_{(0,\infty)}\int_{(-\pi,\pi)} re^{-r^2}\,d\theta\,dr.\end{aligned}$$

Since re^{-r^2} is positive and improperly Riemann integrable [✍], we know that Lebesgue and Riemann integrals coincide (see Theorem 12.9 and Problem 12.35), and therefore

$$\left(\int_{\mathbb{R}} e^{-x^2}\,dx\right)^2 = \lambda^1(0,2\pi)\int_0^\infty re^{-r^2}\,dr = 2\pi\left[-\tfrac{1}{2}e^{-r^2}\right]_0^\infty = \pi. \qquad \square$$

Polar coordinates also exist in higher dimensions, but the formulae become more complicated. The idea here is that we parametrize $\mathbb{R}^n$ by the radius $r \in [0,\infty)$, and $n-1$ angles $\omega = (\theta_1,\dots,\theta_{n-2}) \in (0,\pi)^{n-2}$ and $\theta = \theta_{n-1} \in (-\pi,\pi)$, so that $x = \Psi(r,\omega,\theta)$. The Jacobian is now of the form $r^{n-1}J(\omega,\theta)$. See Fig. 16.2.

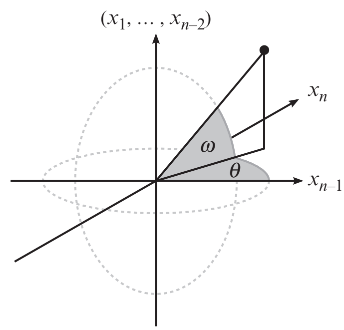


Fig. 16.2. Polar coordinates in higher dimensions.

Definition 16.17 (polar coordinates) Let $x = (x_1,\dots,x_n) \in \mathbb{R}^n$ be Cartesian coordinates and $(r,\boldsymbol{\theta}) = (r,\theta_1,\dots,\theta_{n-1})$, where $r \in (0,\infty)$, $\theta_1,\dots,\theta_{n-2} \in (0,\pi)$ and $\theta_{n-1} \in (-\pi,\pi)$, be polar coordinates. The coordinate transform is given by

$$\Psi : (r,\boldsymbol{\theta}) \mapsto x = \begin{cases} x_1 = r\cos\theta_1,\\ x_2 = r\sin\theta_1\cos\theta_2,\\ x_3 = r\sin\theta_1\sin\theta_2\cos\theta_3,\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ x_{n-1} = r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-2}\cos\theta_{n-1},\\ x_n = r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-2}\sin\theta_{n-1}.\end{cases} \tag{16.15}$$

It is not difficult to see [✍] that

$$\Psi : \underbrace{(0,\infty) \times (0,\pi)^{n-2} \times (-\pi,\pi)}_{=V_n} \longrightarrow \underbrace{\mathbb{R}^n \setminus \{x : x_n = 0,\ x_{n-1} \leqslant 0\}}_{=U_n}.$$

In particular, we find for $n=2$ (see Example 16.15)

$$\Psi : (0,\infty) \times (-\pi,\pi) \to \mathbb{R}^2 \setminus \big((-\infty,0] \times \{0\}\big)$$
$$x_1 = r\cos\theta_1, \quad x_2 = r\sin\theta_1,$$

and for $n=3$

$$\Psi : (0,\infty) \times (0,\pi) \times (-\pi,\pi) \to \mathbb{R}^3 \setminus \{(x_1,x_2,x_3) : x_3 = 0,\ x_2 \leqslant 0\}$$
$$x_1 = r\cos\theta_1, \quad x_2 = r\sin\theta_1\cos\theta_2, \quad x_3 = r\sin\theta_1\sin\theta_2.$$

Lemma 16.18 *Let $\Psi : V_n \to U_n$ denote polar coordinates in $\mathbb{R}^n$. The map Ψ is a C^1-diffeomorphism and the determinant of the Jacobian is given by*

$$\det D\Psi(r,\boldsymbol{\theta}) = r^{n-1}\sin^{n-2}\theta_1 \sin^{n-3}\theta_2 \cdots \sin\theta_{n-2}. \tag{16.16}$$

Proof We check the conditions (a)–(c) of Theorem 16.3.

Differentiability of Ψ is obvious.

Using induction in n we show injectivity. For $n=2$ there is nothing to show. Assume that the $(n-1)$-dimensional polar coordinates are one–to–one. We write $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) = (x_1, \boldsymbol{x}') \in U_n$. By assumption, there are unique polar coordinates $(\rho, \theta_2, \ldots, \theta_{n-1})$ for the point $\boldsymbol{x}' \in \mathbb{R}^{n-1}$. Therefore,

$$r^2 := x_1^2 + x_2^2 + \cdots + x_n^2 = x_1^2 + \rho^2 \quad \text{and} \quad \frac{x_1}{r} \in (-1,1).$$

These equations determine x_1 and ρ uniquely:

$$x_1 = r\cos\theta_1 \quad \text{and} \quad \rho = \sqrt{r^2 - x_1^2} = \sqrt{r^2(1-\cos^2\theta_1)} = r\sin\theta_1, \quad \theta_1 \in (0,\pi).$$

In order to show that $D\Psi(r,\boldsymbol{\theta})$ is invertible, we calculate its determinant. A direct attack is extremely messy. The following elegant proof can be found in Fichtenholz [17, Volume III, XVIII.676.8].

From (16.15) we get the equations

$$\boldsymbol{F}(r,\boldsymbol{\theta},\boldsymbol{x}) = \begin{cases} F_1(r,\boldsymbol{\theta},\boldsymbol{x}) = r^2 - (x_1^2 + \cdots + x_n^2) = 0, \\ F_2(r,\boldsymbol{\theta},\boldsymbol{x}) = r^2 \sin^2\theta_1 - (x_2^2 + \cdots + x_n^2) = 0, \\ F_3(r,\boldsymbol{\theta},\boldsymbol{x}) = r^2 \sin^2\theta_1 \sin^2\theta_2 - (x_3^2 + \cdots + x_n^2) = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ F_n(r,\boldsymbol{\theta},\boldsymbol{x}) = r^2 \sin^2\theta_1 \cdots \sin^2\theta_{n-1} - x_n^2 = 0. \end{cases}$$

Applying the implicit function theorem to the equation $\boldsymbol{F}(r,\boldsymbol{\theta},\boldsymbol{x}) = 0$ shows that

$$D\Psi(r,\boldsymbol{\theta}) = \frac{\partial \boldsymbol{x}}{\partial(r,\boldsymbol{\theta})} = -\left(\frac{\partial \boldsymbol{F}(r,\boldsymbol{\theta},\boldsymbol{x})}{\partial \boldsymbol{x}}\right)^{-1} \frac{\partial \boldsymbol{F}(r,\boldsymbol{\theta},\boldsymbol{x})}{\partial(r,\boldsymbol{\theta})}.$$

The Jacobians appearing in this formula are lower resp. upper triangular matrices: $\partial \boldsymbol{F}(r,\boldsymbol{\theta},\boldsymbol{x})/\partial(r,\boldsymbol{\theta})$ is given by

$$\begin{pmatrix} 2r & 0 & 0 & \dots & 0 \\ * & 2r^2\cos\theta_1\sin\theta_1 & 0 & \dots & 0 \\ * & * & 2r^2\cos\theta_2\sin\theta_2\sin^2\theta_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \dots & 2r^2\cos\theta_{n-1}\sin\theta_{n-1}\prod_{i=1}^{n-2}\sin^2\theta_i \end{pmatrix}$$

and

$$\frac{\partial \boldsymbol{F}(r,\boldsymbol{\theta},\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{pmatrix} -2x_1 & * & * & \cdots & * \\ 0 & -2x_2 & * & \cdots & * \\ 0 & 0 & -2x_3 & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -2x_n \end{pmatrix}.$$

Since the determinants of triangular matrices are easy to compute, we find

$$\begin{aligned} \det D\Psi(r,\boldsymbol{\theta}) &= \frac{2^n r^{2n-1}\sin^{2n-3}\theta_1\cos\theta_1\sin^{2n-5}\theta_2\cos\theta_2\cdots\sin\theta_{n-1}\cos\theta_{n-1}}{2^n x_1\cdots x_n} \\ &= \frac{2^n r^{2n-1}\sin^{2n-3}\theta_1\cos\theta_1\sin^{2n-5}\theta_2\cos\theta_2\cdots\sin\theta_{n-1}\cos\theta_{n-1}}{2^n r^n \sin^{n-1}\theta_1\cos\theta_1\sin^{n-2}\theta_2\cos\theta_2\cdots\sin\theta_{n-1}\cos\theta_{n-1}} \\ &= r^{n-1}\sin^{n-2}\theta_1\sin^{n-3}\theta_2\cdots\sin\theta_{n-2} \neq 0. \end{aligned}$$

This shows that $D\Psi(r,\boldsymbol{\theta})$ is invertible on V_n. □

The next corollary is a direct consequence of Jacobi's transformation theorem (Corollary 16.10), Lemma 16.18 and Fubini's theorem (Corollary 14.9).

Corollary 16.19 *Let $\boldsymbol{x} = \Psi(r, \boldsymbol{\theta})$ denote Cartesian and $(r, \boldsymbol{\theta})$ polar coordinates in $\mathbb{R}^n$. If $|J(r, \boldsymbol{\theta})|$ is the determinant of the Jacobian* (16.16)*, then $u(\boldsymbol{x}) \in \mathcal{L}^1(d\boldsymbol{x})$ if, and only if, $|J(r, \boldsymbol{\theta})|u(\Psi(r, \boldsymbol{\theta})) \in \mathcal{L}^1(dr \times d\boldsymbol{\theta})$. If this is the case, then*

$$\int_{\mathbb{R}^n} u(\boldsymbol{x}) d\boldsymbol{x} = \int_0^\infty \int_0^\pi \cdots \int_{-\pi}^\pi u(\Psi(r, \theta_1, \dots, \theta_{n-1}))|J(r, \boldsymbol{\theta})| d\theta_{n-1} \cdots d\theta_1 \, dr.$$

If $\mathbb{S}^{n-1} = \partial B_1(0)$ is the $(n-1)$-dimensional sphere in $\mathbb{R}^n$ and $\Gamma \in \mathscr{B}(\mathbb{S}^{n-1})$, then

$$\sigma_{n-1}(\Gamma) = \int_0^\pi \cdots \int_{-\pi}^\pi \mathbb{1}_\Gamma(\Psi(1, \theta_1, \dots, \theta_{n-1}))|J(1, \boldsymbol{\theta})| d\theta_{n-1} \cdots d\theta_1$$

is the canonical surface measure on $\mathbb{S}^{n-1}$. Moreover, $u \in \mathcal{L}^1(d\boldsymbol{x})$ if, and only if, $r^{n-1}u(r\boldsymbol{s}) \in \mathcal{L}^1(dr \times d\sigma_{n-1})$. If this is the case, then

$$\int_{\mathbb{R}^n} u(\boldsymbol{x}) d\boldsymbol{x} = \int_0^\infty \int_{\mathbb{S}^{n-1}} r^{n-1} u(r\boldsymbol{s}) \sigma_{n-1}(d\boldsymbol{s}) dr.$$

The fact that σ_{n-1} is the standard surface measure on the sphere $\mathbb{S}^{n-1}$ follows from Cavalieri's principle (Fubini's theorem) and from the observation that $\mathbb{R}^n \setminus \{0\} = \bigcup_{r>0} \partial B_r(0)$ – see also Theorem 16.22 below.

An important special case is the following formula for the integral of rotationally invariant functions.

Corollary 16.20 *Let $f(\boldsymbol{x}) = \phi(|\boldsymbol{x}|)$ be a rotationally symmetric function on $\mathbb{R}^n$. One has $f(\boldsymbol{x}) \in \mathcal{L}^1(d\boldsymbol{x})$ if, and only if, $r^{n-1}\phi(r) \in \mathcal{L}^1((0, \infty), dr)$; in this case,*

$$\int_{\mathbb{R}^n} f(\boldsymbol{x}) d\boldsymbol{x} = n \cdot \omega_n \int_0^\infty r^{n-1} \phi(r) dr,$$

where $n \cdot \omega_n = \sigma_{n-1}(\partial B_1(0))$ is the surface volume of the $(n-1)$-dimensional unit sphere and $\omega_n = \lambda^n(B_1(0))$ is the volume of the n-dimensional unit ball.

In particular, we get for the functions $f_\alpha(x) := |x|^\alpha$, $\alpha \in \mathbb{R}$,

$$\begin{aligned} f_\alpha \in \mathcal{L}^1(B_1(0) \setminus \{0\}) &\iff \alpha > -n, \\ f_\alpha \in \mathcal{L}^1(\mathbb{R}^n \setminus B_1(0)) &\iff \alpha < -n. \end{aligned}$$

Proof Using Corollary 16.19 we get that

$$\sigma_{n-1}(\partial B_1(0)) = \int_0^\pi \int_0^\pi \cdots \int_{-\pi}^\pi |J(1, \boldsymbol{\theta})| d\theta_{n-1} \cdots d\theta_2 \, d\theta_1$$

and

$$\begin{aligned}\lambda^n(B_1(0)) &= \int_0^1 \int_0^\pi \int_0^\pi \cdots \int_{-\pi}^{\pi} |J(r,\boldsymbol{\theta})|d\theta_{n-1}\cdots d\theta_2\,d\theta_1\,dr\\ &= \int_0^1 r^{n-1}\,dr \int_0^\pi \int_0^\pi \cdots \int_{-\pi}^{\pi} |J(1,\boldsymbol{\theta})|d\theta_{n-1}\cdots d\theta_2\,d\theta_1\\ &= \frac{1}{n}\,\sigma_{n-1}(\partial B_1(0)).\end{aligned}$$

The integrability of f_α follows now from Example 12.13. $\square$

Although we can determine ω_n with the formulae in the proof of Corollary 16.20, we prefer the following direct approach which we have already used in Example 16.16:

$$\begin{aligned}\left(\sqrt{\pi}\right)^n \overset{16.16}{=} \left(\int e^{-t^2}\,\lambda^1(dt)\right)^n &= \int\cdots\int e^{-(x_1^2+\cdots+x_n^2)}\,\lambda^1(dx_1)\cdots\lambda^1(dx_n)\\ &\overset{16.20}{=} n\omega_n \int_{(0,\infty)} r^{n-1}e^{-r^2}\,\lambda^1(dr).\end{aligned}$$

Since $r^{n-1}e^{-r^2}$ is positive and improperly Riemann integrable, [✍] Riemann and Lebesgue integrals coincide (use Problems 12.9 and 12.35), and we find after a change of variables according to $s = r^2$

$$\left(\sqrt{\pi}\right)^n = n\omega_n \int_0^\infty r^{n-1}e^{-r^2}dr = \omega_n \frac{n}{2}\int_0^\infty s^{n/2-1}e^{-s}ds = \omega_n\frac{n}{2}\Gamma\left(\frac{n}{2}\right),$$

see Example 12.15. Since $\frac{n}{2}\Gamma(\frac{n}{2}) = \Gamma(\frac{n}{2}+1)$, we have finally established the following corollary.

Corollary 16.21 *The unit ball $B_1(0) \subset \mathbb{R}^n$ has the volume $\omega_n = \pi^{n/2}/\Gamma\left(\frac{n}{2}+1\right)$ and its $(n-1)$-dimensional surface volume is $n\cdot\omega_n = 2\pi^{n/2}/\Gamma\left(\frac{n}{2}\right)$.*

Surface Measure on the Sphere

We will now discuss the canonical surface measure on the unit sphere $\mathbb{S}^{n-1} := \{x\in\mathbb{R}^n : |x|^2 = 1\}$ in $\mathbb{R}^n$. For this we introduce spherical coordinates which are slightly simpler than polar coordinates but lead to similar formulae. Set

$$\Phi : \mathbb{R}^n \setminus \{0\} \to (0,\infty)\times\mathbb{S}^{n-1}, \quad x \mapsto (|x|, \phi(x)),$$

where $\phi(x) := x/|x| \in \mathbb{S}^{n-1}$ is the directional unit vector for x. Obviously, Φ is bijective and differentiable, and has a differentiable inverse $\Phi^{-1}(r,s) = r\cdot s$.

Theorem 16.22 *On $\mathscr{B}(\mathbb{S}^{n-1}) = \mathbb{S}^{n-1} \cap \mathscr{B}(\mathbb{R}^n)$ there exists a measure σ_{n-1} which is invariant under rotations and satisfies*

$$\int_{\mathbb{R}^n} u(x)dx = \int_{(0,\infty)} \int_{\mathbb{S}^{n-1}} r^{n-1} u(rs)\, \sigma_{n-1}(ds)dr \tag{16.17}$$

for all $u \in \mathcal{L}^1(\lambda^n)$.

In other words, $\Phi(\lambda^n) = \mu \times \sigma_{n-1}$ where $\mu(dr) = r^{n-1}\;{}_{(0,\infty)}(r)\lambda^1(dr)$; in particular,

$$u \in \mathcal{L}^1(\mathbb{R}^n, \lambda^n) \iff r^{n-1}u(rs) \in \mathcal{L}^1\big((0,\infty) \times \mathbb{S}^{n-1}, \lambda^1 \times \sigma_{n-1}\big).$$

Proof We define σ_{n-1} by

$$\sigma_{n-1}(A) := n\lambda^n\big(\phi^{-1}(A) \cap B_1(0)\big) \quad \forall A \in \mathscr{B}(\mathbb{S}^{n-1}),$$

which is an image measure, and hence a measure, see Theorem 7.6. Since ϕ^{-1} and λ^n are invariant w.r.t. rotations around the origin, see Theorem 7.9, it is obvious that σ_{n-1} inherits this property, too.

Both Φ and Φ^{-1} are continuous, and hence measurable. Therefore,

$$\Phi^{-1}\big(\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{S}^{n-1})\big) \subset \mathscr{B}(\mathbb{R}^n) \quad \text{and} \quad \Phi\big(\mathscr{B}(\mathbb{R}^n)\big) \subset \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{S}^{n-1}),$$

which shows that $\mathscr{B}(\mathbb{R}^n) = \Phi^{-1}(\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{S}^{n-1}))$. In order to see (16.17), fix $A \in \mathscr{B}(\mathbb{S}^{n-1})$ and consider first the set $B := \{x \in \mathbb{R}^n : |x| \in [a,b),\ \phi(x) \in A\} = \phi^{-1}(A) \cap \{x : a \leqslant |x| < b\}$, which is clearly a Borel set of $\mathbb{R}^n$. Thus

$$\begin{aligned}
\lambda^n(B) &= \lambda^n\big(\phi^{-1}(A) \cap \{x : a \leqslant |x| < b\}\big) \\
&= \lambda^n\big(\phi^{-1}(A) \cap B_b(0)\big) - \lambda^n\big(\phi^{-1}(A) \cap B_a(0)\big) \\
&= b^n\lambda^n\big(\phi^{-1}(A) \cap B_1(0)\big) - a^n\lambda^n\big(\phi^{-1}(A) \cap B_1(0)\big) \\
&= (b^n - a^n)\lambda^n\big(\phi^{-1}(A) \cap B_1(0)\big),
\end{aligned}$$

where we use that $\lambda^n(a \cdot B) = a^n\lambda^n(B)$, see Theorem 7.10 or Problems 5.9 and 7.10, and that ϕ^{-1} is invariant under dilations. This shows that

$$\begin{aligned}
\lambda^n(B) = \frac{1}{n}(b^n - a^n)\sigma_{n-1}(A) &= \int_{[a,b)} r^{n-1}\sigma_{n-1}(A)\lambda^1(dr) \\
&= \mu \times \sigma_{n-1}\big([a,b) \times A\big).
\end{aligned}$$

The family $\{[a,b) \times A : a < b,\ A \in \mathscr{B}(\mathbb{S}^{n-1})\}$ generates $\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{S}^{n-1})$, see Lemma 14.3, and satisfies the conditions of the uniqueness theorem (Theorem 5.7), so the above relation extends to all sets $B' \in \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{S}^{n-1})$. Since

$\mathscr{B}(\mathbb{R}^n) = \Phi^{-1}(\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{S}^{n-1}))$, we have $B' = \Phi(B)$ for some $B \in \mathscr{B}(\mathbb{R}^n)$, so that

$$\lambda^n(B) = \lambda^n(\Phi^{-1} \circ \Phi(B)) = \lambda^n(\Phi^{-1}(B')) = \mu \times \sigma_{n-1}(B').$$

All other assertions follow now from Theorem 15.1 on image integrals and Fubini's theorem (Theorem 14.9). $\square$

Problems

16.1. Let $F, F_1, F_2, F_3, \ldots$ be F_σ-sets in $\mathbb{R}^n$. Show that

(i) $F_1 \cap F_2 \cap \cdots \cap F_N$ is for every $N \in \mathbb{N}$ an F_σ-set;
(ii) $\bigcup_{n\in\mathbb{N}} F_n$ is an F_σ-set;
(iii) F^c and $\bigcap_{n\in\mathbb{N}} F_n^c$ are G_δ-sets;
(iv) all closed sets are F_σ-sets.

16.2. Prove the following corollary to Lemma 16.12: *Lebesgue measure λ^n on $\mathbb{R}^n$ is outer regular, i.e.*

$$\lambda^n(B) = \inf\{\lambda^n(U) : U \supset B,\ U \text{ open}\} \qquad \forall B \in \mathscr{B}(\mathbb{R}^n),$$

and inner regular, i.e.

$$\begin{aligned} \lambda^n(B) &= \sup\{\lambda^n(F) : F \subset B,\ F \text{ closed}\} &&\forall B \in \mathscr{B}(\mathbb{R}^n) \\ &= \sup\{\lambda^n(K) : K \subset B,\ K \text{ compact}\} &&\forall B \in \mathscr{B}(\mathbb{R}^n). \end{aligned}$$

16.3. **Completion (6).** Combine Problems 16.2 and 11.6 to show that the completion $\bar{\lambda}^n$ of n-dimensional Lebesgue measure is again inner and outer regular.

16.4. Let C be Cantor's ternary set, see page 4 and Problem 7.12.

(i) Show that $C - C := \{x - y : x, y \in C\}$ is the interval $[-1, 1]$.
(ii) Show that this proves that the result of Corollary 16.14 is the best possible.

[Hint: (i) Use the ternary expansion of $x \in C$ from Problem 7.12 and show that $\frac{1}{2}(C - C + 1) = [0, 1]$. (ii) Consider the map $\mathbb{R}^2 \ni (x, y) \mapsto \alpha(x, y) := x + y \in \mathbb{R}$.]

16.5. Consider the Borel σ-algebra $\mathscr{B}[0, \infty)$ and write $\lambda = \lambda^1|_{[0,\infty)}$ for Lebesgue measure on the half-line $[0, \infty)$.

(i) Show that $\mathscr{G} := \{[a, \infty) : a \geqslant 0\}$ generates $\mathscr{B}[0, \infty)$.
(ii) Show that $\mu(B) := \int_B \mathbb{1}_{[2,4]}\lambda(dx)$ and $\rho(B) := \mu(5 \cdot B)$, $B \in \mathscr{B}[0, \infty)$ are measures on $\mathscr{B}[0, \infty)$ such that $\rho|_{\mathscr{G}} \leqslant \mu|_{\mathscr{G}}$ but $\rho \leqslant \mu$ fails.

Why does this not contradict Lemma 16.6?

16.6. Use Jacobi's transformation formula to recover Theorem 5.8(i), Problem 5.9 and Theorem 7.10. In particular, for all integrable functions $u : \mathbb{R}^n \to [0, \infty)$

$$\begin{aligned} \int u(x+y)\lambda^n(dx) &= \int u(x)\lambda^n(dx) &&\forall y \in \mathbb{R}^n, \\ \int u(tx)\lambda^n(dx) &= \frac{1}{t^n}\int u(x)\lambda^n(dx) &&\forall t > 0, \\ \int u(Ax)\lambda^n(dx) &= \frac{1}{|\det A|}\int u(x)\lambda^n(dx) &&\forall A \in \mathbb{R}^{n\times n},\ \det A \neq 0. \end{aligned}$$

The left-hand side of any of the above equalities exists and is finite if, and only if, the right-hand side exists and is finite.

Why can't we use Theorem 16.4 to *prove* these formulae?

16.7. **Arc-length.** Let $f:\mathbb{R}\to\mathbb{R}$ be a twice continuously differentiable function and denote by $\Gamma_f := \{(t,f(t)) : t\in\mathbb{R}\}$ its graph. Define a function $\Phi:\mathbb{R}\to\mathbb{R}^2$ by $\Phi(x) := (x,f(x))$. Then

(i) $\Phi:\mathbb{R}\to\Gamma_f$ is a C^1-diffeomorphism and $|D\Phi(x)| = \sqrt{1+(f'(x))^2}$;

(ii) $\sigma := \Phi(|D\Phi|\,\lambda^1)$ is a measure on Γ_f;

(iii) $\int_{\Gamma_f} u(x,y)d\sigma(x,y) = \int_{\mathbb{R}} u(t,f(t))\sqrt{1+(f'(t))^2}\,d\lambda^1(t)$ with the understanding that, whenever one side of the equality makes sense (measurability!) and is finite, so does the other.

The measure σ is called *canonical surface measure* on Γ_f. This name is justified by the following compatibility property w.r.t. λ^2. Let $n(x)$ be a unit normal vector to Γ_f at point $(x,f(x))$ and define a map $\widetilde{\Phi}:\mathbb{R}\times\mathbb{R}\to\mathbb{R}^2$ by $\widetilde{\Phi}(x,r) := \Phi(x)+rn(x)$.

(iv) Then $n(x) = (-f'(x),1)/\sqrt{1+(f'(x))^2}$ and

$$\det D\widetilde{\Phi}(x,r) = \sqrt{1+(f'(x))^2} - \frac{rf''(x)}{1+(f'(x))^2}.$$

Conclude that for every compact interval $[c,d]$ there exists some $\epsilon>0$ such that $\widetilde{\Phi}|_{(c,d)\times(-\epsilon,\epsilon)}$ is a C^1-diffeomorphism.

(v) Let $C\subset\Gamma_{f|(c,d)}$ and $r<\epsilon$ with ϵ as in (iv). Make a sketch of the set $C(r) := \widetilde{\Phi}\big(\Phi^{-1}(C)\times(-r,r)\big)$ and show that it is Borel measurable.

(vi) Show that for every $x\in(c,d)$

$$\lim_{r\downarrow 0}\frac{1}{2r}\int_{(-r,r)}\big|\det D\widetilde{\Phi}(x,s)\big|\,\lambda^1(ds) = \big|\det D\widetilde{\Phi}(x,0)\big|.$$

(vii) Use the general transformation theorem, Tonelli's theorem, (vi) and dominated convergence to show that

$$\lim_{r\downarrow 0}\frac{1}{2r}\lambda^2(C(r)) = \int_{\Phi^{-1}(C)}|\det D\widetilde{\Phi}(x,0)|\lambda^1(dx).$$

(viii) Conclude that $\int\sqrt{1+(f'(t))^2}\,dt$ is the arc-length of the graph of Γ_f.

16.8. Let $\Phi:\mathbb{R}^d\to M\subset\mathbb{R}^n$, $d\leqslant n$, be a C^1-diffeomorphism.

(i) Show that $\lambda_M := \Phi\big(|\det D\Phi|\lambda^d\big)$ is a measure on M. Find a formula for $\int_M u\,d\lambda_M$.

(ii) Show that for a dilation $\theta_r:\mathbb{R}^n\to\mathbb{R}^n$, $x\mapsto rx$, $r>0$, we have

$$\int_M u(r\xi)r^n\,d\lambda^n(\xi) = \int_{\theta_r(M)} u(\xi)d\lambda_M(\xi).$$

(iii) Let $M = \{\|x\| = r\} = \mathbb{S}_r^{n-1}$ be a sphere in $\mathbb{R}^n$, so that $d = n-1$. Show that for every integrable $u\in\mathcal{L}^1(\mathbb{R}^n)$ and $\sigma := \lambda_M$

$$\begin{aligned}\int u(x)\lambda^n(dx) &= \int_{(0,\infty)}\int_{\{\|x\|=r\}} u(x)\sigma(dx)\lambda^1(dr)\\ &= \int_{(0,\infty)}\int_{\{\|x\|=1\}} r^{n-1}u(rx)\sigma(dx)\lambda^1(dr).\end{aligned}$$

Remark. With somewhat more effort it is possible to show the analogue of the approximation formula in Problem 16.7(vii) for λ_M; all that changes are technical details, the idea of the proof is the same, see Stroock [54, pp. 94–101] for a nice presentation.

16.9. In Example 12.15 we introduced Euler's gamma function:

$$\Gamma(t) = \int_{(0,\infty)} x^{t-1} e^{-x} \lambda^1(dx).$$

Show that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

16.10. **3D polar coordinates.** Define $\Phi : [0,\infty) \times [0, 2\pi) \times [-\pi/2, \pi/2) \to \mathbb{R}^3$ by

$$\Phi(r, \theta, \omega) := \big(r\cos\theta\cos\omega, r\sin\theta\cos\omega, r\sin\omega\big).$$

Show that $|\det D\Phi(r, \theta, \omega)| = r^2 \cos\omega$ and find the integral formula for the coordinate change from Cartesian to polar coordinates $(x, y, z) \rightsquigarrow (r, \theta, \omega)$.

16.11. **Euler's Inegrals.** Euler's gamma and beta functions are the following parameter-dependent integrals for $x, y > 0$:

$$\Gamma(x) := \int_0^\infty e^{-t} t^{x-1}\, dt \quad \text{and} \quad B(x,y) := \int_0^1 t^{x-1}(1-t)^{y-1}\, dt.$$

(i) Show that

$$\Gamma(x)\Gamma(y) = 4\int_{(0,\infty)^2} e^{-u^2-v^2} u^{2x-1} v^{2y-1}\, d(u,v).$$

(ii) Use part (i) to prove that

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

16.12. Compute for $m, n \in \mathbb{N}$ the integral

$$\int_{B_1(0)} x^m y^n\, d(x,y).$$

17

Dense and Determining Sets

We have seen that the spaces of integrable functions $\mathcal{L}^p(\mu)$, $1 \leqslant p < \infty$, can be fairly big and that pth-order integrable functions can be quite irregular. It is important to ask whether we can reduce assertions on $\mathcal{L}^p(\mu)$ to a smaller subset of nicely behaved functions. The blueprint for this is the proof of Theorem 15.8, where we reduce the assertion to continuous functions with compact support. We will also establish for how many functions f one needs to know the integral $\int f\,d\mu$ in order to characterize the measure μ.

Speaking of continuity requires more structure on X, at least a topology. We prefer to work in a metric space X where $d: X \times X \to [0, \infty]$ is the metric which describes the distance between the points $x, y \in X$, see Appendix B. In particular, we have open balls $B_r(x) := \{y \in X : d(x, y) < r\}$, open sets $\mathscr{O}$ and a Borel σ-algebra $\mathscr{B}(X) = \sigma(\mathscr{O})$. If you prefer to work in a Euclidean setting you may take $d(x, y) = |x - y|$ in $\mathbb{R}$ or $d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^2\right)^{1/2}$ in $\mathbb{R}^n$.

Dense Sets

Definition 17.1 Let $(X, \mathscr{A}, \mu)$ be a measure space. A set $\mathcal{D} \subset \mathcal{L}^p(\mu)$, $p \in [0, \infty]$, is called *dense* if for every $u \in \mathcal{L}^p(\mu)$ there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathcal{D}$ such that $\lim_{n\to\infty} \|u - f_n\|_p = 0$.

The following observation is an immediate consequence of the definition of the spaces $\mathcal{L}^p(\mu)$.

Lemma 17.2 *Let $(X, \mathscr{A}, \mu)$ be a measure space. $\mathcal{E}(\mathscr{A}) \cap \mathcal{L}^p(\mu)$ is a dense subset of $\mathcal{L}^p(\mu)$, $1 \leqslant p < \infty$.*

Proof Assume that $u \in \mathcal{L}^p(\mu)$ is positive. Because of the sombrero lemma (Theorem 8.8) there is a sequence of simple functions $f_n \uparrow u$, $f_n \in \mathcal{E}(\mathscr{A})$.

Since $0 \leqslant f_n \leqslant u \in \mathcal{L}^p(\mu)$, we can use the dominated convergence theorem (Theorem 12.2 or 13.9), and find that

$$\lim_{n\to\infty} \|f_n - u\|_p = 0.$$

If $u \in \mathcal{L}^p(\mu)$ is not necessarily positive, we use $u = u^+ - u^-$ and approximate $u^\pm$ separately. [✍] □

Caution Lemma 17.3 does not hold for $p = \infty$. This can be seen on $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$ using $u(x) = \sum_{n=-\infty}^{\infty} \mathbb{1}_{[2n,2n+1]}(x)$.

Lemma 17.2 is mostly of theoretical interest. For practical purposes, it is better to consider continuous functions. Let $C_b(X)$ denote the bounded, continuous functions $u: X \to \mathbb{R}$, and $C_b^+(X) = \{u \in C_b(X) : u \geqslant 0\}$. The distance of $x \in X$ to a set $A \subset X$ is defined as $d(x, A) := \inf_{a\in A} d(x, a)$. From

$$d(x, A) = \inf_{a\in A} d(x, a) \leqslant \inf_{a\in A} (d(x, y) + d(y, a)) = d(x, y) + d(y, A)$$

we conclude that

$$|d(x, A) - d(y, A)| \leqslant d(x, y) \quad \forall x, y \in X, \tag{17.1}$$

which proves that $x \mapsto d(x, A)$ is (Lipschitz) continuous.

Lemma 17.3 *Let μ be a measure on $(X, d, \mathscr{B}(X))$ and $U \subset X$ be an open set such that $\mu(U) < \infty$. For every Borel set $B \in \mathscr{B}(U)$ there exists a sequence $(u_n)_{n\in\mathbb{N}} \subset C_b(X) \cap \mathcal{L}^p(\mu)$, such that* $\lim_{n\to\infty} \|\mathbb{1}_B - u_n\|_p = 0$.

Proof Step 1. Let $U \subset X$ be an open set with finite μ-measure. The functions $u_n(x) := \min\{nd(x, U^c), 1\}$ are bounded and continuous, and $u_n \uparrow \mathbb{1}_U$. From dominated convergence we get $\lim_{n\to\infty} \|u_n - \mathbb{1}_U\|_p = 0$.

Step 2. Let $U \subset X$ be as in the first step. Define

$$\mathscr{D} := \left\{ D \in \mathscr{B}(U) : \exists (u_n^D)_{n\in\mathbb{N}} \subset C_b(X) \cap \mathcal{L}^p(\mu),\ \lim_{n\to\infty} \|u_n^D - \mathbb{1}_D\|_p = 0 \right\}.$$

We are going to show that $\mathscr{D}$ is a Dynkin system for the basis set U.

(D_1) Step 1 shows that $U \in \mathscr{D}$.

(D_2) If $D \in \mathscr{D}$, then there are sequences $u_n^U \to \mathbb{1}_U$ and $u_n^D \to \mathbb{1}_D$ in $\mathcal{L}^p$. Therefore, $u_n^U - u_n^D \to \mathbb{1}_U - \mathbb{1}_D = \mathbb{1}_{U\setminus D}$ in $\mathcal{L}^p$, and we get $D^c = U \setminus D \in \mathscr{D}$.

(D_3) Assume that $(D_k)_{k\in\mathbb{N}} \subset \mathscr{D}$ is a sequence of mutually disjoint sets and define $D := \bigcupdot_k D_k$. For fixed $\epsilon > 0$ and every $k \in \mathbb{N}$ we find a function

$$u_\epsilon^{D_k} \in C_b(X) \cap \mathcal{L}^p(\mu) : \left\| \mathbb{1}_{D_k} - u_\epsilon^{D_k} \right\|_p \leqslant \epsilon 2^{-k}.$$

Since $\sum_{k=1}^{n} \mathbb{1}_{D_k} \uparrow \mathbb{1}_D \in \mathcal{L}^p(\mu)$, there is some $N(\epsilon) \in \mathbb{N}$ such that

$$\left\| \mathbb{1}_D - \sum_{k=1}^{N(\epsilon)} \mathbb{1}_{D_k} \right\|_p \leqslant \epsilon$$

(use e.g. the dominated convergence theorem). Therefore,

$$\begin{aligned}\left\| \mathbb{1}_D - \sum_{k=1}^{N(\epsilon)} u_\epsilon^{D_k} \right\|_p &\leqslant \left\| \mathbb{1}_D - \sum_{k=1}^{N(\epsilon)} \mathbb{1}_{D_k} \right\|_p + \left\| \sum_{k=1}^{N(\epsilon)} \mathbb{1}_{D_k} - \sum_{k=1}^{N(\epsilon)} u_\epsilon^{D_k} \right\|_p \\ &\leqslant \epsilon + \sum_{k=1}^{N(\epsilon)} \left\| \mathbb{1}_{D_k} - u_\epsilon^{D_k} \right\|_p \\ &\leqslant \epsilon + \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = 2\epsilon.\end{aligned}$$

Since $\sum_{k=1}^{N(\epsilon)} u_\epsilon^{D_k} \in C_b(X)$, this proves that $D \in \mathscr{D}$.

Step 3. From Step 1 we know that the open sets $U \cap \mathscr{O} \subset \mathscr{D}$. Since $U \cap \mathscr{O}$ is a generator of $\mathscr{B}(U)$ which is stable under finite intersections, we see from Theorem 5.5 that $\mathscr{D} = \sigma(U \cap \mathscr{O}) = \mathscr{B}(U)$, and the claim follows from the definition of the family $\mathscr{D}$. □

The essential ingredient in Lemma 17.3 is the assumption that we can include every Borel set B with finite measure into an open set $U \supset B$ with finite measure $\mu(U) < \infty$. This is, in general, a non-trivial property.

Theorem 17.4 *Let μ be a finite measure on $(X, d, \mathscr{B}(X))$. Then $C_b(X) \subset \mathcal{L}^p(\mu)$ is dense.*

Proof If $w \in C_b(X)$ or $w \in \mathcal{E}(\mathscr{B}(X))$, then $w \in \mathcal{L}^p(\mu)$. This follows at once from

$$\int |w|^p \, d\mu \leqslant \|w\|_\infty^p \mu(X) < \infty.$$

Let $f \in \mathcal{E}(\mathscr{B}(X))$ be given by the standard representation $f = \sum_{m=0}^{M} \alpha_m \mathbb{1}_{B_m}$. Taking $U = X$ in Lemma 17.3 we find sequences $\left(\phi_n^{B_m}\right)_{n\in\mathbb{N}} \subset C_b(X)$ such that

$$\lim_{n\to\infty} \left\| \phi_n^{B_m} - \mathbb{1}_{B_m} \right\|_p = 0, \quad 0 \leqslant m \leqslant M.$$

Consequently, $\lim_{n\to\infty} \sum_{m=0}^{M} \alpha_m \phi_n^{B_m} = f$ in $\mathcal{L}^p(\mu)$.

Fix $\epsilon > 0$. If $u \in \mathcal{L}^p(\mu)$, then Lemma 17.2 shows that there must exist some $f_\epsilon \in \mathcal{E}(\mathscr{B}(X))$ such that

$$\|u - f_\epsilon\|_p < \epsilon.$$

Thus, the above argument proves that there is some $\phi_\epsilon \in C_b(X) \cap \mathcal{L}^p(\mu)$ with

$$\|f_\epsilon - \phi_\epsilon\|_p < \epsilon,$$

and the claim follows as $\|u - \phi_\epsilon\|_p \leqslant \|u - f_\epsilon\|_p + \|f_\epsilon - \phi_\epsilon\|_p \leqslant 2\epsilon$. □

We can relax the finiteness assumption on the measure μ to a 'local' finiteness on large balls.

Corollary 17.5 *Let μ be a measure on $(X, d, \mathscr{B}(X))$ such that $\mu(B_R(0)) < \infty$ for every $R > 0$. Then $C_b(X) \cap \mathcal{L}^p(\mu)$ is dense in $\mathcal{L}^p(\mu)$.*

Proof Step 1. Fix $\epsilon > 0$ and $u \in \mathcal{L}^p(\mu)$. Using the dominated convergence theorem (Theorem 12.2 or 13.9) we see that there is some $R(\epsilon) > 0$ such that

$$\|u - u\mathbb{1}_{B_R(0)}\|_p < \epsilon \quad \forall R \geqslant R(\epsilon).$$

Step 2. Fix $R \geqslant R(\epsilon)$. We can now apply Theorem 17.4 with $\mu_{4R} := \mathbb{1}_{B_{4R}(0)}\mu$, $u\mathbb{1}_{B_R(0)}$ and $\mathcal{L}^p(X, \mu_{4R})$. This shows that there is some $\phi_\epsilon \in C_b(X) \cap \mathcal{L}^p(X, \mu_{4R})$ such that

$$\|(u\mathbb{1}_{B_R(0)} - \phi_\epsilon)\mathbb{1}_{B_{4R}(0)}\|_{\mathcal{L}^p(\mu)} = \|u\mathbb{1}_{B_R(0)} - \phi_\epsilon\|_{\mathcal{L}^p(\mu_{4R})} < \epsilon.$$

Without loss of generality we can assume that $\operatorname{supp}\phi_\epsilon \subset \overline{B_{2R}(0)}$, otherwise we could multiply ϕ_ϵ by the function

$$\chi_R(x) := \frac{d(x, B^c_{2R}(0))}{d(x, B^c_{2R}(0)) + d(x, B_R(0))},$$

which is bounded and continuous, and satisfies $\chi_R|_{B_R(0)} \equiv 1$ and $\chi_R|_{B^c_{2R}(0)} \equiv 0$. Thus,

$$\|u\mathbb{1}_{B_R(0)} - \phi_\epsilon\|_{\mathcal{L}^p(\mu)} \leqslant \|(u\mathbb{1}_{B_R(0)} - \phi_\epsilon)\mathbb{1}_{B_{4R}(0)}\|_{\mathcal{L}^p(\mu)} < \epsilon.$$

Step 3. The assertion of the corollary follows now from

$$\|u - \phi_\epsilon\|_{\mathcal{L}^p(\mu)} \leqslant \|u - u\mathbb{1}_{B_R(0)}\|_{\mathcal{L}^p(\mu)} + \|u\mathbb{1}_{B_R(0)} - \phi_\epsilon\|_{\mathcal{L}^p(\mu)} < 2\epsilon. \qquad \square$$

We can replace the condition $\mu(B_R(0)) < \infty$ with a topological assumption which is known as outer regularity, see Appendix H.

Definition 17.6 Let μ be a measure on the space $(X, d, \mathscr{B}(X))$. The measure is said to be *outer regular*, if

$$\mu(B) = \inf\{\mu(U) : U \supset B,\ U \text{ open}\} \quad \forall B \in \mathscr{B}(X). \tag{17.2}$$

Theorem 17.7 *Let μ be an outer regular measure on the space $(X, d, \mathscr{B}(X))$. Then $C_b(X) \cap \mathcal{L}^p(\mu)$ is dense in $\mathcal{L}^p(\mu)$.*

Proof Fix $\epsilon > 0$ and $u \in \mathcal{L}^p(\mu)$. From Lemma 17.2 we know that there is some $f_\epsilon \in \mathcal{E}(\mathscr{B}(X)) \cap \mathcal{L}^p(\mu)$ such that

$$\|u - f_\epsilon\|_p < \epsilon.$$

A simple function $f \in \mathcal{E}(\mathscr{B}(X))$ is in $\mathcal{L}^p(\mu)$ if, and only if, $\mu\{f \neq 0\} < \infty$. Therefore, outer regularity guarantees that there is some open set $U \supset \{f \neq 0\}$ with finite μ-measure. As in the second half of the proof of Theorem 17.4, we see that there is some $\phi_\epsilon \in C_b(X) \cap \mathcal{L}^p(\mu)$ such that

$$\|f_\epsilon - \phi_\epsilon\|_p < \epsilon.$$

This finishes the proof, since $\|u - \phi_\epsilon\|_p \leqslant \|u - f_\epsilon\|_p + \|f_\epsilon - \phi_\epsilon\|_p \leqslant 2\epsilon$. □

Assume now that the metric space (X, d) is *σ-compact*, i.e. there is a sequence of compact sets $K_i \uparrow X$. This is, for example, the case if $X = \mathbb{R}^n$ (use $K_i = \overline{B_i(0)}$), or if X is locally compact and contains a countable dense subset (see Problem 17.6(ii)). Denote by $C_c(X)$ the continuous functions with compact support $\operatorname{supp} u = \overline{\{u \neq 0\}}$.

Caution In contrast to $\mathbb{R}^n$, we cannot expect in a general metric space (X, d) that the closures of the balls $\overline{B_R(0)}$ are compact sets.

Theorem 17.8 *Let μ be a measure on the σ-compact space $(X, d, \mathscr{B}(X))$ and assume that either μ is outer regular or it satisfies $\mu(B_R(0)) < \infty$ for all $R > 0$. If $\mu(K) < \infty$ for all compact sets $K \subset X$, then $C_c(X) \subset \mathcal{L}^p(\mu)$ is dense.*

Proof Let $u \in C_c(X)$ and set $K := \operatorname{supp} u$. From

$$\int |u|^p \, d\mu = \int_K |u|^p \, d\mu \leqslant \|u\|_\infty^p \mu(K) < \infty$$

we see that $C_c(X) \subset \mathcal{L}^p(\mu)$. Let K_n be a sequence of compact sets such that $K_n \uparrow X$. Observe that the functions

$$\chi_n(x) := \frac{d(x, K_{n+1}^c)}{d(x, K_{n+1}^c) + d(x, K_n)}$$

are continuous, see (17.1), and satisfy $\operatorname{supp} \chi_n \subset K_{n+1}$, $\chi_n|_{K_n} \equiv 1$ and $\chi_n \uparrow 1$.

Because of Corollary 17.5 or Theorem 17.7 there must exist a sequence $(u_n)_n \subset C_b(X) \cap \mathcal{L}^p(\mu)$ such that $u_n \to u$ in $\mathcal{L}^p$. Therefore,

$$\begin{aligned}\|u - u_n\chi_n\|_p &\leqslant \|u - u\chi_n\|_p + \|u\chi_n - u_n\chi_n\|_p \\ &\leqslant \|(1 - \chi_n)u\|_p + \|u - u_n\|_p \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

The first term converges because of the dominated convergence theorem, the second because of the way we have constructed the sequence $(u_n)_{n\in\mathbb{N}}$. The claim follows since $\operatorname{supp}(u_n\chi_n)\subset K_{n+1}$ is compact. $\square$

If we combine Theorem 17.8 with the Friedrichs mollifier (Theorem 15.11), then we see that the test functions are dense in $\mathcal{L}^p(\mu)$.

Corollary 17.9 *Let μ be a measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ which assigns finite measure to compact sets. The compactly supported, smooth functions $C_c^\infty(\mathbb{R}^n)$ are dense in $\mathcal{L}^p(\mu)$.*

Proof Fix $\epsilon>0$ and $u\in\mathcal{L}^p(\mu)$. Since the closed balls $\overline{B_R(0)}$ are compact subsets of $\mathbb{R}^n$, we can use Theorem 17.8 to construct some $u_\epsilon\in C_c(\mathbb{R}^n)$ such that

$$\|u-u_\epsilon\|_p<\epsilon.$$

Now we use the Friedrichs mollifier from Lemma 15.10 and Theorem 15.11: $\phi_\delta\star u_\epsilon$ is a C^∞-function with compact support contained in $\operatorname{supp}u_\epsilon+\overline{B_\delta(0)}$. Thus,

$$\|u-\phi_\delta\star u_\epsilon\|_p\leqslant\|u-u_\epsilon\|_p+\|u_\epsilon-\phi_\delta\star u_\epsilon\|_p.$$

Letting first $\delta\to 0$ and then $\epsilon\to 0$, we get $\lim_{\epsilon\to 0}\lim_{\delta\to 0}\|u-\phi_\delta\star u_\epsilon\|_p=0$. Thus, a standard diagonal argument shows that $\phi_{\delta(\epsilon)}\star u_\epsilon\to u$ in $\mathcal{L}^p$ as $\epsilon\to 0$. $\square$

Determining Sets

We will now turn to the question of how many functions one has to integrate in order to characterize a measure.

Definition 17.10 Let μ,ν be measures on a measurable space $(X,\mathscr{A})$. A set $\mathcal{D}\subset\mathcal{L}^1_{\mathbb{C}}(\mu)\cap\mathcal{L}^1_{\mathbb{C}}(\nu)$ is called *determining*, if

$$\int f\,d\mu=\int f\,d\nu\quad\forall f\in\mathcal{D}\implies\mu=\nu.$$

In the definition of a determining set and the following examples it is always assumed that functions in $\mathcal{D}$ are integrable for the measures μ and ν.

Example 17.11 (i) $\mathcal{D}=\{\mathbb{1}_G : G\in\mathscr{G}\}$ is a determining family if $\mathscr{A}=\sigma(\mathscr{G})$ for any $\cap$-stable generator $\mathscr{G}$ containing a sequence $(G_k)_{k\in\mathbb{N}}\subset\mathscr{G}$ such that $G_k\uparrow X$.
Indeed: use the uniqueness theorem for measures, Theorem 5.7.

(ii) $\mathcal{D} = \{\mathbb{1}_K : K \subset \mathbb{R}^n \text{ compact}\}$ is determining on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$.
Indeed: this follows from (i) since the compact sets generate the Borel σ-algebra $\mathscr{B}(\mathbb{R}^n)$ and $[-k,k]^n \uparrow \mathbb{R}^n$.

(iii) $\mathcal{D} = \left\{\mathbb{1}_{H_1 \cap \dots \cap H_n} : n \geqslant 1, H_k \in \mathscr{H} \cup \{X\}\right\}$ is determining for $\mathscr{A} = \sigma(\mathscr{H})$.
Indeed: use (i) with $\mathscr{G} = \{H_1 \cap \dots \cap H_n : n \geqslant 1,\ H_k \in \mathscr{H} \cup \{X\}\}$.

(iv) $\mathcal{D} = \mathcal{M}_b^+(\mathscr{A})$ is determining.
Indeed: observe that $\mathbb{1}_A \in \mathcal{M}_b^+(\mathscr{A})$ for all $A \in \mathscr{A}$.

(v) If $\mathcal{D}$ is determining, so is every $\mathcal{D}' \supset \mathcal{D}$.

In order to get more interesting examples, let us assume that (X,d) is a metric space. As before we write $C_b(X)$ and $C_c(X)$ for the bounded, resp., compactly supported continuous functions $u : X \to \mathbb{R}$.

Theorem 17.12 *Let (X,d) be a σ-compact metric space and μ, ν be measures which are finite on compact sets. Then the families $C_c^+(X)$ and $C_c(X)$ are determining.*

If μ, ν are finite measures with $\mu(X) = \nu(X)$, σ-compactness is not needed.

Proof Let μ, ν be measures with $\int u\,d\mu = \int u\,d\nu$ for all $u \in C_c(X)$; in particular, $\mathbb{1}_K \in \mathcal{L}^1(\mu) \cap \mathcal{L}^1(\nu)$ for all compact sets K. From Theorem 17.8 we know that $C_c(X)$ is dense in $\mathcal{L}^1(\mu) \cap \mathcal{L}^1(\nu)$, i.e. there is a sequence $(u_n)_{n \in \mathbb{N}}$ such that $u_n \to \mathbb{1}_K$ both in $\mathcal{L}^1(\mu)$ and $\mathcal{L}^1(\nu)$. Therefore,

$$\mu(K) = \int \mathbb{1}_K\,d\mu = \lim_{n\to\infty} \int u_n\,d\mu \overset{\text{assumption}}{=} \lim_{n\to\infty} \int u_n\,d\nu = \int \mathbb{1}_K\,d\nu = \nu(K).$$

Since X is σ-compact, there is an increasing sequence of compacts $K_n \uparrow X$. This means that the assumptions of the uniqueness theorem of measures (Theorem 5.7) are satisfied, and we get $\mu = \nu$. If $\mu(X) = \nu(X) < \infty$, we can apply Theorem 5.7 to the family {compact sets, X} and take the increasing sequence $K_n := X \uparrow X$. □

Since $C_c^\infty(\mathbb{R}^n)$ is dense in $\mathcal{L}^1(\mu)$ (for measures μ which are finite on compact sets), see Corollary 17.9, a small change in the proof of Theorem 17.12 shows the following.

Corollary 17.13 *On $\mathbb{R}^n$ the family $C_c^\infty(\mathbb{R}^n)$ is determining for all measures which are finite on compact sets.*

We close this section with the proof that the functions

$$e_\xi(x) = e^{ix\cdot\xi},\ \ x, \xi \in \mathbb{R}^n,\ x \cdot \xi = \sum_{k=1}^n x_k \xi_k$$

are determining. For this we need an auxiliary result.

Lemma 17.14 *For all $x, \xi \in \mathbb{R}^n$ and $t > 0$*

$$g_t(x) := (2\pi t)^{-n/2}\, e^{-|x|^2/2t} \tag{17.3}$$

$$\breve{g}_t(\xi) := \int_{\mathbb{R}^n} g_t(x) e_\xi(x) dx = e^{-t|\xi|^2/2}. \tag{17.4}$$

Proof Consider first the one-dimensional case. We use the differentiability lemma (Theorem 12.2) and integration by parts to get

$$\begin{aligned}
\frac{d}{d\xi}\breve{g}_t(\xi) &= \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} e^{-x^2/2t}\, \underbrace{\frac{d}{d\xi} e^{ix\xi}}_{=ixe^{ix\xi}}\, dx \\
&= \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} (-it) \frac{d}{dx} \left[e^{-x^2/2t} \right] e^{ix\xi}\, dx \\
&\overset{\text{parts}}{=} \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} (it) e^{-x^2/2t} \frac{d}{dx} e^{ix\xi}\, dx \\
&= \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} (-t\xi) e^{-x^2/2t} e^{ix\xi}\, dx \\
&= -(t\xi)\breve{g}_t(\xi).
\end{aligned}$$

This differential equation has the unique solution $\breve{g}_t(\xi) = \breve{g}_t(0)e^{-t\xi^2/2}$. From Example 14.11 (or Example 16.16) we know that

$$\breve{g}_t(0) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} e^{-x^2/2t}\, dx \underset{dy=\sqrt{t}dx}{\overset{y=\sqrt{t}x}{=}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2}\, dy = 1,$$

and we are done if $n = 1$. In higher dimensions we use Fubini's theorem to get

$$\begin{aligned}
\int_{\mathbb{R}^n} (2\pi t)^{-n/2} e^{-|x|^2/2t} e^{ix\cdot\xi} dx &= \prod_{k=1}^n \int_{\mathbb{R}} e^{-x_k^2/2t} e^{ix_k\xi_k} \frac{dx_k}{\sqrt{2\pi t}} \\
&= \prod_{k=1}^n e^{-t\xi_k^2/2} = e^{-t|\xi|^2/2}.
\end{aligned}$$
□

Theorem 17.15 $\mathcal{D} = \{e_\xi(\cdot) : \xi \in \mathbb{R}^n\}$ *is a determining family in* $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$.

Proof Let $u \in C_c^+(\mathbb{R}^n)$ and let μ, ν be finite measures. The finiteness of the measures μ, ν is needed in order to guarantee that the integrals $\int e_\xi\, d\mu$, $\int e_\xi\, d\nu$ exist.

From Lemma 17.14 we get

$$\begin{aligned}
\int u \star \check{g}_t \, d\mu &= \iint u(y) e^{-\frac{1}{2}t|x-y|^2} \, dy \, \mu(dx) \\
&= \iint u(y) \int g_t(\eta) e_{x-y}(\eta) d\eta \, dy \, \mu(dx) \\
&\overset{\text{Fubini}}{=} \iint u(y) g_t(\eta) \underbrace{\int e_{x-y}(\eta) \mu(dx)}_{=\int e_{x-y}(\eta)\nu(dx)} \, d\eta \, dy.
\end{aligned}$$

Using that $\int e_\xi \, d\mu = \int e_\xi \, d\nu$ we see that

$$\int u \star \check{g}_t \, d\mu = \int u \star \check{g}_t \, d\nu \quad \forall u \in C_c(\mathbb{R}^n), \quad t > 0.$$

This allows us to apply the dominated convergence theorem

$$\begin{aligned}
t^{n/2} \int u \star \check{g}_t \, d\mu &= \iint u(x-y) t^{n/2} e^{-\frac{1}{2}t|y|^2} dy \, \mu(dx) \\
&\underset{dz = t^{n/2}\,dy}{\overset{z=\sqrt{t}y}{=}} \iint u\Big(x - t^{-1/2} z\Big) e^{-\frac{1}{2}|z|^2} dz \, \mu(dx) \\
&\xrightarrow[t\to\infty]{\text{dom. conv.}} \int u(x) \mu(dx) \underbrace{\int e^{-\frac{1}{2}|z|^2} dz}_{=(2\pi)^{n/2}}
\end{aligned}$$

(we use $\|u\|_\infty \mathbb{1}_{\operatorname{supp} u}(x) e^{-|z|^2/2}$ as integrable majorant) and the claim follows from Theorem 17.12. □

Problems

17.1. Let $(X, \mathscr{A}, \mu)$ be a measure space. Assume that $\mathcal{D} \subset \mathcal{L}^p(\mu)$ is dense. If $\mathcal{C} \subset \mathcal{D}$ is dense w.r.t. the norm $\|\cdot\|_p$, then $\mathcal{C} \subset \mathcal{L}^p(\mu)$ is dense.

17.2. The following exercise provides an independent proof of Theorem 17.12 in $\mathbb{R}^n$. Set $d(x, A) := \inf_{a \in A} |x - a|$ and assume that all measures are finite on compact sets.

(i) (Urysohn's lemma) Let $K \subset \mathbb{R}^n$ be a compact set and $U_k := K + B_{1/k}(0)$. Show that U_k is open and that $u_k(x) := d(x, U_k^c)/(d(x, U_k^c) + d(x, K))$ are continuous, compactly supported functions such that $u_k \downarrow \mathbb{1}_K$.

(ii) Use monotone convergence to show that $\int u \, d\mu = \int u \, d\nu$ for $u \in C_c^+(\mathbb{R}^n)$ implies that $\mu(K) = \nu(K)$ on all compact sets K.

(iii) Use the uniqueness of measures theorem to show that $\mu = \nu$.

(iv) Replace $\mathbb{R}^n$ by (X, d) such that X is locally compact, i.e. each $x \in X$ has a compact neighbourhood. Show that $\overline{U}_k$ from (i) is compact and that the other steps go through without changes.
[Hint: the compactness of K gives $K + B_{1/k}(0) = \bigcup_{i<\infty} K \cup B_{1/k}(x_i)$, with finitely many x_i from the boundary of K.]

17.3. Consider on $(\mathbb{R}, \mathscr{B}(\mathbb{R}), dx)$ the Lebesgue space $\mathcal{L}^p(dx)$, $1 \leqslant p < \infty$. We set $\tau_h f(x) := f(x-h)$, $h \in \mathbb{R}$. Show that

(i) τ_h is an isometry on $\mathcal{L}^p(dx)$,
(ii) $\lim_{h\to 0} \|\tau_h f - f\|_p = 0$ and $\lim_{h\to\infty} \|\tau_h f - f\|_p = 2^{1/p}\|f\|_p$.

17.4. Denote by dx one-dimensional Lebesgue measure and let $f \in \mathcal{L}^1(dx)$. Define the mean value

$$M_h f(x) := \frac{1}{2h}\int_{x-h}^{x+h} f(t)\,dt$$

and show that

(i) $M_h f(x)$ is continuous and $\|M_h f\|_1 \leqslant \|f\|_1$,
(ii) $\lim_{h\to 0} \|M_h f - f\|_1 = 0$.

17.5. Let μ be an outer regular measure on $(X, d, \mathscr{B}(X))$, $1 \leqslant p < \infty$, and denote by $C_{\text{Lip}}(X)$ the Lipschitz continuous functions $u : X \to \mathbb{R}$.

(i) Let $A \in \mathscr{B}(X)$ such that $f = \mathbb{1}_A \in \mathcal{L}^p(\mu)$. Show that for every $\epsilon > 0$ there is some $\phi_\epsilon \in C_{\text{Lip}}(X)$ such that $\|f - \phi_\epsilon\|_p < \epsilon$.
(ii) Show that for $f \in \mathcal{L}^p(\mu)$, $f \geqslant 0$, and $\epsilon > 0$ there is some $\phi_\epsilon \in C_{\text{Lip}}(X)$ satisfying $\|f - \phi_\epsilon\|_p < \epsilon$.
[*Hint:* use (i) and the sombrero lemma, Theorem 8.8.]
(iii) Show that $C_{\text{Lip}}(X) \cap \mathcal{L}^p(\mu)$ is dense in $\mathcal{L}^p(\mu)$.

17.6. Let (X, d) be a metric space which is *separable* (i.e. it contains a countable dense subset) and *locally compact* (i.e. every $x \in X$ has an open neighbourhood U such that $\overline{U}$ is compact); denote by $\mathscr{O}$ the open sets of X and assume that μ is a measure on X equipped with its Borel sets $\mathscr{B}(X) = \sigma(\mathscr{O})$ such that $\mu(K) < \infty$ for all compact sets $K \subset X$.

(i) Show that there is a sequence of open sets $(U_n)_{n\in\mathbb{N}} \subset \mathscr{O}$ such that $\overline{U}_n$ is compact and each $U \in \mathscr{O}$ can be written as a union of sets from the sequence $(U_n)_{n\in\mathbb{N}}$.
(ii) Show that X is σ-compact, i.e. there is a sequence of compact sets $K_n \uparrow X$.
(iii) Set $\mathcal{D} = \operatorname{span}\{\mathbb{1}_U : U = \bigcup_{n\in F} U_n,\ F \subset \mathbb{N} \text{ finite}\}$ and denote by $\overline{\mathcal{D}}$ the closure of $\mathcal{D}$ in $\mathcal{L}^p(\mu)$, $1 < p < \infty$. Show that $\mathbb{1}_U \in \overline{\mathcal{D}}$ for any $U \in \mathscr{O}$ such that $\mu(U) < \infty$.
[*Hint:* a criterion of outer regularity is given in Appendix H.]
(iv) Show that μ is outer regular and that (ii) remains valid for $B \in \mathscr{B}(X)$ such that $\mu(B) < \infty$.
(v) Show that $\overline{\mathcal{D}} = \mathcal{L}^p(\mu)$ and conclude that $\mathcal{L}^p(\mu)$ is separable.

17.7. Lusin's theorem. The following steps furnish a proof of the following result.

Theorem (Lusin). *Let μ be an outer regular measure on the space $(X, d, \mathscr{B}(X))$. For every $f \in \mathcal{L}^p(\mu)$, $1 \leqslant p < \infty$, and $\epsilon > 0$ there is some $\phi_\epsilon \in \mathcal{L}^p(\mu) \cap C_b(X)$, such that*

$$\|\phi_\epsilon\|_\infty \leqslant \|f\|_\infty \leqslant \infty, \quad \mu\{f \neq \phi_\epsilon\} \leqslant \epsilon \quad \textit{and} \quad \|f - \phi_\epsilon\|_p \leqslant \epsilon.$$

(i) Let $A \in \mathscr{B}(X)$ such that $\mathbb{1}_A \in \mathcal{L}^p(\mu)$. Then there is an open set $U \supset A$ such that $\mu(U) < \infty$ and we can construct ϕ_ϵ as in Lemma 17.3.
(ii) Let $f \in \mathcal{L}^p(\mu)$ such that $0 \leqslant f \leqslant 1$. By the sombrero lemma (Corollary 8.9) there is a uniformly convergent sequence of simple functions which can be 'smoothed' using part (i).
(iii) Let $f \in \mathcal{L}^p(\mu)$ such that $c = \|f\|_\infty < \infty$. Apply part (ii) to $f^\pm$.
(iv) Let $f \in \mathcal{L}^p(\mu)$. Apply part (iii) to $f_R := (-R) \vee f \wedge R$.

17.8. Consider Lebesgue measure λ on the space $([a,b], \mathscr{B}[a,b])$ and assume there on that $f \in \mathcal{L}^1([a,b], \lambda)$ satisfies $\int_a^b x^n f(x)dx = 0$ for all $n = 0, 1, 2, \ldots$. Show that $f|_{[a,b]} = 0$ is Lebesgue almost everywhere.
[*Hint:* use Weierstraß' approximation theorem, Theorem 28.6.]

17.9. The steps below show that the family $\epsilon_\lambda(t) := e^{-\lambda t}$, $\lambda, t > 0$, is determining for $([0,\infty), \mathscr{B}[0,\infty))$.

(i) Weierstraß' approximation theorem shows that the polynomials on $[0,1]$ are uniformly dense in $C[0,1]$, see Theorem 28.6.

(ii) Define for $u \in C_c[0,\infty)$ the function $u \circ (-\log) : (0,1] \to \mathbb{R}$ and approximate it with a sequence of polynomials p_n, $n \in \mathbb{N}$.

(iii) Show that $\int \epsilon_\lambda \, d\mu = \int \epsilon_\lambda \, d\nu$ implies $\int p_n(e^{-t})\mu(dt) = \int p_n(e^{-t})\nu(dt)$; thus, $\int u \, d\mu = \int u \, d\nu$.

18

Hausdorff Measure

Hausdorff measures were introduced by Hausdorff in 1919 to measure the size of lower-dimensional objects in Euclidean n-space, e.g. the length of a line in $\mathbb{R}^2$ or the area of an $(n-1)$-dimensional hypersurface embedded in $\mathbb{R}^n$. Such objects are, from the perspective of n-dimensional Lebesgue measure, null sets. A similar phenomenon happens with fractal sets like Cantor's ternary set, see page 4, which is uncountable and fairly big, but nevertheless has Lebesgue 'length' zero. In order to overcome this difficulty, Hausdorff uses Carathéodory's construction, as in Theorem 6.1, but he measures the size of the covers by using a more geometric object, namely a function of the diameter. Recall the notion of an outer measure from Chapter 6: this is a set function $\mu^* : \mathscr{P}(X) \to [0, \infty]$ satisfying

$$\mu^*(\emptyset) = 0, \tag{OM$_1$}$$

$$A \subset B \implies \mu^*(A) \leqslant \mu^*(B), \tag{OM$_2$}$$

$$\mu^*\left(\bigcup_{i\in\mathbb{N}} A_i\right) \leqslant \sum_{i\in\mathbb{N}} \mu^*(A_i). \tag{OM$_3$}$$

Constructing (Outer) Measures

Carathéodory's theorem (Theorem 6.1) is more than an extension theorem. When read appropriately, it provides a method to construct an outer measure and a notion of measurability starting from any set function $C : \mathscr{P}(X) \to [0, \infty]$ such that $C(\emptyset) = 0$. Later on, we will use $X = \mathbb{R}^n$ and $C(A) = |\operatorname{diam} A|^s$. Before you continue reading, you should scrutinize the proof of Theorem 6.1, because we will frequently refer to it. The main observation is that Theorem 6.1 states that we need only a pre-measure μ and a semi-ring $\mathscr{S}$ in order to show that the outer measure *extends* μ and that $\mathscr{A}^*$ is 'large' in the sense that it contains $\sigma(\mathcal{S})$ (Steps 2, 3a, 4). For the construction of μ^* and $\mathscr{A}^*$ this is not needed.

Definition 18.1 Let $\mathcal{S} \subset \mathscr{P}(X)$ be any family of subsets of X containing $\emptyset$, and let $A \subset X$. $\mathscr{C}(A, \mathcal{S})$ denote the family of *countable $\mathcal{S}$-covers*

$$\mathscr{C}(A, \mathcal{S}) := \Big\{ (S_i)_{i\in\mathbb{N}} \subset \mathcal{S} : \underbrace{\textstyle\bigcup_{i\in\mathbb{N}} S_i}_{\text{need not be disjoint or in } \mathcal{S}} \supset A \Big\}. \tag{18.1}$$

For any $C : \mathscr{P}(X) \to [0, \infty]$ with $C(\emptyset) = 0$, $\mu^* : \mathscr{P}(X) \to [0, \infty]$ is the set function defined by

$$\mu^*(A) := \inf \left\{ \sum_{i\in\mathbb{N}} C(S_i) : (S_i)_{i\in\mathbb{N}} \in \mathscr{C}(A, \mathcal{S}) \right\}. \tag{18.2}$$

If A admits no $\mathcal{S}$-cover, we have $\mathscr{C}(A, \mathcal{S}) = \emptyset$ and $\mu^*(A) = \inf \emptyset = \infty$.

The family $\mathscr{A}^*$ of *μ^*-measurable sets* is

$$\mathscr{A}^* := \{ A \subset X : \mu^*(Q) = \mu^*(Q \cap A) + \mu^*(Q \setminus A) \quad \forall Q \subset X \}. \tag{18.3}$$

Definition 18.1 is almost identical to the corresponding definitions on page 40 – except that we do not impose structural requirements on $\mathcal{S}$ and C.

Theorem 18.2 *Let $\mathcal{S}$, $C(\,\cdot\,)$, $\mu^*(\,\cdot\,)$ and $\mathscr{A}^*$ be as in Definition 18.1.*

(i) *μ^* is an outer measure such that $\mu^*(S) \leqslant C(S)$ for all $S \in \mathcal{S}$;*

(ii) *$\mathscr{A}^*$ is a σ-algebra and $\mu^*|\mathscr{A}^*$ is a measure;*

(iii) *μ^* is maximal: if ν^* is another outer measure satisfying $\nu^*(S) \leqslant C(S)$ for all $S \in \mathcal{S}$, then $\nu^*(A) \leqslant \mu^*(A)$ for all $A \subset X$.*

Proof Large parts of the proof of Theorem 6.1 do not require structural assumption on $\mathcal{S}$ and $C = \mu$. In fact, Step 1 (of the proof of Theorem 6.1, pages 40–41) remains literally valid and proves that μ^* is an outer measure. Step 3b (pages 44–45) remains literally valid and proves that $\mathscr{A}^*$ is a σ-algebra and $\mu^*|\mathscr{A}^*$ is a measure.

This proves (i) and (ii). In order to see (iii), assume that the outer measure ν^* satisfies $\nu^*(S) \leqslant C(S)$. If $(S_i)_{i\in\mathbb{N}} \in \mathscr{C}(A, \mathcal{S})$, then

$$\nu^*(A) \overset{(\mathrm{OM}_2)}{\leqslant} \nu^*\Big(\bigcup_{i\in\mathbb{N}} S_i \Big) \overset{(\mathrm{OM}_3)}{\leqslant} \sum_{i=1}^{\infty} \nu^*(S_i) \leqslant \sum_{i=1}^{\infty} C(S_i).$$

We can now take the infimum over all such covers and find $\nu^*(A) \leqslant \mu^*(A)$. $\square$

The problem with the construction in Theorem 18.2 is that we lose control over the size of $\mathscr{A}^*$ – this did not happen in Theorem 6.1 as $\sigma(\mathscr{S}) \subset \mathscr{A}^*$ is guaranteed *both* by $\mathscr{S}$ being a semiring *and* by C being a pre-measure. Things are quite different now.

Example 18.3 Consider $X=\mathbb{R}$, the semi-ring $\mathscr{S}=\mathscr{J}=\{[a,b): a<b\}$ and the set function $C(S):=(\operatorname{diam} S)^\alpha$, $\alpha\in(0,1)$; here, $\operatorname{diam} S=\sup_{x,y}|x-y|$ is the diameter of the set S. Observe that C fails to be a pre-measure on $\mathscr{J}$: for any $\mathscr{J}$-cover $([a_i,b_i))_{i\in\mathbb{N}}$ of $[0,1)$ we have

$$1=\lambda[0,1)\leqslant\lambda\left(\bigcup_{i=1}^\infty[a_i,b_i)\right)\leqslant\sum_{i=1}^\infty(b_i-a_i)$$

and by the concave Jensen inequality for infinite series we get

$$\sum_{i=1}^\infty C[a_i,b_i)=\sum_{i=1}^\infty(b_i-a_i)^\alpha>\left(\sum_{i=1}^\infty(b_i-a_i)\right)^\alpha\geqslant 1.$$

This shows that $\mu^*[0,1)\geqslant 1$ and, incidentally, that C is not σ-additive on $\mathscr{J}$ (just take a partition $\dot{\bigcup}_i[a_i,b_i)=[0,1)$). On the other hand, $\{[0,1)\}$ is a trivial $\mathscr{J}$-cover of $[0,1)$, i.e. $\mu^*[0,1)\leqslant 1$, hence $\mu^*[0,1)=1$.

The same reasoning applies to $[-1,0)$ and shows $\mu^*[-1,0)=1$. Finally, observe that $\{[-1,1)\}$ is a trivial $\mathscr{S}$-cover of $[-1,1)$, hence $\mu^*[-1,1)\leqslant C[-1,1)=2^\alpha$.

Let $A=[0,1)$ and $Q=[-1,1)$. The previous calculations show that

$$\mu^*(Q\cap A)+\mu^*(Q\setminus A)=\mu^*[0,1)+\mu^*[-1,0)=2>2^\alpha\geqslant\mu^*[-1,1)=\mu^*(Q),$$

which means that $[0,1)$ is not μ^*-measurable; in particular, $\mathscr{B}(\mathbb{R})$ is not contained in $\mathscr{A}^*$.

We can overcome the problem discussed in Example 18.3 by the following construction.[1] From now on, (X,d) is a metric space[2] together with its open sets $\mathscr{O}$ and the Borel σ-algebra $\mathscr{B}(X)=\sigma(\mathscr{O})$. By $\operatorname{diam} A:=\sup_{x,y\in A}d(x,y)$ we denote the diameter of the set A.

Definition 18.4 Let (X,d) be a metric space, $C:\mathscr{P}(X)\to[0,\infty]$ with $C(\emptyset)=0$, $\mathscr{S}\subset\mathscr{P}(X)$ any family of subsets of X containing $\emptyset$. For $A\subset X$ and $\epsilon>0$ we

[1] Often called 'Method II' (as opposed to 'Method I' from Theorem 18.2); the terminology is due to Munroe [31].

[2] See Appendix B, but you may also safely think of X as $\mathbb{R}^n$ and $d(x,y)=|x-y|$.

denote by $\mathscr{C}_\epsilon(A,\mathcal{S})$ the family of *countable $\mathcal{S}$–ϵ-covers*

$$\mathscr{C}_\epsilon(A,\mathcal{S}) := \Big\{(S_i)_{i\in\mathbb{N}} \subset \mathcal{S} : \bigcup_{i\in\mathbb{N}} S_i \supset A,\ \operatorname{diam} S_i \leqslant \epsilon\Big\}. \tag{18.4}$$

If $\mu_\epsilon^* : \mathscr{P}(X) \to [0,\infty]$ is the outer measure constructed in Theorem 18.2 for C using $\mathcal{S}$–ϵ-coverings, then

$$\mu^*(A) := \sup_{\epsilon>0} \mu_\epsilon^*(A) \quad \forall A \subset X. \tag{18.5}$$

Note that $\mathscr{C}_\epsilon(A,\mathcal{S}) \subset \mathscr{C}_\delta(A,\mathcal{S})$ whenever $\epsilon < \delta$, i.e. the supremum appearing in (18.5) is even an increasing limit.

Theorem 18.5 *Let (X,d) be a metric space and $\mathcal{S}$ a family of sets which is 'rich' in the sense that for every $x \in X$ and $\epsilon > 0$ there is some $A \in \mathcal{S}$ with* $\operatorname{diam} A < \epsilon$ *and $x \in A$. The set function μ^* from Definition 18.4 is a metric outer measure, i.e. an outer measure with the additional property that*

$$\begin{aligned} A, B \subset X,\ d(A,B) := \inf_{x\in A, y\in B} d(x,y) > 0 \\ \implies \mu^*(A\cup B) = \mu^*(A) + \mu^*(B). \end{aligned} \tag{OM$_4$}$$

Proof We verify the properties (OM$_1$)–(OM$_4$).

- (OM$_1$) $\mu^*(\emptyset) = \sup_{\epsilon>0} \mu_\epsilon^*(\emptyset) = \sup_{\epsilon>0} 0 = 0$.
- (OM$_2$) If $A \subset B$, then $\mu_\epsilon^*(A) \leqslant \mu_\epsilon^*(B)$ for all $\epsilon > 0$, and the inequality is preserved as $\epsilon \to 0$.
- (OM$_3$) If $A = \bigcup_{i\in\mathbb{N}} A_i$, then $\mu_\epsilon^*(A) \leqslant \sum_{i=1}^\infty \mu_\epsilon^*(A_i) \leqslant \mu^*(A)$ and we can let $\epsilon \to 0$.
- (OM$_4$) Assume that $d(A,B) \geqslant \delta > 0$, pick $\epsilon < \delta/3$ and $(U_n)_n \in \mathscr{C}_\epsilon(A\cup B,\mathcal{S})$. Define $S_n := U_n$ if $U_n \cap A \neq \emptyset$ and $T_n := U_n$ if $U_n \cap B \neq \emptyset$. This defines $\mathcal{S}$–ϵ-covers $(S_n)_n \in \mathscr{C}_\epsilon(A,\mathcal{S})$ and $(T_n)_n \in \mathscr{C}_\epsilon(B,\mathcal{S})$ such that $S_m \cap T_n = \emptyset$ and $d(S_n, T_m) \geqslant \delta/3$ for all $m,n \in \mathbb{N}$. Consequently,

$$\sum_{n=1}^\infty C(U_n) \geqslant \sum_{i=1}^\infty C(S_i) + \sum_{k=1}^\infty C(T_k) \geqslant \mu_\epsilon^*(A) + \mu_\epsilon^*(B).$$

Since the cover $(U_n)_n$ is arbitrary, this yields $\mu_\epsilon^*(A\cup B) \geqslant \mu_\epsilon^*(A) + \mu_\epsilon^*(B)$ and, as $\epsilon \to 0$, we get

$$\mu^*(A\cup B) \geqslant \mu^*(A) + \mu^*(B) \quad \forall A, B \subset X,\ d(A,B) > 0.$$

The reverse inequality follows from (OM$_3$), hence we get equality. □

The property (OM$_4$) ensures that $\mathscr{A}^*$ is sufficiently rich. The proof is based on the following *increasing sets lemma*, which is also due to Carathéodory [18].

Lemma 18.6 (Carathéodory) *Let (X, d) be a metric space, μ^* a metric outer measure (in the sense of Theorem 18.5) and $A_i \uparrow A$ an increasing sequence of arbitrary sets such that $d(A_i, A \setminus A_{i+1}) > 0$ for all $i \in \mathbb{N}$. Then*

$$\mu^*(A) = \lim_{i\to\infty} \mu^*(A_i).$$

Proof Assume that the sets $(A_i)_{i\in\mathbb{N}}$ are as described in the statement of the lemma.

Step 1. By monotonicity, $\mu^*(A_i) \leqslant \mu^*(A)$, so $\lim_{i\to\infty} \mu^*(A_i) \leqslant \mu^*(A)$ is clear.

Step 2. For the converse inequality, we may assume that $\lim_{i\to\infty} \mu^*(A_i) < \infty$, otherwise there is nothing to be shown. Define

$$B_1 := A_1, \quad B_{i+1} := A_{i+1} \setminus A_i, \ i \in \mathbb{N}.$$

Note that $B_i \subset A_i$ and $B_{i+2} = A_{i+2} \setminus A_{i+1} \subset A \setminus A_{i+1}$, and the thus enlarged sets have less distance between them, i.e. $d(B_i, B_{i+2}) \geqslant d(A_i, A \setminus A_{i+1}) > 0$, see Fig. 18.1.

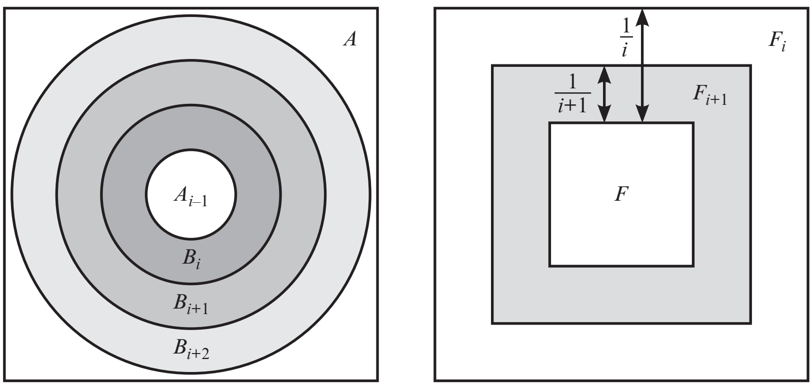


Fig. 18.1. *Left:* The sets B_i and B_{i+2} are separated by B_{i+1} and have strictly positive distance. *Right:* The sets $F_i \downarrow F$ are such that F_i^c and F_{i+1} have strictly positive distance.

Since μ^* is a metric outer measure and $d(B_i, B_{i+2}) > 0$, we see that

$$\mu^*(A_{2i-1}) \geqslant \mu^*\left(\dot{\bigcup_{i=1}^{n}} B_{2i-1}\right) = \sum_{i=1}^{n} \mu^*(B_{2i-1}), \tag{18.6}$$

$$\mu^*(A_{2i}) \geqslant \mu^*\left(\dot{\bigcup_{i=1}^{n}} B_{2i}\right) = \sum_{i=1}^{n} \mu^*(B_{2i}). \tag{18.7}$$

Observe that $\lim_{i\to\infty} \mu^*(A_i)$ is an upper bound for (18.6) and (18.7). This shows that the series $\sum_{i=1}^{\infty} \mu^*(B_i) = \sum_{i=1}^{\infty} \mu^*(B_{2i}) + \sum_{i=1}^{\infty} \mu^*(B_{2i-1})$ converges.

Step 3. By construction and σ-subadditivity,

$$\mu^*(A) = \mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu^*\left(A_n \cup \bigcup_{i=n+1}^{\infty}\!\!\!\!\cdot\;\; B_i\right) \overset{(\text{OM}_3)}{\leqslant} \mu^*(A_n) + \sum_{i=n+1}^{\infty} \mu^*(B_i),$$

but the sum on the right tends to zero as $n \to \infty$ since it is the tail of a convergent series, see Step 2. This proves that $\mu^*(A) \leqslant \lim_{n\to\infty} \mu^*(A_n)$. □

Theorem 18.7 *Let (X, d) be a metric space and μ^* a metric outer measure* (*in the sense of Theorem 18.5*). *Then $\mathscr{B}(X) \subset \mathscr{A}^*$.*

Proof Denote by $\mathscr{C}$ the closed sets and let $F \in \mathscr{C}$. We show that

$$\mu^*(Q) \geqslant \mu^*(Q \cap F) + \mu^*(Q \setminus F) \quad \forall Q \subset X. \tag{18.8}$$

The other inequality '$\leqslant$' is always satisfied since μ^* is subadditive. From (18.8) we see that $\mathscr{C} \subset \mathscr{A}^*$ and, as $\mathscr{A}^*$ is a σ-algebra, we conclude that $\mathscr{B}(X) = \sigma(\mathscr{C}) \subset \mathscr{A}^*$, see Remark 3.5(iii).

Pick any closed set F and any $Q \subset X$, and define $F_i := \{x \in X : d(x, F) \leqslant 1/i\}$. Since F is a closed set, $F = \bigcap_{i=1}^{\infty} F_i$. [✍]

On the other hand, $d(Q \setminus F_i, Q \cap F) \geqslant d(F_i^c, F) \geqslant 1/i$, see Fig. 18.1, and so

$$\mu^*(Q \cap F) + \mu^*(Q \setminus F_i) \overset{(\text{OM}_4)}{=} \mu^*\big(\underbrace{(Q \cap F) \cup (Q \setminus F_i)}_{\subset Q}\big) \leqslant \mu^*(Q). \tag{18.9}$$

We want to apply Lemma 18.6 to the sets $A_i := Q \setminus F_i \uparrow Q \setminus F =: A$. This is indeed possible, since – see Fig. 18.1 – we have

$$d(A_i, A \setminus A_{i+1}) = d(\underbrace{Q \setminus F_i}_{\subset F_i^c}, \underbrace{(Q \setminus F) \setminus (Q \setminus F_{i+1})}_{= Q \cap F^c \cap F_{i+1} \subset F_{i+1}}) \geqslant d(F_i^c, F_{i+1}) \geqslant \frac{1}{i} - \frac{1}{i+1} > 0$$

using again that enlarged sets have smaller distance than the original sets. Therefore, we can let $i \to \infty$ in (18.9) and find

$$\mu^*(Q \cap F) + \mu^*(Q \setminus F) = \mu^*(Q \cap F) + \lim_{i\to\infty} \mu^*(Q \setminus F_i) \leqslant \mu^*(Q). \qquad \square$$

Hausdorff Measures

Throughout this section (X, d) is a metric space[3] with its open sets $\mathscr{O}$, closed sets $\mathscr{C}$ and Borel sets $\mathscr{B}(X) = \sigma(\mathscr{O})$. Moreover, $\phi : [0, \infty) \to [0, \infty)$ is an increasing,

[3] See Appendix B, but you may also safely think of X as $\mathbb{R}^n$ and $d(x, y) = |x - y|$.

right-continuous function; later on, we will frequently use $\phi(x) = x^s$, for $x \geqslant 0$ and $s \geqslant 0$; as usual, $0^0 := 1$. With the help of ϕ we define a set function

$$C(A) = C^\phi(A) = \begin{cases} \phi(\operatorname{diam} A) & \text{if } A \neq \emptyset, \\ 0 & \text{if } A = \emptyset. \end{cases} \tag{18.10}$$

Using Theorem 18.5 with C and $\mathcal{S} = \mathscr{P} = \mathscr{P}(X)$ it is easy to see that

$$\begin{aligned} &\overline{\mathcal{H}}^\phi_\epsilon(A) := \inf\left\{\textstyle\sum_{i=1}^\infty C(S_i) : (S_i)_{i\in\mathbb{N}} \in \mathscr{C}_\epsilon(A;\mathscr{P})\right\}, \quad \epsilon > 0, \\ &\overline{\mathcal{H}}^\phi(A) := \lim_{\epsilon\to 0} \overline{\mathcal{H}}^\phi_\epsilon(A) \end{aligned} \tag{18.11}$$

defines a metric outer measure $\overline{\mathcal{H}}^\phi$ on $\mathscr{P}(X)$ whose restriction $\mathcal{H}^\phi$ to $\mathscr{B}(X)$ (see Theorem 18.7) is a measure.

Definition 18.8 The outer measure $\overline{\mathcal{H}}^\phi$ and the measure $\mathcal{H}^\phi$ are called *Hausdorff measure*[4] with measure function ϕ. If $\phi(x) = x^s$ for some $s \geqslant 0$, we write $\overline{\mathcal{H}}^s$ and $\mathcal{H}^s$, and speak of *s-dimensional Hausdorff measure*.

If we use only open or closed covers in (18.11), we arrive at the same notion of Hausdorff measure. Write, for a moment,

$$\overline{\mathcal{H}}^\phi_{\epsilon,\mathcal{S}}(A) := \inf\left\{\sum_{i=1}^\infty C(S_i) : (S_i)_{i\in\mathbb{N}} \in \mathscr{C}_\epsilon(A;\mathcal{S})\right\}.$$

Lemma 18.9 *Let (X,d) be a metric space, $\mathscr{O}$ the open sets, $\mathscr{C}$ the closed sets and $\mathscr{P} = \mathscr{P}(X)$ the power set. If $\overline{\mathcal{H}}^\phi_{\epsilon,\mathcal{S}}$ is the Hausdorff measure obtained by $\mathcal{S}$-ϵ-covers, then the following inequalities hold for all $\epsilon > \delta$:*

$$\overline{\mathcal{H}}^\phi_{\epsilon,\mathscr{O}}(A) \overset{\text{(i)}}{\leqslant} \overline{\mathcal{H}}^\phi_{\delta,\mathscr{C}}(A) \overset{\text{(ii)}}{\leqslant} \overline{\mathcal{H}}^\phi_{\delta,\mathscr{P}}(A) \overset{\text{(iii)}}{\leqslant} \overline{\mathcal{H}}^\phi_{\delta,\mathscr{O}}(A) \quad \forall A \subset X.$$

In particular, $\overline{\mathcal{H}}^\phi$ and $\mathcal{H}^\phi$ can be defined using open or closed covers only.

Proof Recall that $C(A) = \phi(\operatorname{diam} A)$ for a right-continuous increasing function ϕ.

(i) Fix $\epsilon > \delta$. We may assume that $\overline{\mathcal{H}}^\phi_{\delta,\mathscr{C}}(A) < \infty$, otherwise the inequality becomes trivial. By the definition of $\overline{\mathcal{H}}^\phi_{\delta,\mathscr{C}}(A)$, for every $\eta > 0$ there is a δ-cover of closed sets $(F_i)_{i\in\mathbb{N}}$ such that

$$\sum_{i=1}^\infty \phi(\operatorname{diam} F_i) \leqslant \overline{\mathcal{H}}^\phi_{\delta,\mathscr{C}}(A) + \eta.$$

[4] Although $\overline{\mathcal{H}}^\phi$ is an outer measure, it is customarily called 'measure'.

Consider the open [✍] sets $U_i := \{x \in X : d(x, F_i) < \eta_i\}$, where η_i is chosen in such a way that

- $\delta + 2\eta_i < \epsilon$,
- $\phi(\operatorname{diam} F_i + 2\eta_i) < \phi(\operatorname{diam} F_i) + \eta 2^{-i}$ (this is possible because of the right-continuity of ϕ).

Thus, $(U_i)_{i\in\mathbb{N}}$ is an ϵ-cover of open sets such that $F_i \subset U_i$; moreover,

$$\sum_{i=1}^{\infty} \phi(\operatorname{diam} U_i) \leqslant \sum_{i=1}^{\infty} \left(\phi(\operatorname{diam} F_i) + \eta 2^{-i}\right) \leqslant \overline{\mathcal{H}}^{\phi}_{\delta,\mathscr{C}}(A) + 2\eta,$$

and taking infima we get

$$\overline{\mathcal{H}}^{\phi}_{\epsilon,\mathscr{O}}(A) \leqslant \overline{\mathcal{H}}^{\phi}_{\delta,\mathscr{C}}(A) + 2\eta \xrightarrow[\eta\to 0]{} \overline{\mathcal{H}}^{\phi}_{\delta,\mathscr{C}}(A)$$

and the claim follows as $0 < \delta < \epsilon \to 0$.

(ii) For any δ-cover $(S_i)_{i\in\mathbb{N}} \subset \mathscr{P}$ the closures $(\overline{S}_i)_{i\in\mathbb{N}}$ form a closed δ-cover since the closure does not increase the diameters [✍]. Thus,

$$\overline{\mathcal{H}}^{\phi}_{\delta,\mathscr{C}}(A) \leqslant \sum_{i=1}^{\infty} \phi(\operatorname{diam} \overline{S}_i) = \sum_{i=1}^{\infty} \phi(\operatorname{diam} S_i).$$

Taking infima over all δ-covers shows the claimed estimate.

(iii) follows from the monotonicity of the infimum and the observation that every δ-cover with open sets is, in particular, a δ-cover with arbitrary sets.

Finally, the claim for Hausdorff measures follows by letting $\delta \to 0$. □

With similar arguments we can also show that Hausdorff measure can be defined by coverings of *convex sets* – this is due to the fact that the convex hull of any set A has again diameter $\operatorname{diam} A$.

Caution If we use $\mathcal{S} = \{B_r(x) : x \in X, r > 0\}$ in the definition of Hausdorff measure, we will, in general, not get $\overline{\mathcal{H}}^{\phi}$, see Besicovitch [5, p. 458 *et. seq.*]. This is basically due to the fact that a set $A \subset X$ with $\operatorname{diam} A = 2r$ cannot, in general, be covered by a ball with radius r – think of an equilateral triangle in $\mathbb{R}^2$.

Corollary 18.10 *Let (X, d) be a metric space. Hausdorff measure $\overline{\mathcal{H}}^{\phi}$ is outer regular in the sense that for each $A \subset X$ there is a G_δ-set G (i.e. a countable intersection of open sets) such that $\overline{\mathcal{H}}^{\phi}(A) = \mathcal{H}^{\phi}(G)$.*

Proof Throughout the proof we use the fact that $\overline{\mathcal{H}}^{\phi}|\mathscr{B}(X)=\mathcal{H}^{\phi}$. We can assume that $\overline{\mathcal{H}}^{\phi}(A)<\infty$, otherwise $G=X$ would do the job. Fix $k\in\mathbb{N}$ and $\eta>0$. Because of Lemma 18.9, there is a $1/k$-cover of open sets $(U_i^k)_{i\in\mathbb{N}}$ such that

$$\overline{\mathcal{H}}^{\phi}_{1/k}(A)\leqslant\sum_{i=1}^{\infty}C(U_i^k)\leqslant\overline{\mathcal{H}}^{\phi}_{1/k}(A)+\eta.$$

Since the sets $U^k:=\bigcup_{i=1}^{\infty}U_i^k$ are open, $G:=\bigcap_{k\in\mathbb{N}}U^k$ is a G_δ-set. Monotonicity and σ-subadditivity yield

$$\mathcal{H}^{\phi}_{1/k}(G)\leqslant\mathcal{H}^{\phi}_{1/k}(U^k)\leqslant\sum_{i=1}^{\infty}\mathcal{H}^{\phi}_{1/k}(U_i^k)\leqslant\sum_{i=1}^{\infty}C(U_i^k)\leqslant\overline{\mathcal{H}}^{\phi}_{1/k}(A)+\eta.$$

Letting $k\to\infty$ and then $\eta\to 0$ shows that $\mathcal{H}^{\phi}(G)\leqslant\overline{\mathcal{H}}^{\phi}(A)$. Since $G\supset A$, the other inequality is trivial. □

It is possible to show, using arguments similar to those of Appendix H, that any Borel set with $\mathcal{H}^{\phi}(A)<\infty$ contains an F_σ-set F (i.e. a countable union of closed sets) with $\overline{\mathcal{H}}^{\phi}(A)=\mathcal{H}^{\phi}(F)$, see e.g. [40, Chapter 1, Theorem 22] or Problem 18.3.

If $X=\mathbb{R}^n$, we can use 'nets' to define Hausdorff measure. Consider the lattice $2^{-r}\mathbb{Z}^n$, $r\in\mathbb{N}$, and denote in all lattices by $\mathscr{Q}$ the family of right-open cells $Q(r,\vec{z}):=2^{-r}([0,1)^n+\vec{z})$, $\vec{z}\in\mathbb{Z}^n$, $r\in\mathbb{N}$.

Lemma 18.11 *Let $X=\mathbb{R}^n$ and $\mathscr{P}=\mathscr{P}(\mathbb{R}^n)$. Then*

$$\overline{\mathcal{H}}^{\phi}_{\epsilon,\mathscr{P}}(A)\leqslant\overline{\mathcal{H}}^{\phi}_{\epsilon,\mathscr{Q}}(A)\leqslant 3^n2^{n(n+1)}\overline{\mathcal{H}}^{\phi}_{\epsilon,\mathscr{P}}(A)\quad\forall A\subset X,\ \epsilon>0.$$

In particular, we can define a measure which is equivalent to $\overline{\mathcal{H}}^{\phi}$ using $\mathscr{Q}$.

Proof The first estimate is obvious since every $\mathscr{Q}$–ϵ-cover is also a $\mathscr{P}$–ϵ-cover. For the second inequality we may safely assume that $\overline{\mathcal{H}}^{\phi}_{\epsilon,\mathscr{P}}(A)<\infty$, otherwise the estimate would be trivial. For every $\eta>0$ we can find an ϵ-cover $(S_i)_{i\in\mathbb{N}}\subset\mathscr{P}$ such that

$$\sum_{i=1}^{\infty}\phi(\operatorname{diam}S_i)\leqslant\overline{\mathcal{H}}^{\phi}_{\epsilon,\mathscr{P}}(A)+\eta.$$

Note that there is some $r(i) \in \mathbb{N}$ with $2^{-r(i)-1} \leqslant \operatorname{diam} S_i < 2^{-r(i)}$, and this means that

- S_i hits some cube $Q(r(i), \vec{z})$;
- S_i is contained in the union of 3^n cubes $Q(r(i), \vec{z}')$, where $|z_k - z_k'| \leqslant 1$ for $k = 1, \ldots, n$;
- each $Q(r(i), \vec{z}')$ can be split into $(2^{n+1})^n$ cubes of side-length $2^{-r(i)-(n+1)}$,

i.e. S_i is covered by $3^n(2^{n+1})^n$ cubes Q_{ik} of side-length $2^{-r(i)-(n+1)}$ and diameter

$$\operatorname{diam} Q_{ik} = 2^{-r(i)-(n+1)}\sqrt{n} < 2^{-r(i)-1} \leqslant \operatorname{diam} S_i.$$

If we take this $\mathscr{Q}$–ϵ-cover $(Q_{ik})_{ik}$ of A, we find due to the monotonicity of ϕ

$$\begin{aligned}\overline{\mathcal{H}}^{\phi}_{\epsilon,\mathscr{Q}}(A) \leqslant \sum_{i=1}^{\infty} \sum_{k=1}^{(3\cdot 2^{n+1})^n} \phi(\operatorname{diam} Q_{ik}) &\leqslant 3^n(2^{n+1})^n \sum_{i=1}^{\infty} \phi(\operatorname{diam} S_i)\\ &\leqslant 3^n(2^{n+1})^n \big(\overline{\mathcal{H}}^{\phi}_{\epsilon,\mathscr{P}}(A) + \eta\big).\end{aligned}$$

The asserted estimate follows if we let $\eta \to 0$; the claim about Hausdorff measures follows with $\epsilon \to 0$. □

In order to keep the presentation simple, from now on we will use $\phi(x) = x^s$ only with $s \geqslant 0$. Here are a few elementary properties of s-dimensional Hausdorff measure.

Theorem 18.12 *Let $\mathcal{H}^s$, $s \geqslant 0$ be Hausdorff measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$.*

(i) *$\mathcal{H}^0$ is the counting measure on $\mathscr{P}(\mathbb{R}^n)$.*
(ii) *$\mathcal{H}^1$ is one-dimensional Lebesgue measure on $\mathscr{B}(\mathbb{R})$.*
(iii) *$\mathcal{H}^s(\mathbb{R}^n) = 0$ if $s > n$.*
(iv) *$\overline{\mathcal{H}}^s(A) < \infty \implies \overline{\mathcal{H}}^t(A) = 0$ for all $A \subset \mathbb{R}^n$ and $t > s \geqslant 0$.*
(v) *$\overline{\mathcal{H}}^s(\lambda A) = \lambda^s \overline{\mathcal{H}}^s(A)$ for all $A \subset \mathbb{R}^n$ and $\lambda \geqslant 0$.*
(vi) *$\overline{\mathcal{H}}^s(f(A)) \leqslant L_f \cdot \overline{\mathcal{H}}^s(A)$ for all $A \subset \mathbb{R}^n$, if $f: \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous, where $L_f := \sup_{x,y \in \mathbb{R}^n} |f(x) - f(y)|/|x - y|$ is the (optimal) Lipschitz constant.*
(vii) *$\overline{\mathcal{H}}^s(\jmath(A)) = \overline{\mathcal{H}}^s(A)$ for all $A \subset \mathbb{R}^n$ if $\jmath: \mathbb{R}^n \to \mathbb{R}^n$ is an isometry.*
(viii) *$\overline{\mathcal{H}}^s(\pi(A)) \leqslant \overline{\mathcal{H}}^s(A)$ for all $A \subset \mathbb{R}^n$ if $\pi: \mathbb{R}^n \to \mathbb{R}^k$ is a projection.*

Proof We have $C(A) = (\operatorname{diam} A)^s$ for all $A \subset \mathbb{R}^n$, $A \neq \emptyset$ and $C(\emptyset) = 0$.

(i) Since $C\{x\} = 1$, it is clear that for every finite set $\mathcal{H}^0\{x_1, \ldots, x_m\} = m$. This follows from the 'optimal' covering of the finite set by the balls $B_\epsilon(x_i)$, where $\epsilon < \frac{1}{5} \min_{1 \leqslant i,k \leqslant m} |x_i - x_k|$. Therefore, $\mathcal{H}^0$ is the counting measure.

It is a simple exercise to show that $\mathscr{P}(\mathbb{R}^n)$ are indeed the $\overline{\mathcal{H}}^0$-measurable sets. [✍]

(ii) By Theorem 18.7 $\mathscr{B}(\mathbb{R}) \subset \mathscr{A}^*$. From its definition, it is clear that $\mathcal{H}^s$ is invariant under translations. In view of Theorem 5.8 we have to show $\mathcal{H}^1[0,1) = 1$. Let $0 = t_0 < t_1 < \cdots < t_N = 1$ and $\max(t_i - t_{i-1}) < \epsilon$. Since $\{[t_{i-1}, t_i), 1 \leqslant i \leqslant N\}$ is an ϵ-cover of $[0,1)$, we see that

$$\mathcal{H}^1_\epsilon[0,1) \leqslant \sum_{i=1}^N C[t_{i-1}, t_i) = \sum_{i=1}^N (t_i - t_{i-1}) = 1,$$

which proves that $\mathcal{H}^1[0,1) \leqslant 1$. Let $(U_i)_{i\in\mathbb{N}}$ be any open ϵ-cover of $[0,1)$. We can decompose every open set in $\mathbb{R}$ into countably many disjoint open intervals,[5] so we may assume that the U_i are themselves intervals (a_i, b_i). In particular,

$$\sum_{i=1}^\infty (b_i - a_i) = \sum_{i=1}^\infty \lambda^1(a_i, b_i) \geqslant \lambda^1\left(\bigcup_{i=1}^\infty (a_i, b_i)\right) \geqslant \lambda^1[0,1) = 1.$$

Taking infima over all open covers reveals that $\mathcal{H}^1[0,1) \geqslant \mathcal{H}^1_\epsilon[0,1) \geqslant 1$.

(iii) Let $Q_r = [-r, r)^n$ and decompose Q_r into m^n disjoint cubes of side-length $2r/m$ and diameter $2r\sqrt{n}/m$. By definition

$$\mathcal{H}^s_{2r\sqrt{n}/m}(Q_r) \leqslant \sum_{i=1}^{m^n} \left(\frac{2r\sqrt{n}}{m}\right)^s = (2r)^s n^{s/2} m^{n-s} \xrightarrow[m\to\infty]{} 0.$$

This shows that $\mathcal{H}^s(Q_r) = 0$ and $\mathcal{H}^s(\mathbb{R}^n) = \lim_{r\to\infty} \mathcal{H}^s(Q_r) = 0$ follows from the continuity of measures.

(iv) We know for any ϵ-cover $(S_i)_{i\in\mathbb{N}}$ of A and $t > s \geqslant 0$ that

$$\sum_{i=1}^\infty (\operatorname{diam} S_i)^t \leqslant \epsilon^{t-s} \sum_{i=1}^\infty (\operatorname{diam} S_i)^s,$$

and so $\overline{\mathcal{H}}^t_\epsilon(A) \leqslant \epsilon^{t-s} \overline{\mathcal{H}}^s_\epsilon(A) \leqslant \epsilon^{t-s} \overline{\mathcal{H}}^s(A)$. The claim follows as $\epsilon \to 0$.

(v) is obvious as $\operatorname{diam}(\lambda A) = \lambda \operatorname{diam} A$.

(vi) Let $(S_i)_{i\in\mathbb{N}}$ be an ϵ-cover of $A \subset \mathbb{R}^n$. Because of the Lipschitz continuity we know that

$$\operatorname{diam} f(S_i) \leqslant L_f \operatorname{diam} S_i \leqslant L_f \epsilon.$$

[5] We assume that U is open and take $x \in U$. By definition, there exists some interval $x \in (a,b) \subset U$. Therefore $a(x) := \inf\{a \in U : (a, x] \subset U\}$ and $b(x) := \sup\{b \in U : [x, b) \subset U\}$ are well-defined, $a(x) < b(x)$, $a(x), b(x) \notin U$, due to the openness of U. Moreover, if $y \in (a(x), b(x))$, then $a(y) = a(x)$ and $b(y) = b(x)$. Since each interval $(a(x), b(x))$ contains a rational, we see that $U = \bigcup_{x\in U}(a(x), b(x))$ is actually an at most countable union of mutually disjoint sets (we disregard all multiple countings of intervals ...).

Moreover, $f(A) \subset f\big(\bigcup_i S_i\big) \subset \bigcup_i f(S_i)$, so $(f(S_i))_i$ is an $L_f\epsilon$-cover of $f(A)$. Therefore

$$\overline{\mathcal{H}}^s_{L_f\epsilon}(f(A)) \leqslant \sum_{i=1}^{\infty} (\operatorname{diam} f(S_i))^s \leqslant L_f \sum_{i=1}^{\infty} (\operatorname{diam} S_i)^s.$$

Taking infima reveals that $\overline{\mathcal{H}}^s_{L_f\epsilon}(f(A)) \leqslant \overline{\mathcal{H}}^s_{\epsilon}(A)$ and the limit $\epsilon \to 0$ finishes the proof.

Finally, (vii) and (viii) are obvious from (vi) since isometries $\jmath$, their inverse $\jmath^{-1}$ and projections π are Lipschitz with Lipschitz constant 1. □

Calculating the Hausdorff measure even of simple sets can be tricky, because it routinely requires clever covering arguments or the isodiametric inequality.

Lemma 18.13 (isodiametric inequality) *The inequality*

$$\lambda^n(A) \leqslant \lambda^n(B_1(0))2^{-n}(\operatorname{diam} A)^n = \lambda^n\big(B_{\frac{1}{2}\operatorname{diam} A}(0)\big)$$

holds for every Borel set $A \subset \mathbb{R}^n$.

In other words, the geometric volume of a set A of diameter $\operatorname{diam} A$ is always smaller than the geometric volume of a ball of diameter $\operatorname{diam} A$. This is non-trivial since A need not be contained in any ball of diameter $\operatorname{diam} A$. For a proof we refer the reader to Evans and Gariepy [16, Section 2.2, p. 69]. If we combine this with the maximality assertion for the outer measure constructed in Theorem 18.2,[6] it is immediately evident that the isodiametric inequality implies that

$$\lambda^n(A) \leqslant \lambda^n(B_1(0))2^{-n}\mathcal{H}^n(A) \quad \forall A \in \mathscr{B}(\mathbb{R}^n). \tag{18.12}$$

Example 18.14 Consider in $\mathbb{R}^n$ the n-dimensional Hausdorff measure $\mathcal{H}^n$.

(i) $\mathcal{H}^n(B_1(0)) = 2^n$ and $\mathcal{H}^n(B_r(x)) = (2r)^n$ for any $x \in \mathbb{R}^n$ and $r > 0$.
(ii) $\mathcal{H}^n([0,1)^n) = 2^n/\omega_n$, where $\omega_n = \lambda^n(B_1(0)) = \pi^{n/2}/\Gamma\big(\frac{1}{2}+1\big)$ is the geometric volume of the n-dimensional unit ball.

Proof In order to show (i) we observe that

$$\overline{\mathcal{H}}^n_{2\epsilon}(B_1(0)) = \overline{\mathcal{H}}^n_{2\epsilon}\big(\tfrac{1}{\epsilon}B_\epsilon(0)\big) = \epsilon^{-n}\overline{\mathcal{H}}^n_{2\epsilon}(B_\epsilon(0)) \leqslant \epsilon^{-n}(\operatorname{diam} B_\epsilon(0))^n = 2^n,$$

[6] Use $\nu^* = \kappa\lambda^n$ with $\kappa = 2^n/\lambda^n(B_1(0))$, $\mu^* = \overline{\mathcal{H}}^n$ and $C(A) = (\operatorname{diam} A)^n$.

since $\{B_\epsilon(0)\}$ is an admissible open 2ϵ-cover of $B_\epsilon(0)$. Letting $\epsilon \to 0$ yields $\mathcal{H}^n(B_1(0)) \leqslant 2^n$. For the other estimate we use the isodiametric inequality (18.12) with $A = B_1(0)$ to see

$$\lambda^n(B_1(0)) \leqslant \omega_n 2^{-n} \mathcal{H}^n(B_1(0)) \implies 2^n \leqslant \mathcal{H}^n(B_1(0)).$$

The assertion for $B_r(0)$ follows from $B_r(0) = rB_1(0)$, scaling (Theorem 18.12(v)) and the fact that $\mathcal{H}^n$ is invariant under translations (Theorem 18.12(vii)).

The second example is even more difficult. Set $Q = [0,1)^n$ and use a clever covering of Q. Because of the translation invariance of $\mathcal{H}^n$, we know from Theorem 5.8 that $\mathcal{H}^n = \kappa\lambda^n$, where $\kappa = \mathcal{H}^n(Q)$. By part (i), $0 < \kappa < \infty$, i.e. $\mathcal{H}^n$ and λ^n have the same Borel null sets.

By Vitali's covering theorem (Theorem F.1 in Appendix F) there exist disjoint closed balls $(K_i)_{i\in\mathbb{N}}$ with $\operatorname{diam} K_i < \epsilon$, $\bigcup_{i=1}^\infty K_i \subset Q$ and

$$\sum_{i=1}^\infty \lambda^n(K_i) = \lambda^n(Q).$$

Using the σ-additivity and translation invariance of Hausdorff and Lebesgue measure

$$\begin{aligned}
\mathcal{H}^n(Q) \overset{\text{same}}{\underset{\text{null sets}}{=}} \mathcal{H}^n\left(\bigcup_{i=1}^\infty K_i\right) &= \sum_{i=1}^\infty \mathcal{H}^n(K_i) \overset{\text{(i)}}{=} \sum_{i=1}^\infty (\operatorname{diam} K_i)^n \\
= \frac{2^n}{\omega_n} \sum_{i=1}^\infty \omega_n \left(\frac{\operatorname{diam} K_i}{2}\right)^n &= \frac{2^n}{\omega_n} \sum_{i=1}^\infty \lambda^n(K_i) = \frac{2^n}{\omega_n} \lambda^n(Q). \quad \square
\end{aligned}$$

The best way to express the findings of Example 18.14 is to say that Lebesgue measure assigns unit volume to a cube of side 1, while Hausdorff measure assigns unit volume to a ball of diameter 1. A bit more is true, as follows.

Theorem 18.15 $\lambda^n = \omega_n 2^{-n} \mathcal{H}^n$ *on* $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ *with* $\omega_n = \lambda^n(B_1(0))$.

Proof Since Hausdorff measure is invariant under translations, we know from Theorem 5.8 that $\mathcal{H}^n = \kappa\lambda^n$ with $\kappa = \mathcal{H}^n([0,1)^n)$. The claim follows now from the formula in Example 18.14(ii). $\square$

Hausdorff Dimension

We saw in Theorem 18.12(iv) that

$$\overline{\mathcal{H}}^s(A) < \infty \implies \overline{\mathcal{H}}^t(A) = 0 \quad \forall t > s \geqslant 0,\ A \subset \mathbb{R}^n.$$

In a similar way we get $\overline{\mathcal{H}}^r(A) = \infty$ if $r \in (0, s)$. This means that Hausdorff measure takes a finite and non-zero value $0 < \overline{\mathcal{H}}^s(A) < \infty$ for at most one index s.

The value at which Hausdorff measure changes from $+\infty$ to 0 is commonly called the Hausdorff dimension.

Definition 18.16 The *Hausdorff dimension* of a set $A \subset \mathbb{R}^n$ is

$$\dim_{\mathcal{H}}(A) = \inf\left\{s \in (0,\infty) : \overline{\mathcal{H}}^s(A) = 0\right\}. \tag{18.13}$$

The considerations preceding Definition 18.16 show that

$$\begin{aligned}\dim_{\mathcal{H}}(A) &= \sup\left\{s \in (0,\infty) : \overline{\mathcal{H}}^s(A) = \infty\right\}\\ &= \sup\left\{s \in (0,\infty) : \overline{\mathcal{H}}^s(A) > 0\right\}\\ &= \inf\left\{s \in (0,\infty) : \overline{\mathcal{H}}^s(A) < \infty\right\}.\end{aligned}$$

Here are a few simple rules for the calculation of Hausdorff dimensions.

Lemma 18.17 *Let $A, B, A_1, A_2, \ldots \subset \mathbb{R}^n$ be arbitrary sets.*

(i) $A \subset B \implies \dim_{\mathcal{H}} A \leqslant \dim_{\mathcal{H}} B$.
(ii) $\dim_{\mathcal{H}}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sup_{i\in\mathbb{N}} \dim_{\mathcal{H}} A_i$.
(iii) $\overline{\mathcal{H}}^s(A) \in (0,\infty) \implies \dim_{\mathcal{H}} A = s$.

Proof (i) Let $s > \dim_{\mathcal{H}} B$. By definition, we have that $\overline{\mathcal{H}}^s(B) = 0$ and so $\overline{\mathcal{H}}^s(A) \leqslant \overline{\mathcal{H}}^s(B) = 0$. Therefore, $s \geqslant \dim_{\mathcal{H}} A$ and the claim follows on letting $s \downarrow \dim_{\mathcal{H}} B$.

(ii) Set $A := \bigcup_{i=1}^{\infty} A_i$. Part (i) shows that $\dim_{\mathcal{H}}(A) \geqslant \dim_{\mathcal{H}}(A_i)$ for all i, hence $\dim_{\mathcal{H}} A \geqslant \sup_{i\in\mathbb{N}} \dim_{\mathcal{H}} A_i$.

Assume that $s > \sup_{i\in\mathbb{N}} \dim_{\mathcal{H}} A_i$. This implies that $\overline{\mathcal{H}}^s(A_i) = 0$ for all i, and by the σ-subadditivity, $\overline{\mathcal{H}}^s(A) \leqslant \sum_{i=1}^{\infty} \overline{\mathcal{H}}^s(A_i) = 0$. From this we conclude that $s \geqslant \dim_{\mathcal{H}}(A)$ and the claim follows as $s \downarrow \sup_{i\in\mathbb{N}} \dim_{\mathcal{H}} A_i$.

(iii) is clear from the discussion preceding Definition 18.16. □

Calculating Hausdorff dimension can be quite tricky. We close this chapter with a few typical examples.

Example 18.18 (integer dimensions) $\dim_{\mathcal{H}} \mathbb{R} = 1$ and $\dim_{\mathcal{H}} \mathbb{R}^n = n$.

Indeed: $\mathcal{H}^1[0,1) \stackrel{18.12(\mathrm{ii})}{=} \lambda^1[0,1) = 1$, which implies that $\dim_{\mathcal{H}}[0,1) = 1$. In the same fashion we see that $\dim_{\mathcal{H}}[i,i+1) = 1$ for any $i \in \mathbb{Z}$. Since $\mathbb{R} = \bigcup_{i\in\mathbb{Z}}[i,i+1)$, we get $\dim_{\mathcal{H}} \mathbb{R} = \sup_{i\in\mathbb{Z}} \dim_{\mathcal{H}}[i,i+1) = 1$ from Lemma 18.17(ii).

Using Example 18.14, the multivariate case follows in the same way; just replace the dimension 1 by n and the intervals $[i,i+1)$ by the n-cells $\vec{z} + [0,1)^n$, where $\vec{z} \in \mathbb{Z}^n$.

Example 18.19 (self-similar structures) Let $C = \bigcap_{n=0}^{\infty} C_n$ be the Cantor ternary set which we constructed in the prologue, on page 4. Recall that the set C_n of the nth generation is made up of 2^n disjoint closed intervals J_i^n, $i = 1, \dots, 2^n$, of length 3^{-n}. Moreover, C is *self-similar* in the sense that the 2^n pieces of C_n are each miniature copies of C, scaled by the factor 3^{-n}.

The Hausdorff dimension is $\dim_{\mathcal{H}} C = \log 2/\log 3 \approx 0.6309\dots$.

Heuristically this is easily explained using self-similarity. Let $s = \log 2/\log 3$ and *assume* that $\mathcal{H}^s(C) \in (0, \infty)$ (this is not obvious). Then we get from $C = (C \cap [0, 1/3]) \cup (C \cap [2/3, 1])$ and the fact that the sets on the left and right are downscaled versions of C that

$$\mathcal{H}^s(C) = \mathcal{H}^s(C \cap [0, 1/3]) + \mathcal{H}^s(C \cap [2/3, 1]) = \frac{1}{3^s}\mathcal{H}^s(C) + \frac{1}{3^s}\mathcal{H}^s(C).$$

On dividing by $\mathcal{H}^s(C)$ and solving the resulting equation $1 = 2 \cdot 3^{-s}$ for s we obtain $s = \log 2/\log 3$.

Let us show that $\mathcal{H}^s(C) \in (0, \infty)$. First of all it is clear that the intervals J_i^n, $i = 1, \dots, 2^n$, are a 3^{-n}-cover of C. Thus,

$$\mathcal{H}^s_{3^{-n}}(C) \leqslant \sum_{i=1}^{2^n} (\operatorname{diam} J_i^n)^s = 2^n (3^{-n})^s = 1 \implies \mathcal{H}^s(C) \leqslant 1.$$

We will now show that $\mathcal{H}^s(C) \geqslant \frac{1}{2}$. For this we have to check that for any 3^{-n}-cover with open sets $(U_i)_{i \in \mathbb{N}}$ we have

$$\sum_{i=1}^{\infty} (\operatorname{diam} U_i)^s \geqslant \frac{1}{2}.$$

Without loss of generality, we can assume that the U_i are open intervals – otherwise we would take their convex hulls without changing the value of the sum. Since C is compact – it is an intersection of a decreasing sequence of closed and bounded sets C_n – it is already fully covered by $U_1, \dots, U_N$ for some $N \in \mathbb{N}$. For every $i = 1, \dots, N$ there is some $n(i)$ such that

$$3^{-n(i)-1} \leqslant \operatorname{diam} U_i < 3^{-n(i)}, \quad i = 1, \dots, N,$$

which means that U_i meets at most at most $2^{m-n(i)}$ intervals J_i^m, $i = 1, \dots, 2^m$, where $m \geqslant n(i)$. If we take $m > \max_{1 \leqslant i \leqslant N} n(i)$, we have $3^{-m-1} \leqslant \operatorname{diam} U_i$ for all $i = 1, \dots, N$, and since $U_1 \cup \dots \cup U_N \supset C$, the U_i intersect *all* of the J_i^m, $i = 1, \dots, 2^m$, *at least once*. Thus, counting the intervals intersected by the U_i gives

$$2^m \leqslant \sum_{i=1}^{N} 2^{m-n(i)} = \sum_{i=1}^{N} 2^m 3^{-sn(i)} = \sum_{i=1}^{N} 2^m 3^s (\operatorname{diam} U_i)^s$$

and from this we easily get $\sum_{i=1}^{N} (\operatorname{diam} U_i)^s \geqslant 3^{-s} = \frac{1}{2}$.

Example 18.20 (graphs) Let $f: \mathbb{R}^n \to \mathbb{R}^m$ be a function and $A \subset \mathbb{R}^n$. We denote by $\Gamma(f; A) = \{(x, f(x)) : x \in A\} \subset \mathbb{R}^n \times \mathbb{R}^m$ the *graph of f over the set A*. If $\lambda^n(A) > 0$, then

$$\dim_{\mathcal{H}} \Gamma(f; A) \geqslant n,$$

and equality holds, if f is Lipschitz continuous.

The lower bound is easy: because of Theorem 18.12(viii), we know that

$$\mathcal{H}^n(\Gamma(f; A)) \geqslant \mathcal{H}^n(\pi(\Gamma(f; A))) = \mathcal{H}^n(A) > 0$$

for the projection $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $(x, y) \mapsto x$. Thus, $\dim_{\mathcal{H}} \Gamma(f; A) \geqslant n$.

Let $L_f := \sup_{x,y} |f(x) - f(y)|/|x - y|$ be the (optimal) Lipschitz constant. If $Q \subset \mathbb{R}^n$ with side ϵ, then $f(Q)$ is contained in a cube $Q' \subset \mathbb{R}^m$ with side $L_f \epsilon$. So, $\Gamma(f; Q) \subset Q \times Q'$ and $\operatorname{diam}(Q \times Q') \leqslant \kappa\epsilon$ where $\kappa = \kappa(m, n, L^f)$.

Now take the cube $I := [0, 1)^n$ (or any translate of it), $\epsilon = 1/m$, and split it into m^n smaller cubes Q_i of side $1/m$. Then

$$\mathcal{H}^n_{\kappa/m}(\Gamma(f; A \cap I)) \leqslant \sum_{i=1}^{m^n} (\operatorname{diam} Q_i \times Q_i')^n \leqslant m^n \kappa^n \frac{1}{m^n} = \kappa^n.$$

On letting $m \to \infty$, we find $\mathcal{H}^n(\Gamma(f; A \cap I)) < \infty$, hence $\dim_{\mathcal{H}} \Gamma(f; A \cap I) \leqslant n$. Since $A = A \cap \bigcup_{\vec{z} \in \mathbb{Z}^n} (\vec{z} + I)$, the claim follows from Lemma 18.17(ii).

Problems

18.1. Show that we even get '=' in the estimate denoted by (ii) in Lemma 18.9.

18.2. Show that the outer regularity from Corollary 18.10 coincides with the usual notion, i.e.

$$\overline{\mathcal{H}}^\phi(A) = \inf \left\{ \mathcal{H}^\phi(U) : U \supset A,\ U \text{ open} \right\}, \tag{18.14}$$

provided that there exists some open set $U \supset A$ with finite Hausdorff measure. Use the example $\mathcal{H}^0$ (counting measure) to show that this condition is essential.

18.3. Let $X = \mathbb{R}^n$ (or a separable metric space). Let B be a Borel set or, more generally, an $\overline{\mathcal{H}}^\phi$-measurable set, such that $\mathcal{H}(B) < \infty$. Show that in the situation of Corollary 18.10 the measure $\mathcal{H}^\phi$ is inner regular in the following sense: B contains an F_σ-set F (this is a countable union of closed sets) with $\mathcal{H}^\phi(F) = \mathcal{H}^\phi(B)$.

[*Instructions*. Open sets in X are F_σ-sets. Thus, Corollary 18.10 gives a decreasing sequence $U_i \supset B$ of open sets and increasing sequences $F_{ik} \uparrow U_i$ of closed sets. Show that $\mathcal{H}^\phi(B \setminus F_{ik(i)}) \leqslant \epsilon/2^i$, define $F = \bigcap_i F_{ik(i)}$ and verify $\mathcal{H}^\phi(F \setminus B) = 0$. According to Corollary 18.10 $F \setminus B$ is contained in a G_δ-set G with $\mathcal{H}^\phi(G) = 0$, the F_σ-set $F \setminus G$ is in B and has the same Hausdorff measure.]

18.4. Finish the proof of Theorem 18.12(i) and show that every set $A \subset \mathbb{R}^n$ is indeed $\overline{\mathcal{H}}^0$-measurable.

18.5. For every $B \subset \mathbb{R}^n$ one has $\dim_{\mathcal{H}} B \leqslant n$. If B contains an open set or a set of strictly positive Lebesgue measure, then $\dim_{\mathcal{H}} B = n$.

18.6. Consider the fractals introduced in Problems 1.4 and 1.5: the Koch snowflake and the Sierpiński triangle. Give a heuristic argument that the Hausdorff dimension of the Sierpiński triangle is $\log 3/\log 2$. What is the dimension of the Koch snowflake?

18.7. Assume that $\phi, \psi : [0, \infty) \to [0, \infty)$ are admissible for the construction of Hausdorff measures and assume that $\lim_{x \to 0} \phi(x)/\psi(x) = 0$. Show that $\overline{\mathcal{H}}^{\psi}(A) < \infty$ implies $\overline{\mathcal{H}}^{\phi}(A) = 0$.

19

The Fourier Transform

Throughout this chapter we consider the measurable space $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and often write dx for the n-dimensional Lebesgue measure. The Fourier transform is an indispensable tool which is used in analysis, probability theory, mathematical physics and many other disciplines. Depending on the setting, various norming conventions are used. Here we adopt 'probabilistic' norming, i.e. the inverse Fourier transform coincides with the characteristic function used in probability theory.

Recall that $C_b(\mathbb{R}^n)$ and $C_c(\mathbb{R}^n)$ are the *bounded continuous* and *continuous and compactly supported* functions, respectively; $\langle x, \xi\rangle = \sum_{k=1}^n x_k \xi_k$ denotes the usual Euclidean scalar product in $\mathbb{R}^n$.

Definition 19.1 The *Fourier transform* of a finite measure μ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ is given by

$$\widehat{\mu}(\xi) := \mathscr{F}\mu(\xi) := (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-i\langle x,\xi\rangle} \mu(dx). \tag{19.1}$$

The *Fourier transform* of a function $f \in L^1(\lambda^n)$ is given by

$$\widehat{f}(\xi) = \mathscr{F}f(\xi) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-i\langle x,\xi\rangle} f(x)dx. \tag{19.2}$$

Note that for $\mu(dx) = f(x)dx$ we have $\widehat{\mu} = \widehat{f}$, i.e. the definition of the Fourier transform for measures and functions is consistent.

Caution The integral of a complex function $u : \mathbb{R}^n \to \mathbb{C}$ is defined by linearity, $\int u\,d\mu = \int \operatorname{Re} u\,d\mu + i \int \operatorname{Im} u\,d\mu$; this definition preserves the usual rules for the integral and we have, additionally,

$$\operatorname{Re}\int u\,d\mu = \int \operatorname{Re} u\,d\mu, \quad \operatorname{Im}\int u\,d\mu = \int \operatorname{Im} u\,d\mu \quad \text{and} \quad \overline{\int u\,d\mu} = \int \bar{u}\,d\mu.$$

Since the map $\mathbb{R}^2 \ni (\operatorname{Re} z, \operatorname{Im} z) \mapsto z \in \mathbb{C}$ is bi-continuous, and hence bi-measurable, there are no measurability problems. For details we refer the reader to Appendix D. In particular, (19.2) still holds for $f \in L^1_{\mathbb{C}}(dx)$: just use $f = \operatorname{Re} f + i \operatorname{Im} f$.

Example 19.2 (i) $\widehat{\mathbb{1}_{[a,b]}}(\xi) = \left(e^{-i\xi a} - e^{-i\xi b}\right) / 2\pi i \xi, \quad a < b.$

(ii) $\widehat{\delta_c}(\xi) = (2\pi)^{-n} e^{-i\langle c,\xi\rangle}, \quad c \in \mathbb{R}^n.$

(iii) For all $x, \xi \in \mathbb{R}^n$ and $t > 0$

$$g_t(x) := (2\pi t)^{-n/2} e^{-|x|^2/2t} \Longrightarrow \widehat{g_t}(\xi) = (2\pi)^{-n} e^{-t|\xi|^2/2}. \tag{19.3}$$

Indeed: see the proof of Lemma 17.14 and use $\widehat{g_t}(\xi) = (2\pi)^{-n} \check{g}_t(\xi)$ (in the notation of Lemma 17.14).

Properties 19.3 (of the Fourier transform) *Let μ be a finite measure on $\mathbb{R}^n$ and $f \in L^1(\lambda^n)$. The Fourier transforms $\widehat{\mu}$ and $\widehat{f}$ are continuous functions which are bounded by*

$$|\widehat{\mu}(\xi)| \leqslant \widehat{\mu}(0) = (2\pi)^{-n} \mu(\mathbb{R}^n) \quad \textit{resp.} \quad |\widehat{f}(\xi)| \leqslant \widehat{f}(0) = (2\pi)^{-n} \|f\|_1.$$

Moreover, the transformation rules in Table 19.1 hold.

Table 19.1. *Transformation rules*

	Function	Fourier transform	Remarks
(i)	$\widetilde{\mu}(dx) = \mu(-dx)$	$\overline{\widehat{\mu}(\xi)}$	Reflection at the origin
	$\widetilde{f}(x) = f(-x)$	$\overline{\widehat{f}(\xi)}$	
(ii)	$\mu \circ T^{-1}(dx)$	$e^{-i\langle b,\xi\rangle} \widehat{\mu}(\lambda \xi)$	$Ty = \lambda y + b$, $\lambda \in \mathbb{R}, b \in \mathbb{R}^n$
(iii)	$f(rx + c)$	$r^{-n} e^{i\langle c, r^{-1}\xi\rangle} \widehat{f}(r^{-1}\xi)$	$r \neq 0$, $c \in \mathbb{R}^n$
(iv)	$f(x) e^{-i\langle x,c\rangle}$	$\widehat{f}(c + \xi)$	$c \in \mathbb{R}^n$
(v)	$f(Rx)$	$\widehat{f}(R\xi)$	$R \in \mathbb{R}^{n\times n}$ orthogonal

Proof The continuity of $\widehat{\mu}$ and $\widehat{f}$ follows from the continuity lemma for parameter-dependent integrals (Theorem 12.4), where we use the following integrable majorants: $|e^{-i\langle x,\xi\rangle}| = 1 \in L^1(\mu)$ and $|f| \in L^1(\lambda^n)$. The bounds on $\widehat{\mu}$ and $\widehat{f}$ are directly obtained from the definition of the Fourier transform.

(i) follows from (ii), resp. (iii), if we take $b=0$, $\lambda=-1$, resp., $r=-1$, $c=0$, and observe that $\widehat{\mu}(-\xi)=\overline{\widehat{\mu}(\xi)}$ and $\widehat{f}(-\xi)=\overline{\widehat{f}(\xi)}$.

(ii) Note that

$$\begin{aligned}\widehat{\mu\circ T^{-1}}(\xi) &= (2\pi)^{-n}\int e^{-i\langle x,\xi\rangle}\mu\circ T^{-1}(dx) \overset{15.1}{=} (2\pi)^{-n}\int e^{-i\langle Tx,\xi\rangle}\mu(dx)\\ &= (2\pi)^{-n}\int e^{-i\langle \lambda x+b,\xi\rangle}\mu(dx) = e^{-i\langle b,\xi\rangle}\,\widehat{\mu}(\lambda\xi).\end{aligned}$$

(iii) follows from (compare with Example 15.3)

$$\begin{aligned}(2\pi)^{-n}\int f(rx+c)e^{-i\langle x,\xi\rangle}\,dx &\underset{dy=r^n\,dx}{\overset{y=rx+c}{=}} (2\pi)^{-n}\int f(y)e^{-i\langle r^{-1}(y-c),\xi\rangle}r^{-n}\,dy\\ &= r^{-n}e^{i\langle c,r^{-1}\xi\rangle}\widehat{f}(r^{-1}\xi).\end{aligned}$$

(iv) is obvious, and (v) goes like this: since R is orthogonal, we have $R^{-1}=R^\top$ and $\det R=\pm 1$, i.e.

$$\begin{aligned}\int f(Rx)e^{-i\langle x,\xi\rangle}\,dx &\overset{4.7}{=} |\det R|^{-1}\int f(y)e^{-i\langle R^{-1}y,\xi\rangle}\,dy\\ &= \int f(y)e^{-i\langle y,(R^{-1})^\top\xi\rangle}\,dy\\ &= (2\pi)^n\widehat{f}(R\xi).\end{aligned}$$

□

Example 19.4 (the Fourier transform of a rotationally symmetric function) Let $u\in L^1(\mathbb{R}^n,dx)$ and assume that $u(x)=f(|x|)$. We have

$$\int_{\mathbb{R}^n} f(|x|)\,e^{i\langle x,\xi\rangle}\,dx=\frac{(2\pi)^{n/2}}{|\xi|^{n/2-1}}\int_0^\infty f(r)r^{n/2}J_{n/2-1}(r|\xi|)dr,\quad \xi\in\mathbb{R}^n, \tag{19.4}$$

where $J_\nu(z)$ is a Bessel function of the first kind of order ν, see the NIST Handbook [33, p. 217 *et seq.*], in particular p. 224, 10.9.4,

$$J_\nu(z)=\frac{(z/2)^\nu}{\sqrt{\pi}\,\Gamma\left(\nu+\frac12\right)}\int_{-1}^1(1-t^2)^{\nu-1/2}\cos(zt)dt,\quad z\geqslant 0,\ \operatorname{Re}\nu>-\tfrac12.$$

Proof (Bochner [7, pp. 187–188]) Fix $\xi\in\mathbb{R}^n$ and then construct a rotation matrix $R\in\mathbb{R}^{n\times n}$ such that $R\xi=|\xi|e_1$, $e_1=(1,0,\dots,0)$. Since R is orthogonal, we know from 19.3(v) that

$$\mathscr{F}[f(|x|)](\xi)=\mathscr{F}[f(|Rx|)](\xi)=\mathscr{F}[f(|x|)](R\xi)=\int_{\mathbb{R}^n}f(|x|)\,e^{ix_1|\xi|}\,dx.$$

Set $x' = (x_2, \ldots, x_n) \in \mathbb{R}^{n-1}$, $|x|^2 = |x'|^2 + x_1^2$ and use Corollary 16.20 for the variable x'.

$$\int_{\mathbb{R}^n} f(|x|)\, e^{ix_1|\xi|}\, dx = (n-1)\omega_{n-1} \int_0^\infty \int_{\mathbb{R}} f\left(\sqrt{\rho^2 + x_1^2}\right) e^{ix_1|\xi|}\, \rho^{n-2}\, dx_1\, d\rho.$$

Now we use polar coordinates $(r, \alpha) \in (0, \infty) \times (0, \pi)$ for $(x_1, \rho) \in \mathbb{R} \times (0, \infty)$ and get

$$\begin{aligned}
&\int_{\mathbb{R}^n} f(|x|)\, e^{ix_1|\xi|}\, dx \\
&\qquad = (n-1)\omega_{n-1} \int_0^\infty \int_0^\pi f(r) e^{ir|\xi|\cos\alpha} r^{n-1} \sin^{n-2}\alpha\, d\alpha\, dr \\
&\qquad = \int_0^\infty f(r) r^{n/2} \left\{ (n-1)\omega_{n-1} \int_0^\pi e^{ir|\xi|\cos\alpha} r^{n/2-1} \sin^{n-2}\alpha\, d\alpha \right\} dr.
\end{aligned}$$

In the expression $\{\cdots\}$ we change variables according to $t = -\cos\alpha$ and observe that $\sin^2\alpha = 1 - t^2$. This shows that the expression $\{\cdots\}$ is real-valued and equals $(2\pi)^{n/2}|\xi|^{-n/2+1} J_{n/2-1}(r|\xi|)$. □

Injectivity and Existence of the Inverse Transform

The Fourier transform of a finite measure is unique, i.e. $\widehat{\mu} = \widehat{\nu} \implies \mu = \nu$. This follows from an explicit inversion formula due to P. Lévy. For this we need the following classical result from Example 14.12.

Lemma 19.5 (sine integral)

$$\lim_{T\to\infty} \int_0^T \frac{\sin\xi}{\xi}\, d\xi = \frac{\pi}{2}.$$

In particular, we get for $x, a \in \mathbb{R}$

$$\lim_{T\to\infty} \int_0^T \frac{\sin((a-x)\xi)}{\xi}\, d\xi = \lim_{T\to\infty} \int_0^{(a-x)T} \frac{\sin\eta}{\eta}\, d\eta = \begin{cases} \frac{\pi}{2}, & x < a, \\ 0, & x = a, \\ -\frac{\pi}{2}, & x > a. \end{cases} \tag{19.5}$$

Theorem 19.6 (Lévy) *If μ is a finite measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, then the following inversion formula holds: for all $a < b$,*

$$\frac{1}{2}\mu\{a\} + \mu(a,b) + \frac{1}{2}\mu\{b\} = \lim_{T\to\infty} \int_{-T}^T \frac{e^{ib\xi} - e^{ia\xi}}{i\xi} \widehat{\mu}(\xi) d\xi. \tag{19.6}$$

Caution The integral on the right-hand side of (19.6) is a so-called *improper Cauchy principal value*. In general, it cannot be written as a Lebesgue integral.

Proof of Theorem 19.6 Using $\operatorname{Im} e^{i\xi} = \sin\xi$ we find for all $a < b$

$$\begin{aligned}
&\int_{-T}^{T} \frac{e^{i(b-x)\xi} - e^{i(a-x)\xi}}{i\xi}\,d\xi \\
&\quad = \int_{-T}^{0} \frac{e^{i(b-x)\xi} - e^{i(a-x)\xi}}{i\xi}\,d\xi + \int_{0}^{T} \frac{e^{i(b-x)\xi} - e^{i(a-x)\xi}}{i\xi}\,d\xi \\
&\quad = \int_{0}^{T} \frac{e^{-i(a-x)\xi} - e^{-i(b-x)\xi}}{i\xi}\,d\xi + \int_{0}^{T} \frac{e^{i(b-x)\xi} - e^{i(a-x)\xi}}{i\xi}\,d\xi \\
&\quad = 2\int_{0}^{T} \frac{\sin((b-x)\xi)}{\xi}\,d\xi - 2\int_{0}^{T} \frac{\sin((a-x)\xi)}{\xi}\,d\xi.
\end{aligned}$$

This yields, in view of (19.5),

$$\lim_{T\to\infty} \int_{-T}^{T} \frac{e^{i(b-x)\xi} - e^{i(a-x)\xi}}{i\xi}\,d\xi = \begin{cases} 0, & \text{if } x < a \text{ or } x > b, \\ \pi, & \text{if } x = a \text{ or } x = b, \\ 2\pi, & \text{if } a < x < b. \end{cases}$$

With a formal calculation we obtain

$$\begin{aligned}
\lim_{T\to\infty} \int_{-T}^{T} \frac{e^{ib\xi} - e^{ia\xi}}{i\xi}\widehat{\mu}(\xi)d\xi &= \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{ib\xi} - e^{ia\xi}}{i\xi} \int e^{-ix\xi}\mu(dx)d\xi \\
&\overset{(F)}{=} \lim_{T\to\infty} \frac{1}{2\pi} \int \int_{-T}^{T} \frac{e^{ib\xi} - e^{ia\xi}}{i\xi} e^{-ix\xi}\,d\xi\,\mu(dx) \\
&= \lim_{T\to\infty} \frac{1}{2\pi} \int \int_{-T}^{T} \frac{e^{i(b-x)\xi} - e^{i(a-x)\xi}}{i\xi}\,d\xi\,\mu(dx) \\
&\overset{(L)}{=} \int \left[\frac{1}{2}\mathbb{1}_{\{a\}} + \mathbb{1}_{(a,b)} + \frac{1}{2}\mathbb{1}_{\{b\}}\right] d\mu \\
&= \tfrac{1}{2}\mu\{a\} + \mu(a,b) + \tfrac{1}{2}\mu\{b\}.
\end{aligned}$$

At the points marked with (F) and (L), we use Fubini's theorem (Corollary 14.9) and Lebesgue's dominated convergence theorem (Theorem 12.2). Let us check the requirements for these theorems:

Justification of (F). From the elementary inequality

$$\left|e^{it} - e^{is}\right| = \left|\int_{s}^{t} ie^{iu}\,du\right| \leqslant \int_{s}^{t} \left|ie^{iu}\right| du = t - s \quad \forall s \leqslant t$$

we conclude that

$$\int_{-T}^{T} \left|\frac{e^{i(b-x)\xi} - e^{i(a-x)\xi}}{i\xi}\right| d\xi \leqslant \int_{-T}^{T} (b-a)\,d\xi = 2(b-a)T \in L^1(\mu),$$

which ensures that we can indeed use Fubini's theorem.

Justification of (L). Lemma 19.5 shows that the expression

$$T \mapsto \int_0^T \frac{\sin u}{u}\,du$$

is bounded, and hence μ-integrable. Therefore, we may use dominated convergence. $\square$

Fubini's theorem allows us to get a multivariate version of Theorem 19.6. As usual, $\mathscr{J}^o$ is the family of open rectangles in $\mathbb{R}^n$,

$$I := \bigtimes_{k=1}^{n} (a_k, b_k) \quad \text{and} \quad \partial I := \bigtimes_{k=1}^{n} [a_k, b_k] \Big\backslash \bigtimes_{k=1}^{n} (a_k, b_k).$$

Corollary 19.7 *If μ is a finite measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, then the following inversion formula holds for all open rectangles $I \in \mathscr{J}^o$ such that $\mu(\partial I) = 0$:*

$$\mu(I) = \lim_{T\to\infty} \int_{-T}^{T} \cdots \int_{-T}^{T} \prod_{k=1}^{n} \frac{e^{ib_k\xi_k} - e^{ia_k\xi_k}}{i\xi_k} \widehat{\mu}(\xi)\, d\xi_1 \ldots d\xi_n. \tag{19.7}$$

Corollary 19.8 (i) *If μ, ν are finite measures, then $\widehat{\mu} = \widehat{\nu} \implies \mu = \nu$.*
(ii) *If $f, g \in L^1(\lambda^n)$, then $\widehat{f} = \widehat{g} \implies f = g$ Lebesgue a.e.*

Proof (i) Consider rectangles $I := \bigtimes_{k=1}^{n} (a_k, b_k) \supset \bigtimes_{k=1}^{n} (a_k + \epsilon, b_k - \epsilon) =: I^\epsilon$, where $\epsilon < \frac{1}{2} \min_{1 \leqslant k \leqslant n} (b_k - a_k)$.

Since $\mu + \nu$ is a finite measure, there exist at most finitely many sets whose boundary ∂I^ϵ has strictly positive $\mu + \nu$-measure $> 1/m$; in particular, there are at most countably many sets ∂I^ϵ such that $\mu(\partial I^\epsilon) + \nu(\partial I^\epsilon) > 0$. [✍] Consequently, there is a sequence

$$I_k \uparrow I, \ \mu(\partial I_k) + \nu(\partial I_k) = 0, \quad \text{so} \quad \mu(I_k) \uparrow \mu(I), \quad \nu(I_k) \uparrow \nu(I).$$

Now Corollary 19.7 implies that $\mu(I) = \nu(I)$ for all $I \in \mathscr{J}^o$, and the claim follows from the uniqueness theorem for measures (Theorem 5.7), since $\mathscr{J}^o$ is a $\cap$-stable generator of the Borel sets.

(ii) is a consequence of Corollary 11.7, since we know from part (i) that

$$\widehat{f} = \widehat{g} \implies \widehat{f^+ + g^-} = \widehat{f^- + g^+} \overset{\text{(i)}}{\implies} (f^+ + g^-)\lambda^n = (f^- + g^+)\lambda^n. \quad \square$$

Theorem 19.6 and Corollary 19.7 also give the following inversion formula for *functions*.

Corollary 19.9 *Let μ be a finite measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and assume that $\widehat{\mu} \in L^1_{\mathbb{C}}(\lambda^n)$. In this case, $\mu(dx) = u(x)dx$, where*

$$u(x) = \int \widehat{\mu}(\xi) e^{i\langle x, \xi\rangle}\, d\xi. \tag{19.8}$$

If both $u \in L^1_{\mathbb{C}}(\lambda^n)$ and $\widehat{u} \in L^1_{\mathbb{C}}(\lambda^n)$, then $u(x) = \int \widehat{u}(\xi) e^{i\langle x, \xi\rangle}\, d\xi$.

Proof Let

$$I = \bigtimes_{k=1}^{n} (a_k, b_k) \in \mathscr{J}^o$$

such that $\mu(\partial I) = 0$. By assumption $\widehat{\mu}(\xi)\mathbb{1}_I(x) \in L^1(\lambda^n \times \lambda^n)$, which allows us to use Fubini's theorem. Because of Corollary 19.7 we see

$$\begin{aligned}\int_I u(x)dx &= \int_I \int \widehat{\mu}(\xi)e^{i\langle x,\xi\rangle}\, d\xi\, dx = \int \widehat{\mu}(\xi) \int_I e^{i\langle x,\xi\rangle}\, dx\, d\xi \\ &= \lim_{T\to\infty} \int_{[-T,T]^n} \widehat{\mu}(\xi) \prod_{k=1}^{n} \frac{e^{ib_k\xi_k} - e^{ia_k\xi_k}}{i\xi_k}\, d\xi = \mu(I).\end{aligned}$$

Arguing as in Corollary 19.8 we see that $\mu(dx) = u(x)dx$ for some $u \geqslant 0$ as $B \mapsto \int_B u(x)dx$ is a (positive) measure. The case of a complex valued $u \in L^1_{\mathbb{C}}(\lambda^n)$ follows in a similar fashion if we decompose u into $(\operatorname{Re} u)^{\pm}$ and $(\operatorname{Im} u)^{\pm}$. [✍] □

Corollary 19.9 justifies the following definition.

Definition 19.10 Let μ be a finite measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. The *inverse Fourier transform* is given by

$$\widetilde{\mu}(x) = \mathscr{F}^{-1}\mu(x) = \int_{\mathbb{R}^n} e^{i\langle x,\xi\rangle}\mu(d\xi). \tag{19.9}$$

The inverse Fourier transform of a function $u \in L^1_{\mathbb{C}}(\lambda^n)$ is

$$\widetilde{u}(x) = \mathscr{F}^{-1}u(x) = \int_{\mathbb{R}^n} e^{i\langle x,\xi\rangle}u(\xi)d\xi. \tag{19.10}$$

The Fourier and inverse Fourier transforms are (structurally) almost identical. In fact, we have the following relations [✍]:

$$\widetilde{\mu}(x) = (2\pi)^n\widehat{\mu}(-x), \quad \overline{\widehat{\mu}(x)} = (2\pi)^{-n}\widetilde{\mu}(x) \quad \text{and} \quad \overline{\widehat{u}(x)} = (2\pi)^{-n}\widetilde{\overline{u}}(x).$$

The Convolution Theorem

An important feature of the (inverse) Fourier transform is the property that it trivializes the convolution product of measures and functions. Recall from Definition 15.4 that the convolution of two (finite) measures μ, ν is the (finite) measure

$$\mu \star \nu(B) = \iint \mathbb{1}_B(x+y)\mu(dx)\nu(dy) \quad \forall B \in \mathscr{B}(\mathbb{R}^n).$$

Theorem 19.11 (convolution theorem) *If μ, ν are finite measures on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, then*

$$\widehat{\mu \star \nu}(\xi) = (2\pi)^n \widehat{\mu}(\xi)\widehat{\nu}(\xi) \quad \textit{and} \quad \widecheck{\mu \star \nu}(\xi) = \widecheck{\mu}(\xi)\widecheck{\nu}(\xi). \tag{19.11}$$

Proof By definition, $\int \mathbb{1}_B(z)\mu \star \nu(dz) = \iint \mathbb{1}_B(x+y)\mu(dx)\nu(dy)$ and paragraph 15.5(iv) extends this equality to $L^1(\mu \star \nu)$ and – considering real and imaginary parts – to $L^1_{\mathbb{C}}(\mu \star \nu)$. Consequently,

$$\int e^{-i\langle z,\xi\rangle} \mu \star \nu(dz) = \iint e^{-i\langle x+y,\xi\rangle} \mu(dx)\nu(dy) = \int e^{-i\langle x,\xi\rangle} \mu(dx) \int e^{-i\langle y,\xi\rangle} \nu(dy),$$

which proves both formulae in (19.11). □

The following theorem shows that the Fourier transform is a symmetric operator.

Theorem 19.12 *For a finite measure μ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and $u \in L^1(\lambda^n)$ we have*

$$\int \widehat{u}(x)\mu(dx) = \int u(\xi)\widehat{\mu}(\xi)d\xi.$$

Proof Since μ is a finite measure and $u \in L^1(\lambda^n)$, we know that $\widehat{\mu} \in C_b(\mathbb{R}^n)$ and $\widehat{u} \in C_b(\mathbb{R}^n)$. Thus, all expressions appearing in the statement of the theorem are well-defined. In particular, we can use Fubini's theorem and get

$$\begin{aligned}\int \widehat{u}(x)\mu(dx) &= (2\pi)^{-n} \iint u(\xi)e^{-i\langle x,\xi\rangle}\, d\xi\, \mu(dx) \\ &= (2\pi)^{-n} \iint e^{-i\langle x,\xi\rangle}\, \mu(dx)u(\xi)\, d\xi = \int u(\xi)\widehat{\mu}(\xi)d\xi. \qquad \square\end{aligned}$$

The Riemann–Lebesgue Lemma

We know from 19.3 that $\mathscr{F}(L^1(\lambda^n)) \subset C_b(\mathbb{R}^n)$. If $f(x) = \mathbb{1}_{[a_1,b_1)\times\cdots\times[a_n,b_n)}(x)$, Example 19.2(i) shows that

$$\widehat{\mathbb{1}}_{[a_1,b_1)\times\cdots\times[a_n,b_n)}(\xi) = \prod_{k=1}^{n} \frac{e^{-i\xi_k a_k} - e^{-i\xi_k b_k}}{2\pi i\, \xi_k}. \tag{19.12}$$

This Fourier transform converges to 0 as $|\xi| \to \infty$, i.e. it is a continuous function which vanishes at infinity,

$$C_\infty(\mathbb{R}^n) := \left\{ u \in C(\mathbb{R}^n) : \lim_{|x|\to\infty} |u(x)| = 0 \right\}.$$

This is true for all Fourier transforms of L^1-functions.

Theorem 19.13 (Riemann–Lebesgue) *If $u \in L^1(\lambda^n)$, then $\widehat{u} \in C_\infty(\mathbb{R}^n)$.*

Proof Because of 19.3 it suffices to show that $\lim_{|\xi|\to\infty} \widehat{u}(\xi) = 0$. Let $u \in L^1(\lambda^n)$ and $I_R := [-R, R)^n$.

Step 1. Fourier transforms of bounded Borel sets are in $C_\infty(\mathbb{R}^n)$. We define $\mathscr{D} := \{B \in \mathscr{B}(I_R) : \widehat{\mathbb{1}}_B \in C_\infty(\mathbb{R}^n)\}$, $R > 0$, and claim that $\mathscr{D}$ is a Dynkin system.

(D_1) Indeed, (19.12) shows that $I_R \in \mathscr{D}$.

(D_2) Assume that $D \in \mathscr{D}$. With (D_1) we see that $\widehat{\mathbb{1}}_{D^c} = \widehat{\mathbb{1}}_{I_R} - \widehat{\mathbb{1}}_D \in C_\infty(\mathbb{R}^n)$, and so $D^c \in \mathscr{D}$.

(D_3) Let $(D_i)_i \subset \mathscr{D}$ be a sequence of disjoint sets and set $D = \bigcupdot_i D_i$. Since $\sum_{i=1}^\infty \mathbb{1}_{D_i} \leqslant \mathbb{1}_{I_R}$, we can use Lebesgue's dominated convergence theorem to see that $\lim_{N\to\infty} \sum_{i=1}^N \mathbb{1}_{D_i} = \mathbb{1}_D$ in $L^1(\lambda^n)$. Therefore,

$$\left\| \widehat{\mathbb{1}}_D - \sum_{i=1}^N \widehat{\mathbb{1}}_{D_i} \right\|_\infty \overset{19.3}{\leqslant} (2\pi)^{-n} \left\| \mathbb{1}_D - \sum_{i=1}^N \mathbb{1}_{D_i} \right\|_1 \xrightarrow[N\to\infty]{} 0,$$

which means that $\widehat{\mathbb{1}}_D$ is the uniform limit of $\sum_{i=1}^N \widehat{\mathbb{1}}_{D_i} \in C_\infty(\mathbb{R}^n)$. In particular, [✍] $\widehat{\mathbb{1}}_D \in C_\infty(\mathbb{R}^n)$ and so $D \in \mathscr{D}$.

Because of (19.12) we know that $I_R \cap \mathscr{J} \subset \mathscr{D}$. Since $I_R \cap \mathscr{J}$ is a $\cap$-stable generator of $\mathscr{B}(I_R) = I_R \cap \mathscr{B}(\mathbb{R}^n)$, we can invoke Theorem 5.5 to see $\mathscr{D} = \mathscr{B}(I_R)$.

Step 2. Because of the first part of the proof we know that $\mathscr{F}(f) \in C_\infty(\mathbb{R}^n)$ for all simple functions $f \in \mathcal{E}(\mathscr{B}(\mathbb{R}^n))$ whose support is in a rectangle of side-length $R > 0$. Let $u \in L^1(\lambda^n)$. Using Corollary 8.9 we can construct a sequence $(f_i)_{i\in\mathbb{N}} \subset \mathcal{E}(\mathscr{B}(\mathbb{R}^n))$ such that $f_i \to u$, $\operatorname{supp} f_i$ is compact and $|f_i| \leqslant |u|$. Lebesgue's dominated convergence theorem applies and shows that

$$\lim_{i\to\infty} \|u - f_i\|_1 = 0.$$

Step 3. For every $\epsilon > 0$ there is some N_ϵ such that for all $i \geqslant N_\epsilon$ and $\xi \in \mathbb{R}^n$

$$\begin{aligned} |\widehat{u}(\xi)| &\leqslant \left|\widehat{u}(\xi) - \widehat{f}_i(\xi)\right| + \left|\widehat{f}_i(\xi)\right| \\ &\overset{19.3}{\leqslant} (2\pi)^{-n} \|u - f_i\|_1 + \left|\widehat{f}_i(\xi)\right| \\ &\leqslant \epsilon + \left|\widehat{f}_i(\xi)\right| \xrightarrow[|\xi|\to\infty]{\text{Step 2}} \epsilon \xrightarrow[\epsilon\to 0]{} 0. \end{aligned}$$

□

The Wiener Algebra, Weak Convergence and Plancherel's Theorem

Let us continue our investigation of the range of the Fourier transform. The space of L^1-functions where Definition 19.10 is indeed the inverse of the Fourier transform is the so-called *Wiener algebra*:

$$\mathcal{A}(\mathbb{R}^n) = \left\{ u \in L^1(\lambda^n) : \widehat{u} \in L^1(\lambda^n) \right\}.$$

The following lemma shows, among other things, that $\mathcal{A}(\mathbb{R}^n)$ is an algebra.

Lemma 19.14 (properties of the Wiener algebra)

(i) $u \in \mathcal{A}(\mathbb{R}^n) \iff \widehat{u} \in \mathcal{A}(\mathbb{R}^n)$.
(ii) $u \in \mathcal{A}(\mathbb{R}^n) \implies u \in C_\infty(\mathbb{R}^n)$.
(iii) $u \in \mathcal{A}(\mathbb{R}^n) \implies u \in L^p(\lambda^n)$, $1 \leqslant p < \infty$.
(iv) $u, v \in \mathcal{A}(\mathbb{R}^n) \implies u \star v \in \mathcal{A}(\mathbb{R}^n)$.
(v) $u, v \in \mathcal{A}(\mathbb{R}^n) \implies uv \in \mathcal{A}(\mathbb{R}^n)$.

Proof (i) follows from Corollary 19.9.
(ii) Since $\mathcal{A}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$, this is a consequence of (i) and Theorem 19.13.
(iii) Because of (ii) we find $\int |u(x)|^p \, dx \leqslant \|u\|_\infty^{p-1} \int |u(x)| dx < \infty$.
(iv) A combination of Young's inequality (Theorem 15.6) and (iii) shows that $u \star v \in L^1(\mathbb{R}^n)$. Using the convolution theorem (Theorem 19.11) and part (ii) now yields

$$\|\widehat{u \star v}\|_1 = (2\pi)^n \|\,\widehat{u}\,\widehat{v}\,\|_1 \leqslant (2\pi)^n \|\,\widehat{u}\,\|_\infty \|\,\widehat{v}\,\|_1 < \infty.$$

(v) Because of (i) and (iv) we have $\widehat{u} \star \widehat{v} \in \mathcal{A}(\mathbb{R}^n)$. With the convolution theorem (Theorem 19.11) we see that

$$\widecheck{\widehat{u} \star \widehat{v}} = \check{\widehat{u}} \cdot \check{\widehat{v}} \overset{19.9}{=} uv \overset{19.9}{\Longrightarrow} \widehat{u} \star \widehat{v} = \widehat{uv}.$$

The claim follows from (i) as $\widehat{u} \star \widehat{v} \in \mathcal{A}(\mathbb{R}^n)$. $\square$

We have seen in Example 19.2(iii) that the function $g_t(x) = (2\pi t)^{-n/2} e^{-|x|^2/2t}$ has the Fourier transform $\widehat{g}_t(\xi) = (2\pi)^{-n} e^{-t|\xi|^2/2}$.

Lemma 19.15 *If $u \in L^1(\lambda^n)$, then $u \star g_t \in \mathcal{A}(\mathbb{R}^n)$.*

Proof Since $u \in L^1(\lambda^n)$ and $g_t \in L^1(\lambda^n)$, Young's inequality (15.7) proves that $u \star g_t \in L^1(\lambda^n)$. Using the fact that $\widehat{g}_t \in L^1(\lambda^n)$, we see that

$$|\widehat{u \star g_t}| \overset{(19.11)}{=} (2\pi)^n |\,\widehat{u}\,| \cdot |\,\widehat{g}_t\,| \overset{19.3}{\leqslant} (2\pi)^{-n} \|u\|_1 e^{-t|\cdot|^2/2} \in L^1(\lambda^n),$$

and the assertion follows from Lemma 19.14(i). $\square$

The next result is an L^∞-version of Theorem 15.11.

Lemma 19.16 (approximation of unity) *If* $u \in C_b(\mathbb{R}^n)$ *is uniformly continuous, then* $\lim_{t\to 0} \|u - u \star g_t\|_\infty = 0$.

Proof Since $u \in C_b(\mathbb{R}^n)$ is uniformly continuous, we find for every $\epsilon > 0$ some $\delta > 0$ such that

$$|u(x) - u(y)| < \epsilon \quad \forall |x-y| < \delta.$$

Observe that $\int g_t(y)dy = \int g_t(x-y)dy = 1$. Therefore,

$$\begin{aligned}
|u(x) - u \star g_t(x)| &= \left| \int (u(x) - u(y)) g_t(x-y)dy \right| \\
&\leqslant \int_{|x-y|<\delta} \underbrace{|u(x) - u(y)|}_{\leqslant \epsilon} g_t(x-y)dy + 2\|u\|_\infty \int_{|x-y|\geqslant\delta} g_t(x-y)dy \\
&\leqslant \epsilon + 2\|u\|_\infty \int_{|z|\geqslant\delta} (2\pi t)^{-n/2} e^{-|z^2|/2t}\, dz \\
&= \epsilon + 2\|u\|_\infty \int_{\sqrt{t}|y|\geqslant\delta} (2\pi)^{-n/2} e^{-|y|^2/2}\, dy \\
&\xrightarrow[t\to 0]{\text{monotone conv.}} \epsilon \xrightarrow[\epsilon\to 0]{} 0,
\end{aligned}$$

where all limits are uniform in the variable x. □

The Fourier transform not only 'trivializes' the convolution product but also can be used to describe the convergence of measures.

Definition 19.17 Let $(\mu_k)_{k\in\mathbb{N}}$ and μ be finite measures on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. The sequence of measures *converges weakly* if

$$\lim_{k\to\infty} \int_{\mathbb{R}^n} u(x)\mu_k(dx) = \int_{\mathbb{R}^n} u(x)\mu(dx) \quad \forall u \in C_b(\mathbb{R}^n).$$

Theorem 19.18 (Fourier transform and weak convergence) *Let* μ *and* $(\mu_k)_{k\in\mathbb{N}}$ *be finite measures on* $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$.

$$\widehat{\mu}_k(\xi) \xrightarrow[k\to\infty]{\textit{for every } \xi} \widehat{\mu}(\xi) \iff \mu_k \xrightarrow[k\to\infty]{\textit{weakly}} \mu. \tag{19.13}$$

Proof The direction '$\Leftarrow$' follows from $e^{-i\langle\cdot,\xi\rangle} \in C_b(\mathbb{R}^n)$.

The implication '$\Rightarrow$' is more complicated.

Step 1. Observe that $\mu_k(\mathbb{R}^n) = (2\pi)^n \widehat{\mu}_k(0) \to (2\pi)^n \widehat{\mu}(0) = \mu(\mathbb{R}^n)$.

Step 2. The assertion holds for $u \in \mathcal{A}(\mathbb{R}^n)$. Pick $\phi \in \mathcal{A}(\mathbb{R}^n)$. Because of Theorem 19.12, we see that

$$\int \widehat{\phi}(x)\mu_k(dx) = \int \phi(\xi)\widehat{\mu}_k(\xi)d\xi \xrightarrow[k\to\infty]{\text{dom. conv.}} \int \phi(\xi)\widehat{\mu}(\xi)d\xi = \int \widehat{\phi}(x)\mu(dx).$$

Note that we can use dominated convergence since $|\widehat{\mu}_k(\xi)| \leqslant \sup_k \mu_k(\mathbb{R}^n) < \infty$ is an integrable majorant. By Lemma 19.14(i) we can obtain any $u \in \mathcal{A}(\mathbb{R}^n)$ as $\widehat{\phi} \in \mathcal{A}(\mathbb{R}^n)$, and the claim follows.

Step 3. The assertion holds for $u \in C_c(\mathbb{R}^n)$. By Lemma 19.16 we find for $u \in C_c(\mathbb{R}^n)$ and $\epsilon > 0$ some $u_\epsilon \in \mathcal{A}(\mathbb{R}^n)$ such that $\|u - u_\epsilon\|_\infty \leqslant \epsilon$. Therefore,

$$\begin{aligned}\left|\int u\,d\mu_k - \int u\,d\mu\right| &\leqslant \int |u - u_\epsilon|d\mu_k + \int |u - u_\epsilon|d\mu + \left|\int u_\epsilon\,d\mu_k - \int u_\epsilon\,d\mu\right| \\ &\leqslant [\mu_k(\mathbb{R}^n) + \mu(\mathbb{R}^n)]\,\epsilon + \left|\int u_\epsilon\,d\mu_k - \int u_\epsilon\,d\mu\right| \\ &\xrightarrow[k\to\infty]{\text{Step 1, 2}} 2\mu(\mathbb{R}^n)\epsilon \xrightarrow[\epsilon\to 0]{} 0.\end{aligned}$$

Step 4. Let $u \in C_b(\mathbb{R}^n)$ and $R > 0$. Pick $\chi_R \in C_c(\mathbb{R}^n)$ such that $\chi_R(x) = 1$ for $x \in [-R, R]^n$ and $0 \leqslant \chi_R \leqslant 1$. For $u_R := u\chi_R$ we see

$$\begin{aligned}&\left|\int u\,d\mu_k - \int u\,d\mu\right| \\ &\quad\leqslant \int |u - u_R|d\mu_k + \int |u - u_R|d\mu + \left|\int u_R\,d\mu_k - \int u_R\,d\mu\right| \\ &\quad\leqslant \|u\|_\infty \int (1 - \chi_R)d\mu_k + \|u\|_\infty \int (1 - \chi_R)d\mu + \left|\int u_R\,d\mu_k - \int u_R\,d\mu\right| \\ &\quad\xrightarrow[k\to\infty]{\text{Step 1, 3}} 2\|u\|_\infty \int (1 - \chi_R)d\mu \xrightarrow[R\to\infty]{\text{dom. conv.}} 0.\end{aligned}$$ □

The Wiener algebra is a fairly large set. This can be seen from the following density result.

Lemma 19.19 *The Wiener algebra $\mathcal{A}(\mathbb{R}^n)$ is dense in $L^p_{\mathbb{C}}(\lambda^n)$, $1 \leqslant p < \infty$, and $C_\infty(\mathbb{R}^n, \mathbb{C})$.*

Proof Without loss of generality we can restrict ourselves to real-valued functions – otherwise we consider real and imaginary parts.

Density in L^p. Let $u \in L^p(\lambda^n)$. Since $C_c(\mathbb{R}^n)$ is dense in $L^p(\lambda^n)$, it suffices to approximate $f \in C_c(\mathbb{R}^n)$ in the L^p-norm by functions from $\mathcal{A}(\mathbb{R}^n)$. Since we have

$f \in L^1(\mathbb{R}^n)$, Lemma 19.15 tells us that $f \star g_t \in \mathcal{A}(\mathbb{R}^n)$. It is not hard to adapt the proof of Theorem 15.11 to show that $\lim_{t\to 0} \|f - f \star g_t\|_p = 0$ holds. Thus

$$\|u - f \star g_t\|_p \leqslant \|u - f\|_p + \|f - f \star g_t\|_p \xrightarrow[t\to 0]{\text{as in 15.11}} \|u - f\|_p \xrightarrow[f\to u]{C_c \subset L^p \text{ dense}} 0.$$

Density in $C_\infty(\mathbb{R}^n)$. Observe, first of all, that $C_c(\mathbb{R}^n)$ is dense in $C_\infty(\mathbb{R}^n)$. Fix $u \in C_\infty(\mathbb{R}^n)$ and $\epsilon > 0$. By the definition of $C_\infty(\mathbb{R}^n)$ there is some $R = R_\epsilon > 0$ such that $|u(x)| < \epsilon$ for all $|x| \geqslant R$. Let $\chi_R \in C_c(\mathbb{R}^n)$ satisfy $\mathbb{1}_{B_R(0)} \leqslant \chi_R \leqslant \mathbb{1}_{B_{2R}(0)}$ and set $u_R := u\chi_R \in C_c(\mathbb{R}^n)$. This shows that

$$\|u - u_R\|_\infty = \sup_{|x|\geqslant R} (|u(x)|(1 - \chi_R(x))) \leqslant \epsilon.$$

Let u and u_R as above. By Lemma 19.16 we have for all $t \leqslant h(R, \epsilon)$

$$\|u - u_R \star g_t\|_\infty \leqslant \|u - u_R\|_\infty + \|u_R - u_R \star g_t\|_\infty \leqslant \epsilon + \|u_R - u_R \star g_t\|_\infty \overset{19.16}{\leqslant} 2\epsilon.$$

Since $u_R \in L^1(\lambda^n)$, Lemma 19.15 shows $u_R \star g_t \in \mathcal{A}(\mathbb{R}^n)$ and the claim follows. □

We close this section with a classic result which allows us to extend the Fourier transform to square integrable functions.

Theorem 19.20 (Plancherel) *If* $u \in L^2_{\mathbb{C}}(\lambda^n) \cap L^1_{\mathbb{C}}(\lambda^n)$*, then*

$$\|\widehat{u}\|_2 = (2\pi)^{-n/2}\|u\|_2. \tag{19.14}$$

In particular, there is a continuous extension $\mathscr{F} : L^2_{\mathbb{C}}(\lambda^n) \to L^2_{\mathbb{C}}(\lambda^n)$.

Proof Let $u \in \mathcal{A}(\mathbb{R}^n)$. Since $u \in L^2_{\mathbb{C}}(\lambda^n) \cap L^1_{\mathbb{C}}(\lambda^n)$ and $\widehat{u} \in L^1_{\mathbb{C}}(\lambda^n)$, we see from Theorem 19.12

$$\begin{aligned} \int |\widehat{u}(\xi)|^2 \, d\xi = \int \widehat{u}(\xi)\overline{\widehat{u}(\xi)} d\xi \;&=\; (2\pi)^{-n} \int \widehat{u}(\xi)\breve{\overline{u}}(\xi) d\xi \\ &\overset{19.12}{=} (2\pi)^{-n} \int u(x)\mathscr{F}\left[\breve{\overline{u}}\right](x) dx \\ &\overset{19.9}{=} (2\pi)^{-n} \int |u(x)|^2 \, dx. \end{aligned}$$

Since the Wiener algebra is dense in $L^2_{\mathbb{C}}(\lambda^n)$, see Lemma 19.19, we can approximate any $u \in L^2_{\mathbb{C}}(\mathbb{R}^n)$ with a sequence $(u_k)_k \subset \mathcal{A}(\mathbb{R}^n)$. Therefore,

$$\|\mathscr{F}u_k - \mathscr{F}u_m\|_2 = (2\pi)^{-n/2}\|u_k - u_m\|_2 \xrightarrow[k,m\to\infty]{} 0.$$

This shows that $(\mathscr{F}u_k)_{k\in\mathbb{N}}$ is a Cauchy sequence in $L^2_{\mathbb{C}}(\lambda^n)$. Because of the completeness, the limit $L^2\text{-}\lim_{k\to\infty} \mathscr{F}u_k$ exists (and it is independent of the approximating sequence [✍]!) and *defines* an element $\mathscr{F}u \in L^2_{\mathbb{C}}(\lambda^n)$; the identity (19.14) remains valid since the L^2-norm is continuous.[✍] □

The Fourier Transform in $\mathcal{S}(\mathbb{R}^n)$

Throughout this section we use the shorthand ∂_k to denote the partial derivative $\partial/\partial x_k$, $k=1,\dots,n$. The following result is an immediate consequence of the continuity and differentiability lemmas for parameter-dependent integrals, Theorem 12.4 and 12.5. [✍]

Lemma 19.21 *Assume that μ is a finite measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and $1 \leqslant k \leqslant n$.*

(i) *If $\int |x_k|\mu(dx) < \infty$, then $\partial_k\widehat{\mu} \in C_b(\mathbb{R}^n)$ and $\partial_k\widehat{\mu}(\xi) = \widehat{(-i)x_k\mu}(\xi)$.*
(ii) *If $u, x_k u \in L^1(\lambda^n)$, then $\partial_k\widehat{u} \in C_b(\mathbb{R}^n)$ and $\partial_k\widehat{u}(\xi) = \widehat{(-i)x_k u}(\xi)$.*
(iii) *If $\partial_k u \in L^1(\lambda^n)$ and $u \in C_\infty(\mathbb{R}^n) \cap L^1(\lambda^n)$, then $\widehat{\partial_k u}(\xi) = i\xi_k\widehat{u}(\xi)$.*

For sufficiently regular functions u we can iterate Lemma 19.21. The following notation involving multi-indices $\alpha = (\alpha_1, \dots, \alpha_n), \beta = (\beta_1, \dots, \beta_n) \in \mathbb{N}_0^n$ will be helpful:

$$x^\alpha := \prod_{k=1}^n x_k^{\alpha_k} \quad \text{and} \quad \partial^\beta := \frac{\partial^{\beta_1+\cdots+\beta_n}}{\partial x_1^{\beta_1}\cdots\partial x_n^{\beta_n}}.$$

Definition 19.22 The Schwartz space $\mathcal{S}(\mathbb{R}^n)$ of rapidly decreasing smooth functions consists of complex-valued $u \in C^\infty(\mathbb{R}^n)$ which decrease, along with their derivatives, faster than any polynomial:

$$\sup_{x\in\mathbb{R}^n}\left|x^\alpha\,\partial^\beta u(x)\right| < \infty \quad \forall \alpha, \beta \in \mathbb{N}_0^n.$$

In particular, $(n + x_1^{2n} + \cdots + x_n^{2n})|u(x)| \leqslant c$ for all $u \in \mathcal{S}(\mathbb{R}^n)$. Using the elementary inequalities

$$\prod_{k=1}^n (1+x_k^2) \leqslant \Big[\max_{1\leqslant k\leqslant n}(1+x_k^2)\Big]^n \leqslant \sum_{k=1}^n (1+x_k^2)^n \overset{\text{Hölder}}{\leqslant} 2^{n-1}\sum_{k=1}^n(1+x_k^{2n})$$

in combination with the Fubini–Tonelli theorem (Theorem 14.8), we get for all $u \in \mathcal{S}(\mathbb{R}^n)$

$$\int |u(x)|dx \leqslant c_d \int\cdots\int \frac{dx_1\cdots dx_n}{\prod_{k=1}^n(1+x_k^2)} = c_d\prod_{k=1}^n\int\frac{dx_k}{(1+x_k^2)} < \infty.$$

Theorem 19.23 *If $u \in \mathcal{S}(\mathbb{R}^n)$, then $\widehat{u} \in \mathcal{S}(\mathbb{R}^n)$, i.e. $\mathscr{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$.*

Proof Since $\mathcal{S}(\mathbb{R}^n) \subset L^1(\lambda^n)$, the Fourier transform $\widehat{u}(\xi)$ exists. Moreover, for all multi-indices $\alpha, \beta, \gamma \in \mathbb{N}_0^n$

$$\partial_x^\beta \left[(-ix)^\alpha u(x)\right] = \sum_{\gamma_1=0}^{\beta_1} \cdots \sum_{\gamma_n=0}^{\beta_n} p_{\alpha,\beta,\gamma}(x) \partial_x^\gamma u(x),$$

where $p_{\alpha,\beta,\gamma}$ denotes a polynomial in x[1] As $\partial^\gamma u$ decays faster than any polynomial,

$$\partial_x^\beta \left[(-ix)^\alpha u(x)\right] \in L^1(\lambda^n).$$

This means that we can use Lemma 19.21 recursively to get[2]

$$(i\xi)^\beta \, \partial_\xi^\alpha \widehat{u}(\xi) = (i\xi)^\beta \mathscr{F}_{x\to\xi} \left[(-ix)^\alpha u(x)\right](\xi) = \mathscr{F}_{x\to\xi} \left[\partial_x^\beta \left\{(-ix)^\alpha u(x)\right\}\right](\xi).$$

Finally, using that $\partial_x^\beta\{(-ix)^\alpha u(x)\}$ is integrable, we conclude that

$$\left|(i\xi)^\beta \, \partial_\xi^\alpha \widehat{u}(\xi)\right| = \left|\mathscr{F}_{x\to\xi}\left[\partial_x^\beta \left\{(-ix)^\alpha u(x)\right\}\right](\xi)\right| \overset{19.3}{\leqslant} (2\pi)^{-n} \left\|\partial_x^\beta \left\{(-ix)^\alpha u(x)\right\}\right\|_{L^1}$$

where the norm on the right-hand side is finite and independent of ξ. □

Corollary 19.24 *The Fourier transform $\mathscr{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ is bijective and for all $u, v \in \mathcal{S}(\mathbb{R}^n)$*

(i) $\mathscr{F}^{-1} v(\xi) = \displaystyle\int_{\mathbb{R}^n} v(x) e^{i\langle x,\xi\rangle} \, dx.$

(ii) $\mathscr{F}^{-1} v(\xi) = (2\pi)^n \mathscr{F} v(-\xi)$ *and* $\mathscr{F} \circ \mathscr{F} u(x) = (2\pi)^{-n} u(-x).$

Proof Assertion (i) follows from the inversion formula (19.8) and Theorem 19.23. In particular, the existence of the inverse transform entails bijectivity. Part (ii) is shown with a simple direct calculation, see the remark following Definition 19.10. □

Problems

19.1. Calculate the Fourier transform of the following functions/measures on $\mathbb{R}$:

(a) $\mathbb{1}_{[-1,1]}(x)$, (b) $\mathbb{1}_{[-1,1]} \star \mathbb{1}_{[-1,1]}(x)$, (c) $e^{-x}\mathbb{1}_{[0,\infty)}(x)$,

(d) $e^{-|x|}$, (e) $1/(1+x^2)$, (f) $(1-|x|)\mathbb{1}_{[-1,1]}(x)$,

(g) $\displaystyle\sum_{k=0}^{\infty} \frac{t^k}{k!} e^{-t} \delta_k$, (h) $\displaystyle\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \delta_k$.

[1] With some effort one can make this explicit using the Leibniz formula for derivatives of products.

[2] For clarity, we use $\mathscr{F}_{x\to\xi}[u(x)](\xi)$ to denote $\mathscr{F}u(\xi) = \widehat{u}(\xi)$.

19.2. Extend Plancherel's theorem (Theorem 19.20) to show that

$$\int \widehat{u}(\xi)\overline{\widehat{v}(\xi)}d\xi = (2\pi)^{-n}\int u(x)\overline{v(x)}dx \quad \forall u, v \in L^2_{\mathbb{C}}(\lambda^n).$$

19.3. Let μ be a finite measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and set $\chi(\xi) := \widehat{\mu}(\xi)$. Show that χ is real-valued if, and only if, μ is symmetric w.r.t. the origin, i.e. $\widetilde{\mu}(B) := \mu(-B) = \mu(B)$.

19.4. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric, positive definite matrix. Find the Fourier transform of the function $e^{-\langle x, Ax\rangle}$.

19.5. Assume that $u \in L^1(\lambda^1) \cap L^\infty(\lambda^1)$ and $\widehat{u} \geqslant 0$. Show that $\widehat{u} \in L^1(\lambda^1)$. Extend the assertion to $u \in L^2(\lambda^1)$
[Hint: estimate $\int \widehat{u}\,\widehat{g}_t\,d\xi$, where g_t is as in Example 19.2(iii). Use monotone convergence for the limit $t \to 0$.]

19.6. Let μ be a finite measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. Prove P. Lévy's *truncation inequality*:

$$\mu\left(\mathbb{R}^n \setminus [-2R, 2R]^n\right) \leqslant 2\left(\frac{R}{2}\right)^n \int_{[-1/R,1/R]^n} \big(\mu(\mathbb{R}^n) - \operatorname{Re}\widecheck{\mu}(\xi)\big)d\xi.$$

[Hint: show that the right-hand side equals

$$2\int \left(1 - \prod_1^n \frac{\sin(x_n/R)}{x_n/R}\right)\mu(dx);$$

shrink the integration domain to $\mathbb{R}^n \setminus [-2R, 2R]^n$ and observe that $0 \leqslant \frac{1}{2}\sin 2 \leqslant \frac{1}{2}$ and $x^{-1}\sin x < 1$, $x \neq 0$.]

19.7. Let μ be a finite measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and denote by $\phi(\xi) := \widehat{\mu}(\xi)$ the Fourier transform

(i) ϕ is positive semidefinite, i.e. for all $m \in \mathbb{N}$, $\xi_1, \ldots, \xi_m \in \mathbb{R}^n$, $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$ we have $\sum_{k,l=1}^m \phi(\xi_k - \xi_l)\lambda_k\bar{\lambda}_l \geqslant 0$. (This means that the $m \times m$ matrices $\big(\phi(\xi_k - \xi_l)\big)_{kl}$ are positive hermitian.)
Remark. The converse assertion also holds, see Problem 21.4.

(ii) Let $m \in \mathbb{N}$. Show that $\int |x|^m\,\mu(dx) < \infty \implies \phi \in C^m(\mathbb{R}^n)$.

(iii) Let $N \in \mathbb{N}$. Show that $\phi \in C^{2N}(\mathbb{R}^n) \implies \int |x|^{2N}\,\mu(dx) < \infty$.
[Hint: It suffices to consider $n = 1$ and $N = 1$. Observe that we have $\phi''(0) = \lim_{h\to 0}(\phi(2h) - 2\phi(0) + \phi(-2h))/(4h^2)$, express this identity using Fourier transforms and use Fatou's lemma.]

(iv) The support of a measure μ is the smallest closed set $K \subset \mathbb{R}^n$ such that we have $\mu(U) = 0$ for all open sets $U \subset K^c$. If $\operatorname{supp}\mu$ is compact, then $z \mapsto \phi(z)$ is well-defined for $z \in \mathbb{C}^n$ and analytic.

19.8. Let $B \in \mathscr{B}(\mathbb{R})$. If $\int_B e^{ix/n}\,dx = 0$ for all $n = 1, 2, \ldots$, then $\lambda^1(B) = 0$.

19.9. Assume that μ is a finite measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Show that

(i) $\exists \xi \neq 0 : \widehat{\mu}(\xi) = \widehat{\mu}(0) \iff \exists \xi \neq 0 : \mu\big(\mathbb{R} \setminus (2\pi/\xi)\mathbb{Z}\big) = 0$;

(ii) $\exists \xi_1, \xi_2,\ \xi_1/\xi_2 \notin \mathbb{Q} : |\widehat{\mu}(\xi_1)| = |\widehat{\mu}(\xi_2)| = \widehat{\mu}(0) \implies |\widehat{\mu}| \equiv \widehat{\mu}(0)$.

20

The Radon–Nikodým Theorem

Let $(X, \mathscr{A}, \mu)$ be a measure space. We saw in Lemma 10.8 that for any positive function $f \in \mathcal{L}^1(\mathscr{A})$ – or indeed for any positive $f \in \mathcal{M}(\mathscr{A})$ – the set function $\nu := f\mu$ given by $\nu(A) := \int_A f(x)\mu(dx)$ is again a measure. If $\mu = \lambda^1$ is one-dimensional Lebesgue measure, $A = [a, x)$ and f is a positive continuous function, then we have $f(x) = d\nu(x)/dx$. This is the blueprint for the following generalization. From Theorem 11.2(ii) we know that

$$N \in \mathscr{A},\ \mu(N) = 0 \implies \nu(N) = 0. \tag{20.1}$$

This condition means that all μ-null sets are ν-null sets: $\mathscr{N}_\mu \subset \mathscr{N}_\nu$.

Definition 20.1 Let μ, ν be two measures on the measurable space $(X, \mathscr{A})$. If (20.1) holds, we call ν *absolutely continuous* w.r.t. μ and write $\nu \ll \mu$.

Measures with densities are always absolutely continuous w.r.t. their base measure: $f\mu \ll \mu$. Remarkably, the converse is also true.

Theorem 20.2 (Radon–Nikodým) *Let μ, ν be two measures on the measurable space $(X, \mathscr{A})$. If μ is σ-finite, then the following assertions are equivalent:*

(i) $\nu(A) = \displaystyle\int_A f(x)\mu(dx)$ *for some a.e. unique $f \in \mathcal{M}(\mathscr{A})$, $f \geqslant 0$;*

(ii) $\nu \ll \mu$.

The unique function f is called the Radon–Nikodým derivative and is traditionally denoted by $f = d\nu/d\mu$.

Proof[1] We have already seen the direction (i) $\Rightarrow$ (ii). For the converse implication (ii) $\Rightarrow$ (i) we construct an $\mathscr{A}/\mathscr{B}(\mathbb{R})$ measurable function $f^* : X \to [0, \infty]$

[1] The first part of this proof (Steps 1–5) is inspired by the presentation of Sellke [44].

such that

$$\int u\,d\nu = \int uf^*\,d\mu \quad \forall u \in \mathcal{M}(\mathscr{A}),\ \ u \geqslant 0. \tag{20.2}$$

If we take $u = \mathbb{1}_A$ for $A \in \mathscr{A}$, we get (i). We divide the proof into several steps.

Assume that μ and ν are finite measures.

Step 1. Set $\rho := \mu + \nu$ and define a functional

$$\Phi(u) := \int u^2\,d\mu + \int (1-u)^2\,d\nu \quad \forall u \in L^2(\rho) = L^2(\mu) \cap L^2(\nu).$$

Observe that

$$d^2 := \inf_{u \in L^2(\rho)} \Phi(u) \leqslant \Phi(0) = \int 1\,d\nu = \nu(X) < \infty,$$

and

$$\begin{aligned}
&\Phi(u+w) + \Phi(u-w)\\
&= \int \left[(u+w)^2 + (u-w)^2\right] d\mu + \int \left[((1-u)+w)^2 + ((1-u)-w)^2\right] d\nu\\
&= 2\int u^2\,d\mu + 2\int w^2\,d\mu + 2\int (1-u)^2\,d\nu + 2\int w^2\,d\nu\\
&= 2\Phi(u) + 2\|w\|^2_{L^2(\rho)}.[2]
\end{aligned} \tag{20.3}$$

Step 2. There is a minimizer $f \in L^2(\rho) : d^2 = \Phi(f)$. By the very definition of the infimum, there is a sequence $(f_n)_{n\in\mathbb{N}} \subset L^2(\rho)$ such that

$$d^2 = \lim_{n\to\infty} \Phi(f_n).$$

If we can show that $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, we know because of completeness (Theorem 13.7) that the limit $f = \lim_{n\to\infty} f_n$ exists and defines an element in $L^2(\rho)$. Using (20.3) with $u = \frac{1}{2}(f_n + f_m)$ and $w = \frac{1}{2}(f_n - f_m)$, we find

$$\begin{aligned}
d^2 + \left\|\frac{f_n - f_m}{2}\right\|^2_{L^2(\rho)} &\overset{\inf}{\leqslant} \Phi\left(\frac{f_n + f_m}{2}\right) + \left\|\frac{f_n - f_m}{2}\right\|^2_{L^2(\rho)}\\
&\overset{(20.3)}{=} \frac{1}{2}\Phi(f_n) + \frac{1}{2}\Phi(f_m) \xrightarrow[n,m\to\infty]{} d^2.
\end{aligned}$$

Thus, $\lim_{n,m\to\infty} \|f_n - f_m\|_{L^2(\rho)} = 0$, i.e. $(f_n)_n$ is a Cauchy sequence in $L^2(\rho)$.

[2] By $\|u\|_{L^2(\rho)}$ we denote the norm in $L^2(\rho)$, i.e. $\left(\int |u|^2\,d\rho\right)^{1/2}$.

Using (20.3) for the functions $u = f_n$ and $w = f - f_n$ we get

$$\begin{aligned}\Phi(f) + d^2 &\leqslant \Phi(f) + \Phi(2f_n - f) \\ &\overset{(20.3)}{=} 2\Phi(f_n) + 2\|f - f_n\|^2_{L^2(\rho)} \xrightarrow[n\to\infty]{} 2d^2 + 0,\end{aligned}$$

which means that $\Phi(f) = d^2$, i.e. f is indeed a minimizer of $\Phi(\cdot)$.

Step 3. The minimizer f is μ-a.e. and ν-a.e. unique. Assume that g is another minimizer. Using $u = \frac{1}{2}(f + g)$ and $w = \frac{1}{2}(f - g)$ in (20.3) we get

$$\begin{aligned}d^2 + \left\|\frac{f-g}{2}\right\|^2_{L^2(\rho)} &\leqslant \Phi\left(\frac{f+g}{2}\right) + \left\|\frac{f-g}{2}\right\|^2_{L^2(\rho)} \\ &\overset{(20.3)}{=} \frac{1}{2}\Phi(f) + \frac{1}{2}\Phi(g) = d^2.\end{aligned}$$

This shows that $\|f - g\|_{L^2(\rho)} = 0$, hence $f = g$ $(\mu + \nu)$-a.e.

Step 4. The minimizer takes values in $[0, 1]$. From

$$(0 \vee f \wedge 1)^2 = 0 \vee f^2 \wedge 1 \leqslant f^2,$$
$$(0 \vee (1 - f) \wedge 1)^2 = 0 \vee (1 - f)^2 \wedge 1 \leqslant (1 - f)^2$$

we conclude that $d^2 \leqslant \Phi(0 \vee f \wedge 1) \leqslant \Phi(f) = d^2$. Since the minimizer is unique, we get $0 \leqslant f \leqslant 1$.

Step 5. Construction of the density. Let $u \in \mathcal{M}(\mathscr{A})$, $u \geqslant 0$ and set $u_n := u \wedge n$. For all $t \in \mathbb{R}$ we find

$$\begin{aligned}0 &\leqslant \Phi(f + tu_n) - \Phi(f) \\ &= \int \left[(f + tu_n)^2 - f^2\right] d\mu + \int \left[(1 - f - tu_n)^2 - (1 - f)^2\right] d\nu \\ &= t^2 \int u_n^2 \, d\mu + t^2 \int u_n^2 \, d\nu + 2t \int fu_n \, d\mu - 2t \int (1 - f)u_n \, d\nu.\end{aligned}$$

Dividing by t^2 we conclude from

$$0 \leqslant \int u_n^2 \, d\mu + \int u_n^2 \, d\nu + \frac{2}{t}\left(\int fu_n \, d\mu - \int (1 - f)u_n \, d\nu\right) \quad \forall t \in \mathbb{R} \setminus \{0\}$$

that the expression in braces must vanish. Using Beppo Levi's theorem, we see that for any $u \in \mathcal{M}(\mathscr{A})$, $u \geqslant 0$,

$$\int fu \, d\mu = \sup_n \int fu_n \, d\mu = \sup_n \int (1 - f)u_n \, d\nu = \int (1 - f)u \, d\nu.$$

Now pick $u = \mathbb{1}_N$, where $N = \{f = 1\} \in \mathscr{A}$. This shows that

$$\mu(N) = \int_N 1\,d\mu = \int_N 0\,d\nu = 0$$

and so $\nu(N) = 0$ since $\nu \ll \mu$. Set $f^* := f/(1-f)\mathbb{1}_{\{f\neq 1\}}$, which is clearly a positive measurable function. By construction,

$$\begin{aligned}\int uf^*\,d\mu = \int f\frac{u}{1-f}\mathbb{1}_{N^c}\,d\mu &= \int (1-f)\frac{u}{1-f}\mathbb{1}_{N^c}\,d\nu \\ &= \int u\mathbb{1}_{N^c}\,d\nu \\ &\overset{\nu(N)=0}{=} \int u\,d\nu.\end{aligned}$$

From now on, the proof follows standard arguments.

$\boxed{\textbf{Assume that } \mu \textbf{ is a finite and } \nu \textbf{ an arbitrary measure.}}$

Step 6. Denote by $\mathscr{F} := \{F \in \mathscr{A} : \nu(F) < \infty\}$ the sets with finite ν-measure. Obviously, $\mathscr{F}$ is $\cup$-stable, and the constant

$$c := \sup_{F\in\mathscr{F}} \mu(F) \leqslant \mu(X) < \infty$$

can be approximated by an increasing sequence defined by $(F_n)_{n\in\mathbb{N}} \subset \mathscr{F}$ such that $F_\infty := \bigcup_{n\in\mathbb{N}} F_n$ and $c = \mu(F_\infty) = \sup_{n\in\mathbb{N}} \mu(F_n)$. [✍] The restriction of ν to F_∞ is, by definition, σ-finite, while for $A \subset F_\infty^c$, $A \in \mathscr{A}$, we have

$$\text{either} \quad \mu(A) = \nu(A) = 0 \quad \text{or} \quad 0 < \mu(A) < \nu(A) = \infty. \tag{20.4}$$

In fact, if $\nu(A) < \infty$, then $F_n \cup A \in \mathscr{F}$ for all $n \in \mathbb{N}$, which implies that

$$c \geqslant \sup_{n\in\mathbb{N}} \mu(F_n \uplus A) = \sup_{n\in\mathbb{N}}(\mu(F_n) + \mu(A)) = \mu(F_\infty) + \mu(A) = c + \mu(A),$$

that is, $\mu(A) = 0$, hence $\nu(A) = 0$ by absolute continuity; if, however, $\nu(A) = \infty$ we have again by absolute continuity that $\mu(A) > 0$. Define now

$$\nu_n := \nu\big(\cdot \cap (F_n \setminus F_{n-1})\big), \quad \mu_n := \mu\big(\cdot \cap (F_n \setminus F_{n-1})\big) \qquad (F_0 := \emptyset)$$

and it is clear that $\nu_n \ll \mu_n$ for every $n \in \mathbb{N}$. Since μ_n, ν_n are finite measures, the first part of this proof shows that $\nu_n = f_n\mu_n$. Obviously, the function

$$f(x) := \begin{cases} f_n(x) & \text{if } x \in F_n \setminus F_{n-1}, \\ \infty & \text{if } x \in F_\infty^c, \end{cases} \tag{20.5}$$

fulfils $\nu = f\mu$. By construction, f is unique on the set F_∞. But since *every* density $\tilde{f}$ of ν with respect to μ satisfies

$$\nu\big(\{\tilde{f} \leqslant r\} \cap F_\infty^c\big) = \int_{\{\tilde{f} \leqslant r\} \cap F_\infty^c} \tilde{f}\, d\mu \leqslant n\mu\big(\{\tilde{f} \leqslant r\} \cap F_\infty^c\big) < \infty,$$

the alternative (20.4) reveals that $\nu\big(\{\tilde{f} \leqslant r\} \cap F_\infty^c\big) = \mu\big(\{\tilde{f} \leqslant r\} \cap F_\infty^c\big) = 0$ for all $r \in \mathbb{N}$, i.e. that $\tilde{f}|_{F_\infty^c} = \infty$. In other words: f, as defined in (20.5), is also unique on F_∞^c.

Assume that μ is a σ-finite and ν an arbitrary measure.

Step 7. Since μ is σ-finite, there is a sequence $(A_n)_{n \in \mathbb{N}} \subset \mathscr{A}$ such that $A_n \uparrow X$ and $\mu(A_n) < \infty$. Setting

$$h(x) := \sum_{n=1}^{\infty} \frac{2^{-n}}{1 + \mu(A_n)} \mathbb{1}_{A_n}(x) > 0 \quad \forall x$$

we see that the measures $h\mu$ and μ have the same null sets. [✍] Therefore,

$$\nu \ll \mu \iff \nu \ll h\mu.$$

By Beppo Levi's theorem (Theorem 9.6)

$$\int_X h\, d\mu = \sum_{n=1}^{\infty} \frac{2^{-n}}{1 + \mu(A_n)} \int_X \mathbb{1}_{A_n}\, d\mu = \sum_{n=1}^{\infty} \frac{\mu(A_n)}{1 + \mu(A_n)} 2^{-n} \leqslant \sum_{n=1}^{\infty} 2^{-n} = 1,$$

i.e. $h\mu$ is a finite measure and the first two parts of the proof show that there is some $f \in \mathcal{M}^+(\mathscr{A})$ such that

$$\nu = f \cdot (h\mu) \overset{(*)}{=} (fh)\mu.$$

Let us verify the equality marked $(*)$: for any simple function $f = \sum_{n=0}^{N} y_n \mathbb{1}_{B_n} \geqslant 0$ we see that

$$\nu(A) = \int_A \sum_{n=0}^{N} y_n \mathbb{1}_{B_n}\, d(h\mu) = \sum_{n=0}^{N} y_n \int \mathbb{1}_{B_n \cap A} h\, d\mu = \int_A (fh) d\mu,$$

and the general case follows using the sombrero lemma (Theorem 8.8) and Beppo Levi's theorem.

Uniqueness is clear as f is $h\mu$-a.e. unique, which implies that fh is μ-a.e. unique since $h > 0$. $\square$

The Radon–Nikodým theorem can be strengthened in the sense that every σ-finite measure has a unique decomposition into an absolutely continuous and a singular part. Singularity is, in some sense, the opposite of absolute continuity.

Definition 20.3 Two measures μ, ν on a measurable space $(X, \mathscr{A})$ are called (mutually) *singular* if there is a set $N \in \mathscr{A}$ such that $\nu(N) = 0 = \mu(N^c)$. We write in this case $\mu \perp \nu$ (or $\nu \perp \mu$ as '$\perp$' is symmetric).

On $\mathbb{R}^n$ with the Borel σ-algebra $\mathscr{B}(\mathbb{R}^n)$ the typical examples of singular measures are $\delta_x \perp \lambda^n$ (for any $x \in \mathbb{R}^n$) and $f\mu \perp g\mu$, where f, g are positive measurable functions with disjoint support: $\operatorname{supp} f \cap \operatorname{supp} g = \emptyset$.[3]

The measures μ and ν are singular, if they have disjoint 'supports', that is, if μ lives in a region of X which is not charged by ν and vice versa. In general, however, two measures are neither purely absolutely continuous nor purely singular, but are a mixture of both. The following proof is taken from Titkos [56]. Although we present it only for finite measures, it can be extended to a σ-finite setting. The classical proof of Lebesgue's decomposition theorem is sketched in Problem 20.5 and worked out in Theorem 25.9.

Theorem 20.4 (Lebesgue decomposition) *Let μ, ν be finite measures on a measurable space $(X, \mathscr{A})$. Then there exists a unique (up to null sets) decomposition $\nu = \nu^\circ + \nu^\perp$, where $\nu^\circ \ll \mu$ and $\nu^\perp \perp \mu$.*

Proof Let $\mathcal{N}_\mu := \operatorname{span}\{\mathbb{1}_N : N \in \mathscr{N}_\mu\}$ be the linear space generated by the indicator functions of μ-null sets. Define for all $A \in \mathscr{A}$ the set function

$$\nu^\circ(A) := \inf_{u \in \mathcal{N}_\mu} \int_X |\mathbb{1}_A - u| \, d\nu = \inf_{\substack{u \in \mathcal{N}_\mu \\ \{u \neq 0\} \subset A}} \int_X |\mathbb{1}_A - u| \, d\nu.$$

Obviously,

- ν° is a positive set-function,
- $\nu^\circ(A) \leqslant \nu(A)$ for all $A \in \mathscr{A}$,
- $\mu(A) = 0 \implies \nu^\circ(A) = 0$, i.e. $\nu^\circ \ll \mu$.

If $A_1, A_2 \in \mathscr{A}$ are disjoint, then

$$\{u \in \mathcal{N}_\mu : \{u \neq 0\} \subset A_1 \cup A_2\} = \{u_1 + u_2 \in \mathcal{N}_\mu : \{u_i \neq 0\} \subset A_i, \ i = 1, 2\},$$

and this shows that ν° is finitely additive. Moreover, for any disjoint sequence $(A_n)_{n \in \mathbb{N}} \subset \mathscr{A}$

$$\nu^\circ\left(\bigcupdot_{i=1}^\infty A_i\right) - \sum_{i=1}^n \nu^\circ(A_i) \overset{\text{finitely}}{\underset{\text{additive}}{=}} \nu^\circ\left(\bigcupdot_{i=n+1}^\infty A_i\right) \leqslant \nu\left(\bigcupdot_{i=n+1}^\infty A_i\right) \xrightarrow[n \to \infty]{} 0$$

because of the σ-additivity of the measure ν.

[3] Recall that $\operatorname{supp} f := \overline{\{f \neq 0\}}$.

Set $\nu^{\perp} := \nu - \nu^{\circ}$. We still have to show that $\nu^{\perp} \perp \mu$ and that the decomposition is unique. Let us show that ν° is maximal among all measures ρ such that $\rho \leqslant \nu$ and $\rho \ll \mu$: assume that ρ is such a measure, $u \in \mathcal{N}_{\mu}$ and note that

$$\rho(A) = \int |\mathbb{1}_A| d\rho \leqslant \int |\mathbb{1}_A - u| d\rho + \underbrace{\int |u| d\rho}_{=0,\, \rho \ll \mu} \overset{\rho \leqslant \nu}{\leqslant} \int |\mathbb{1}_A - u|^2 \, d\nu.$$

Taking the infimum over $u \in \mathcal{N}_{\mu}$ yields $\rho \leqslant \nu^{\circ}$.

Orthogonality. Let τ be a measure such that $\tau \leqslant \mu$ and $\tau \leqslant \nu^{\perp}$. Clearly, this implies $\nu^{\circ} + \tau \leqslant \nu$ and $\nu^{\circ} + \tau \ll \mu$. By the maximality, $\nu^{\circ} + \tau \ll \nu^{\circ}$ and we conclude that $\tau = 0$ and $\nu^{\perp} \perp \mu$.

Uniqueness. If $\nu = \nu_1 + \nu_2$ is any other decomposition such that $\nu_1 \ll \mu$ and $\nu_2 \perp \mu$, then $\nu^{\circ} - \nu_1$ is, with respect to μ, both absolutely continuous and singular. This is possible only if $\nu_1 = \nu^{\circ}$, proving uniqueness. □

Problems

20.1. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and let ν be a further measure. Show that $\nu \leqslant \mu$ entails that $\nu = f\mu$ for some (a.e. uniquely determined) density function f such that $0 \leqslant f \leqslant 1$.

20.2. Let μ, ν be two σ-finite measures on $(X, \mathscr{A})$ which have the same null sets. Show that $\nu = f\mu$ and $\mu = g\nu$, where $0 < f < \infty$ a.e. and $g = 1/f$ a.e.
Remark. Measures having the same null sets are called *equivalent*.

20.3. Give an example of a measure μ and a density f such that $f\mu$ is not σ-finite.

20.4. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and assume that $\nu = f\mu$ for a positive measurable function f.

(i) Show that ν is a finite measure if, and only if, $f \in \mathcal{L}^1(\mu)$.
(ii) Show that ν is a σ-finite measure if, and only if, $\mu\{f = \infty\} = 0$.

20.5. Two measures ρ, σ defined on the same measurable space $(X, \mathscr{A})$ are *singular*, if there is a set $N \in \mathscr{A}$ such that $\rho(N) = \sigma(N^c) = 0$. If this is the case, we write $\rho \perp \sigma$. Steps (i)–(iv) below show the so-called **Lebesgue decomposition theorem**: *if μ, ν are σ-finite measures, then there is a decomposition (unique up to null sets) $\nu = \nu^{\circ} + \nu^{\perp}$, where $\nu^{\circ} \ll \mu$ and $\nu^{\perp} \perp \mu$.*

(i) $\nu \ll \nu + \mu$ and there is a density $f = d\nu/d(\nu + \mu)$.
(ii) The density satisfies $0 \leqslant f \leqslant 1$ and $(1 - f)\nu = f\mu$.
(iii) $\nu^{\circ}(A) := \nu(A \cap \{f < 1\})$ and $\nu^{\perp}(A) = \nu(A \cap \{f = 1\})$.
(iv) Uniqueness follows from $d\nu^{\circ}/d\mu = f/(1 - f)\mathbb{1}_{\{f<1\}}$.

20.6. **Bounded variation and absolute continuity.** Let λ be one-dimensional Lebesgue measure. A function $F: [a, b] \to \mathbb{R}$ is said to be

absolutely continuous (AC) if for each $\epsilon > 0$ there is a $\delta > 0$ such that $\sum_{n=1}^{N} |F(y_n) - F(x_n)| \leqslant \epsilon$ for all $a \leqslant x_1 < y_1 < x_2 < y_2 < \cdots < x_N < y_N \leqslant b$ with $\sum_{n=1}^{N} |y_n - x_n| < \delta$; of *bounded variation* (BV) if $V(F, [a, b]) :=$

$\sup\sum_{n=1}^{N}|F(t_n)-F(t_{n-1})|$ is finite, where the sup extends over all finite partitions $t_0=a<t_1<\cdots<t_N=b$, $N\in\mathbb{N}$. $V(F,[a,b])$ is called the *variation* of F.

(i) We have $\mathrm{AC}[a,b]\subset C[a,b]\cap\mathrm{BV}[a,b]$.

(ii) Any $F\in\mathrm{BV}[a,b]$ can be written as a sum of increasing functions f and g. [Hint: try $f(t)=V(f,[a,t])$.]

(iii) $F(x):=\int_{[a,x)}f(t)\lambda(dt)$ is AC for all $f\in\mathcal{L}^1(\mu)$.

(iv) If F is AC, there is some $\phi\in\mathcal{L}^1(\lambda)$ such that $F(x)=\int_{(-\infty,x)}\phi(t)\lambda(dt)$. [Hint: decompose $F=f_1-f_2$ with f_i increasing and AC.]

20.7. Let μ be Lebesgue measure on $[0,2]$ and ν be Lebesgue measure on $[1,3]$. Find the Lebesgue decomposition of ν with respect to μ.

20.8. Stieltjes measure (3). Let $(\mathbb{R},\mathscr{B}(\mathbb{R}),\mu)$ be a finite measure space and denote by F the left-continuous distribution function of μ as in Problem 6.1. Use Lebesgue's decomposition theorem (Theorem 20.4) to show that we can decompose $F=F_1+F_2+F_3$ and, accordingly, $\mu=\mu_1+\mu_2+\mu_3$ in such a way that

(a) F_1 is discrete, i.e. μ_1 is the countable sum of weighted Dirac δ-measures;

(b) F_2 is absolutely continuous (see Problem 20.6);

(c) F_3 is continuous and singular, i.e. $\mu_3\perp\lambda^1$.

20.9. The devil's staircase. Recall the construction of Cantor's ternary set from Problem 7.12. Denote by $I_n^1,\dots,I_n^{2^n-1}$ the intervals which make up $[0,1]\setminus C_n$ arranged in increasing order of their endpoints. We construct a sequence of functions $F_n:[0,1]\to[0,1]$ by

$$F_n(x):=\begin{cases}0, & \text{if } x=0,\\ i2^{-n} & \text{if } x\in I_n^i,\ 1\leqslant i\leqslant 2^n-1,\\ 1, & \text{if } x=1,\end{cases}$$

and interpolate linearly between these values to get $F_n(x)$ for all other x.

(i) Sketch the first three functions F_1,F_2,F_3.

(ii) Show that the limit $F(x):=\lim_{n\to\infty}F_n(x)$ exists.
Remark. F is usually called the *Cantor function*.

(iii) Show that F is continuous and increasing.

(iv) Show that F' exists a.e. and equals 0.

(v) Show that F is not absolutely continuous (in the sense of Problem 20.8(2)) but singular, i.e. the corresponding measure μ with distribution function F is singular w.r.t. Lebesgue measure $\lambda^1|_{[0,1]}$.

21
Riesz Representation Theorems

In this chapter we want to study the structure of linear functionals on the Lebesgue spaces $L^p(\mu)$ over a σ-finite measure space $(X, \mathscr{A}, \mu)$ and on the spaces of continuous functions defined on a (locally compact) metric space (X, d), see Appendix B. If X is a metric space we always use the Borel σ-algebra $\mathscr{A} = \mathscr{B}(X)$.

Bounded and Positive Linear Functionals

Let X be as above and assume that μ is a measure on $(X, \mathscr{A})$. We denote by $(\mathcal{L}, \|\cdot\|)$ any of the following function spaces:

- the space of pth-power integrable functions $(L^p(\mu), \|\cdot\|_p)$, $1 \leqslant p < \infty$;
- the space of continuous functions with compact support $(C_c(X), \|\cdot\|_\infty)$;
- the space of continuous functions vanishing at infinity $(C_\infty(X), \|\cdot\|_\infty)$, $C_\infty(X) := \overline{C_c(X)}^{\|\cdot\|_\infty} = \{u \in C(X) : \lim_{|x|\to\infty} u(x) = 0\}$;
- the space of bounded continuous functions $(C_b(X), \|\cdot\|_\infty)$.

Definition 21.1 A linear map $I : \mathcal{L} \to \mathbb{R}$ is said to be a

bounded linear functional if there is a $c < \infty$ such that $|I(u)| \leqslant c\|u\|$ for all $u \in \mathcal{L}$;
positive linear functional if $I(u) \geqslant 0$ for all $u \in \mathcal{L}$ with $u \geqslant 0$.

Integrals are typical examples of bounded and positive linear functionals.

Example 21.2 $I(u) := \int u\,d\mu$ is a positive linear functional

(i) on $L^1(\mu)$, see Theorem 10.4;
(ii) on $C_\infty(X)$ or $C_b(X)$, if μ is a finite measure;
(iii) on $C_c(X)$, if μ is a measure such that $\mu(K) < \infty$ for all compact sets $K \subset X$.

Using the Hölder inequality we see that $I_f(u) := \int uf\,d\mu$ is a bounded linear functional

(iv) on $L^p(\mu)$ provided that $f \in L^q(\mu)$; moreover, $c = \|f\|_q$.

If I is a linear functional on $\mathcal{L}$ we write

$$\|I\| := \sup_{u \in \mathcal{L}, u \neq 0} \frac{|I(u)|}{\|u\|} \in [0, \infty]$$

for the norm of I. For a bounded linear functional, $\|I\| < \infty$ is the smallest constant c such that the inequality in Definition 21.1 holds. Because of linearity, boundedness implies continuity. This follows immediately from

$$|I(u) - I(v)| = |I(u - v)| \leqslant c\|u - v\| \quad \forall u, v \in \mathcal{L}.$$

We will see in the next sections that *all* bounded and positive linear functionals on $\mathcal{L}$ are integrals. Let us show some properties of positive linear functionals.

Lemma 21.3 *Let $(\mathcal{L}, \|\cdot\|)$ be one of the spaces $(L^p(\mu), \|\cdot\|_{L^p(\mu)})$, $1 \leqslant p < \infty$, $(C_\infty(X), \|\cdot\|_\infty)$ or $(C_b(X), \|\cdot\|_\infty)$ and $I : \mathcal{L} \to \mathbb{R}$ be a positive linear functional.*

(i) *I is monotone, i.e. $I(u) \leqslant I(v)$ for all $u, v \in \mathcal{L}$ such that $u \leqslant v$.*
(ii) *I satisfies the triangle inequality, i.e. $|I(u)| \leqslant I(|u|)$ for all $u \in \mathcal{L}$.*
(iii) *I is bounded, i.e. $|I(u)| \leqslant c\|u\|$ for all $u \in \mathcal{L}$.*

Proof (i) If $u \leqslant v$, then $v - u \geqslant 0$ and, because of linearity and positivity,

$$I(v) - I(u) = I(v - u) \geqslant 0.$$

(ii) follows from (i) if we take $\pm u \leqslant |u| = v$.
(iii) Assume that I is not bounded. This means that there exists a sequence $(u_n)_{n \in \mathbb{N}} \subset \mathcal{L}$ such that

$$\|u_n\| \leqslant 1 \quad \text{and} \quad |I(u_n)| \geqslant 4^n.$$

Since $\||u_n|\| = \|u_n\|$ and $|I(u_n)| \leqslant I(|u_n|)$, we may assume that $u_n \geqslant 0$. Define $u := \sum_{n=1}^\infty 2^{-n} u_n$ and observe that

$$\left\| \sum_{n=1}^N 2^{-n} u_n - \sum_{n=1}^M 2^{-n} u_n \right\| = \left\| \sum_{n=M+1}^N 2^{-n} u_n \right\| \leqslant \sum_{n=M+1}^\infty \|2^{-n} u_n\| \leqslant \sum_{n=M+1}^\infty 2^{-n}$$

tends to zero as $N > M \to \infty$. Since $\mathcal{L}$ is complete, we see that $u \in \mathcal{L}$; moreover,

$$I(u) \geqslant I(2^{-n} u_n) = 2^{-n} I(u_n) \geqslant 2^{-n} 4^n = 2^n \quad \forall n \in \mathbb{N}.$$

This implies that $I(u) = \infty$, contradicting the assumption $I : \mathcal{L} \to \mathbb{R}$. $\square$

Positive linear functionals can be used as building blocks for all bounded linear functionals.

Theorem 21.4 *Let* $(\mathcal{L}, \|\cdot\|)$ *be one of the spaces* $(L^p(\mu), \|\cdot\|_{L^p(\mu)})$, $1 \leqslant p < \infty$, $(C_c(X), \|\cdot\|_\infty)$, $(C_\infty(X), \|\cdot\|_\infty)$ *or* $(C_b(X), \|\cdot\|_\infty)$ *and* $I : \mathcal{L} \to \mathbb{R}$ *be a bounded linear functional. There exist positive linear functionals* $I^\pm : \mathcal{L} \to \mathbb{R}$ *such that* $I = I^+ - I^-$.

Proof Throughout the proof all functions f, f_i, u, u_i are elements of $\mathcal{L}$. For $u \geqslant 0$ we define $I^+(u) := \sup\{I(f) : 0 \leqslant f \leqslant u\}$ and $I^-(u) := I^+(u) - I(u)$.

Step 1. Well-definedness of I^+. In the spaces $\mathcal{L}$ we see that $0 \leqslant f \leqslant u$ implies that $\|f\| \leqslant \|u\|$, and so

$$I(f) \leqslant |I(f)| \leqslant \|I\| \|f\| \leqslant \|I\| \|u\|.$$

If we take the supremum over all $0 \leqslant f \leqslant u$, we see that $I^+(u) \leqslant \|I\| \|u\| < \infty$.

Step 2. I^+ *is additive and positive homogeneous.* $I^+(\gamma u) = \gamma I^+(u)$ follows for $\gamma \geqslant 0$ from the very definition of I^+. Assume, for a moment, that we know that

$$0 \leqslant f \leqslant u_1 + u_2,\ u_1, u_2 \geqslant 0 \iff \exists f_i : 0 \leqslant f_i \leqslant u_i\ (i = 1, 2). \tag{21.1}$$

Let $u_1, u_2 \geqslant 0$. From the definition of I^+ we infer $I^+(u_1 + u_2) \geqslant I(f_1) + I(f_2)$ for $0 \leqslant f_i \leqslant u_i$, and so

$$I^+(u_1 + u_2) \geqslant \sup_{0 \leqslant f_1 \leqslant u_1} I(f_1) + \sup_{0 \leqslant f_2 \leqslant u_2} I(f_2) = I^+(u_1) + I^+(u_2).$$

Conversely, we get

$$\begin{aligned} I^+(u_1 + u_2) &= \sup\{I(f) : 0 \leqslant f \leqslant u_1 + u_2\} \\ &\overset{(21.1)}{=} \sup\{I(f_1 + f_2) : 0 \leqslant f_1 \leqslant u_1, 0 \leqslant f_2 \leqslant u_2\} \\ &\leqslant \sup\{I(f_1) : 0 \leqslant f_1 \leqslant u_1\} + \sup\{I(f_2) : 0 \leqslant f_2 \leqslant u_2\} \\ &= I^+(u_1) + I^+(u_2). \end{aligned}$$

Step 3. Extension of I^+ *to a linear functional.* We extend I^+ by linearity, setting $I^+(u) = I^+(u^+) - I^+(u^-)$. Well-definedness is proved as in Remark 10.2.

Step 4. Proof of (21.1). The direction '$\Leftarrow$' is obvious. In order to see '$\Rightarrow$' we use the elementary inequality

$$(a + b) \wedge c \leqslant (a \wedge c) + (b \wedge c) \quad \forall a, b, c \in [0, \infty). \tag{21.2}$$

Assume that $0 \leqslant f \leqslant u_1 + u_2$ and $u_1, u_2 \geqslant 0$. Define $f_1 := f \wedge u_1$ and $f_2 := f - f_1$ and note that $f_1, f_2 \in \mathcal{L}$ [✍]. Clearly, $f_1 + f_2 = f$, $0 \leqslant f_1 \leqslant u_1$ and, using (21.2),

$$0 \leqslant f_2 = \underbrace{(u_1 + u_2) \wedge f}_{=f} - u_1 \wedge f \leqslant (u_1 \wedge f) + (u_2 \wedge f) - (u_1 \wedge f) \\ = u_2 \wedge f \leqslant u_2. \qquad \square$$

Duality of the Spaces $L^p(\mu)$, $1 \leqslant p < \infty$

Throughout this section $(X, \mathscr{A}, \mu)$ is a σ-finite measure space, i.e. there is a sequence $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ such that $A_n \uparrow X$ and $\mu(A_n) < \infty$. Moreover, we assume that $p \in [1, \infty)$ and $q \in (1, \infty]$ are conjugate indices, i.e. $p^{-1} + q^{-1} = 1$; as usual, $q = \infty$ if $p = 1$. From Hölder's inequality we see that for every $f \in L^q(\mu)$, $f \geqslant 0$,

$$I_f(u) := \int uf \, d\mu \quad \forall u \in L^p(\mu) \tag{21.3}$$

defines a positive linear functional on $L^p(\mu)$ and $|I_f(u)| \leqslant \|f\|_q \|u\|_p$.

Theorem 21.5 (Riesz) *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space, $p \in [1, \infty)$ and $q \in (1, \infty]$ conjugate indices. Every positive linear functional $I : L^p(\mu) \to \mathbb{R}$ is of the form* (21.3)*, i.e. there is some unique $f \in L^q(\mu)$, $f \geqslant 0$, such that $I = I_f$.*

Proof *Step 1. Uniqueness.* If $I_f \equiv I_g$, then $I_{f-g} \equiv 0$. Define $h := f - g$ and observe that $(|h|^{q-1})^p = |h|^{(q-1)p} = |h|^q \in L^1(\mu)$. Thus, $\operatorname{sgn}(h)|h|^{q-1} \in L^p(\mu)$, and we get

$$0 = I_h\Big(\operatorname{sgn}(h)|h|^{q-1}\Big) = \int |h|^q \, d\mu \implies h = 0 \ \mu\text{-a.e.} \implies f = g \ \mu\text{-a.e.}$$

Step 2. Existence if $\mu(X) < \infty$. Since $\mathbb{1}_A \in L^p(\mu)$,

$$\nu : \mathscr{A} \to [0, \infty), \quad \nu(A) := I(\mathbb{1}_A)$$

is a well-defined set function. Since I is a positive linear functional, we have $\nu(\emptyset) = 0$. If $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ are disjoint, the dominated convergence theorem (Theorem 13.9) shows that $\sum_{n=1}^N \mathbb{1}_{A_n} \to \sum_{n=1}^\infty \mathbb{1}_{A_n}$ in $L^p(\mu)$ as $N \to \infty$. Because of the continuity of I (Lemma 21.3) we get

$$\nu\left[\dot{\bigcup_{n=1}^\infty} A_n\right] = I\left[\sum_{n=1}^\infty \mathbb{1}_{A_n}\right] = \lim_{N\to\infty} I\left[\sum_{n=1}^N \mathbb{1}_{A_n}\right] = \lim_{N\to\infty} \sum_{n=1}^N I[\mathbb{1}_{A_n}] = \sum_{n=1}^\infty \nu(A_n),$$

which shows that ν is a measure. Moreover, we have for all $A \in \mathscr{A}$

$$\mu(A) = 0 \implies \mathbb{1}_A = 0 \ \mu\text{-a.e.} \implies \nu(A) = I(\mathbb{1}_A) = 0.$$

This means that $\nu \ll \mu$ is absolutely continuous with respect to μ, and the Radon–Nikodým theorem (Theorem 20.2) shows that there is a positive measurable function $f \in \mathcal{M}(\mathscr{A})$ such that

$$I(\mathbb{1}_A) = \nu(A) = \int \mathbb{1}_A f \, d\mu \quad \forall A \in \mathscr{A}. \tag{21.4}$$

Because of the linearity of I, (21.4) shows that $I(u) = \int uf \, d\mu$ for all simple functions $u \in \mathscr{E}(\mathscr{A}) \cap L^p(\mu)$; since every $u \in L^p(\mu)$ is the limit of simple functions (sombrero lemma, Corollary 8.9), we find with dominated convergence and the continuity of I

$$I(u) = \int uf \, d\mu \quad \forall u \in L^p(\mu). \tag{21.5}$$

Step 3. $f \in L^\infty(\mu)$ if $\mu(X) < \infty$ and $p = 1$. From (21.4) we get

$$\int_A f \, d\mu = I(\mathbb{1}_A) \leqslant \|I\| \cdot \|\mathbb{1}_A\|_{L^1(\mu)} = \|I\| \cdot \mu(A) \quad \forall A \in \mathscr{A}.$$

In particular, setting $A = \{f \geqslant n\}$, then $n\mu(A) \leqslant \int_A f \, d\mu \leqslant \|I\| \cdot \mu(A)$. If $n > \|I\|$, we conclude that $\mu(A) = 0$, i.e. $f \in L^\infty(\mu)$.

Step 4. $f \in L^q(\mu)$ if $\mu(X) < \infty$ and $1 < p < \infty$. Define $B_n := \{f^{q-1} \leqslant n\} \in \mathscr{A}$. The function $u_n := f^{q-1}\mathbb{1}_{B_n}$ is bounded and (21.5) entails with $u = u_n$

$$\begin{aligned}\int f^q \mathbb{1}_{B_n} \, d\mu = \int u_n f \, d\mu = I(u_n) &\leqslant \|I\| \cdot \left(\int u_n^p \, d\mu\right)^{1/p} \\ &= \|I\| \cdot \left(\int f^q \mathbb{1}_{B_n} \, d\mu\right)^{1/p}.\end{aligned}$$

Here we use that p and q are conjugate, i.e. $q^{-1} = 1 - p^{-1}$, and so $p(q-1) = q$. We get with monotone convergence

$$\left(\int f^q \mathbb{1}_{B_n} \, d\mu\right)^{1/q} \leqslant \|I\| \implies \|f\|_q = \sup_{n \in \mathbb{N}} \|f \mathbb{1}_{B_n}\|_q \leqslant \|I\|.$$

Step 5. Assume that $\mu(X) = \infty$. Let $(A_n)_{n \in \mathbb{N}}$ be a sequence in $\mathscr{A}$ with $A_n \uparrow X$ and $\mu(A_n) < \infty$. As in the proof of the Radon–Nikodým theorem (Theorem 20.2) we define,

$$h(x) := \sum_{n=1}^{\infty} \frac{2^{-n}}{1 + \mu(A_n)} \mathbb{1}_{A_n}(x) > 0 \quad \forall x \in X.$$

Notice that $h \in L^1(\mu)$ and $\tilde{\mu} := h \cdot \mu$ is a finite measure. Moreover,

$$\tilde{u} \in L^p(\tilde{\mu}) \iff u := h^{1/p}\tilde{u} \in L^p(\mu).$$

Therefore, $J(\tilde{u}) := I(h^{1/p}\tilde{u})$ is a positive linear functional on $L^p(\tilde{\mu})$, and the first four steps show that there is a unique $\tilde{f} \in L^q(\tilde{\mu})$ satisfying

$$I(h^{1/p}\tilde{u}) = J(\tilde{u}) = \int \tilde{u}\tilde{f}\,d\tilde{\mu} = \int \tilde{u}h\tilde{f}\,d\mu. = \int h^{1/p}\tilde{u}h^{1/q}\tilde{f}\,d\mu.$$

Setting $u = h^{1/p}\tilde{u} \in L^p(\mu)$ we conclude that $I(u) = \int uf\,d\mu$, where $f = h^{1/q}\tilde{f}$ is in $L^q(\mu)$ (as usual, $1/\infty = 0$ and $h^0 \equiv 1$). □

The family of bounded linear functionals $I : L^p(\mu) \to \mathbb{R}$ is called the *dual space* of $L^p(\mu)$ and denoted by $(L^p(\mu))^*$. We can combine Theorems 21.4 and 21.5 in order to characterize the dual space.

Corollary 21.6 *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and let $1 \leqslant p < \infty$, $1 < q \leqslant \infty$ conjugate indices. Then we can identify $(L^p(\mu))^*$ with $L^q(\mu)$ in the sense that every $I \in (L^p(\mu))^*$ is of the form $I_f(u) = \int uf\,d\mu$ for some unique $f \in L^q(\mu)$.*

Proof *Uniqueness* follows exactly as the uniqueness in Theorem 21.5.

Existence. If I is a positive linear functional, the assertion follows from Theorem 21.5. If I is a general bounded linear functional, we use Theorem 21.4 and decompose it into $I = I^+ - I^-$, where $I^\pm$ are positive linear functionals. These are of the form $I^+ = I_f$ and $I^- = I_g$ for uniquely determined, positive $f, g \in L^q(\mu)$. Therefore, $I(u) = \int u \cdot (f - g)d\mu$. Although $I^\pm$ need not be unique, the first part of the proof shows that the difference $f - g \in L^q(\mu)$ is unique. □

Remark 21.7 The assumption of σ-finiteness in Theorem 21.5 and Corollary 21.6 is needed only if $p = 1$; if $p \in (1, \infty)$, one can argue like in Step 7 of the proof of Theorem 20.2 and dispose of σ-finiteness. The quite technical argument can be found in Yosida [59, Section IV.9, Example 3]. The positive linear functionals on $L^\infty(\mu)$ are given by $I_\alpha(u) = \int u\,d\alpha$, where α is a finitely additive positive set-function. The proof uses the axiom of choice, see Dunford and Schwartz [15, Section IV.8.16] or Yosida [59, Section IV.9, Example 5].

The Riesz Representation Rheorem for $C_c(X)$

We will now turn to spaces of continuous functions. We assume that (X, d) is a metric space and we write $\mathscr{O}$, $\mathscr{C}$ and $\mathscr{B}(X)$ for the families of open, closed and Borel sets; $B_r(x) = \{y \in X : d(x, y) < r\}$ is the open ball with radius $r > 0$ and centre x, see Appendix B.

We need the following tools from topology. Recall that a set $A \subset X$ is *relatively compact* if its closure $\overline{A}$ is compact; we write $\widehat{\mathscr{O}}$ for the relatively compact open sets. A metric space (X, d) is *locally compact* if each x has a relatively compact open neighbourhood $V(x)$; without loss of generality we may assume that $V(x) = B_r(x)$ for some small $r = r(x)$. As in (17.1) and the proof of Theorem 17.8, we use the following Urysohn functions

$$f_{K,U}(x) := \frac{d(x, U^c)}{d(x, U^c) + d(x, K)} \quad \text{and} \quad d(x, A) := \inf_{a \in A} d(x, a),$$

where K is a compact set which is contained in the open set $U \subset X$. It is not difficult to see that $f_{K,U}$ is continuous and $\mathbb{1}_K \leqslant f_{K,U} \leqslant \mathbb{1}_U$ [✍]. If (X, d) is locally compact, we can even assume that U is, for a given compact set K, relatively compact. Later on we will need the following two facts which are often called *Urysohn's lemma*, see Lemma B.2 in Appendix B:

$$\forall K \subset X \text{ compact} \quad \exists u_n \in C_c(X) : u_n \downarrow \mathbb{1}_K, \tag{21.6}$$

$$\forall U \subset X \text{ open and relatively compact} \quad \exists v_n \in C_c(X) : v_n \geqslant 0,\ v_n \uparrow \mathbb{1}_U. \tag{21.7}$$

Typically, the functions u_n and v_n are Urysohn functions for suitable pairs K, U_n and K_n, U. We will also need a *continuous partition of unity*. For this we introduce the notation $\chi \prec \mathbb{1}_A$ if $\chi \leqslant \mathbb{1}_A$ and $\operatorname{supp} \chi \subset A$. Every compact set $K \subset U$ which is covered by finitely many open sets $K \subset \bigcup_{i=1}^n U_i$ has the following property:

$$\exists \chi_1, \dots, \chi_n \in C_c(X) : 0 \leqslant \chi_i \prec \mathbb{1}_{U_i} \quad \text{and} \quad \sum_{i=1}^n \chi_i(x) = 1 \quad (x \in K); \tag{21.8}$$

see Lemma B.3 for a proof. Finally, a measure μ on $(X, \mathscr{B}(X))$ is called *regular* if $\mu(K) < \infty$ for all compact sets K and

$$\begin{aligned} \mu(B) &= \inf\{\mu(U) : U \supset B,\ U \text{ open}\} \quad \forall B \in \mathscr{B}(X), \\ \mu(U) &= \sup\{\mu(K) : K \subset U,\ K \text{ compact}\} \quad \forall U \in \mathscr{O}. \end{aligned} \tag{21.9}$$

We are now going to prove the following representation theorem for positive linear functionals on $C_c(X)$.

Theorem 21.8 (Riesz) *Assume that (X, d) is a locally compact metric space and $I : C_c(X) \to \mathbb{R}$ a positive linear functional. There is a uniquely determined regular measure μ on $(X, \mathscr{B}(X))$ such that*

$$I(u) = I_\mu(u) = \int u \, d\mu, \quad u \in C_c(X). \tag{21.10}$$

The proof of Theorem 21.8 is based on a few auxiliary results.

Lemma 21.9 *Let $I : C_c(X) \to \mathbb{R}$ be a positive linear functional. The set function*

$$\nu : \mathscr{O} \to [0, \infty], \quad \nu(U) := \sup\{I(u) : u \prec \mathbb{1}_U\}, \tag{21.11}$$

enjoys the following properties:

(i) $\nu(\emptyset) = 0$ *and* $\nu(U) < \infty$ *for all relatively compact open* $U \in \widehat{\mathscr{O}}$*;*
(ii) $\nu\left(\bigcup_{i=1}^{\infty} U_i\right) \leqslant \sum_{i=1}^{\infty} \nu(U_i)$ *for* $(U_i)_{i\in\mathbb{N}} \subset \mathscr{O}$ (σ*-subadditive*)*;*
(iii) $\nu(U \cup V) = \nu(U) + \nu(V)$ *for all disjoint* $U, V \in \mathscr{O}$ (*additive*)*;*
(iv) $\nu(U) = \sup\left\{\nu(V) : \overline{V} \subset U,\ V \in \widehat{\mathscr{O}}\right\}$.

Proof (i) $\nu(\emptyset) = 0$ is obvious. Since $\overline{U}$ is compact, (21.6) shows that there is a function $f \in C_c(X)$ such that $\mathbb{1}_{\overline{U}} \leqslant f \leqslant \mathbf{1}$. Because of the monotonicity of I,

$$\nu(U) = \sup\{I(u) : u \prec \mathbb{1}_U\} \leqslant I(f) < \infty.$$

(ii) Assume that $(U_i)_{i\in\mathbb{N}} \subset \mathscr{O}$, $U = \bigcup_{i\in\mathbb{N}} U_i$ and $f \in C_c(X)$ such that $f \prec \mathbb{1}_U$. Since $\operatorname{supp} f$ is compact, there is some $n \geqslant 1$ such that

$$f \prec \mathbb{1}_{\bigcup_{i=1}^{n} U_i}.$$

We use (21.8) to construct a partition of unity $\chi_1, \dots, \chi_n \in C_c(X)$ such that

$$\chi_i f \prec \mathbb{1}_{U_i} \quad \text{and} \quad f = \sum_{i=1}^{n} \chi_i f.$$

Therefore,

$$I(f) = \sum_{i=1}^{n} I(\chi_i f) \leqslant \sum_{i=1}^{n} \nu(U_i) \leqslant \sum_{i=1}^{\infty} \nu(U_i),$$

and the claim follows if we take the supremum over all $f \in C_c(X)$ with $f \prec \mathbb{1}_U$.

(iii) Assume that $U, V \in \mathscr{O}$ are disjoint and $f \prec \mathbb{1}_U$, $g \prec \mathbb{1}_V$, $f, g \in C_c(X)$. Since $f + g \prec \mathbb{1}_{U \cup V}$, we get

$$I(f) + I(g) = I(f + g) \leqslant \nu(U \cup V),$$

taking the supremum over all admissible f, g, yields $\nu(U) + \nu(V) \leqslant \nu(U \cup V)$. Equality follows because of (ii).

(iv) Let $U \in \mathscr{O}$ and $\alpha := \sup\left\{\nu(V) : \overline{V} \subset U,\ V \in \widehat{\mathscr{O}}\right\}$. We have to check that $\nu(U) = \alpha$. For any $f_0 \in C_c(X)$ such that $\mathbb{1}_{\overline{V}} \leqslant f_0 \prec \mathbb{1}_U$ we have

$$\nu(V) \leqslant I(f_0) \leqslant \sup\{I(f) : f \prec \mathbb{1}_U\} \stackrel{\text{def}}{=} \nu(U).$$

This proves $\alpha = \sup_{\overline{V} \subset U,\ V \in \widehat{\mathscr{O}}} \nu(V) \leqslant \nu(U)$.

For the other inequality $\nu(U) \leqslant \alpha$ we may, without loss of generality, assume that $\alpha < \infty$. From the very definition of ν we see that for every $\epsilon > 0$ there is some $f_\epsilon \prec \mathbb{1}_U$, $f_\epsilon \in C_c(X)$, and a relatively compact open set $V_\epsilon \in \widehat{\mathscr{O}}$ such that $\operatorname{supp} f_\epsilon \subset V_\epsilon \subset \overline{V}_\epsilon \subset U$ and

$$\nu(U) \leqslant I(f_\epsilon) + \epsilon \leqslant \nu(V_\epsilon) + \epsilon \leqslant \alpha + \epsilon.$$

The second estimate follows, again, from the definition of ν. Letting $\epsilon \to 0$ finishes the proof. $\square$

Lemma 21.10 *Let I and ν be as in Lemma 21.9. The set function*

$$\nu^*(A) := \inf\{\nu(U) : U \supset A,\ U \in \mathscr{O}\}, \quad A \subset X, \tag{21.12}$$

is an outer measure[1] *which is inner regular, i.e.*

$$\nu^*(U) = \sup\{\nu^*(K) : K \subset U,\ K \text{ compact}\}, \quad U \in \mathscr{O}.$$

Proof (OM_1), (OM_2): From the definition of ν^* we get $\nu^*|_{\mathscr{O}} = \nu$ as well as monotonicity: $A \subset B \implies \nu^*(A) \leqslant \nu^*(B)$; in particular, $\nu^*(\emptyset) = 0$.

(OM_3): Let $A_n \subset X$, $n \in \mathbb{N}$. Using (21.12) we see that

$$\forall \epsilon > 0 \quad \exists U_n \in \mathscr{O},\ A_n \subset U_n : \nu(U_n) \leqslant \nu^*(A_n) + \epsilon 2^{-n}.$$

By the σ-subadditivity of ν,

$$\nu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leqslant \nu^*\left(\bigcup_{n=1}^{\infty} U_n\right) = \nu\left(\bigcup_{n=1}^{\infty} U_n\right) \leqslant \sum_{n=1}^{\infty} \nu(U_n) \leqslant \sum_{n=1}^{\infty} \nu^*(A_n) + \epsilon.$$

Letting $\epsilon \to 0$ shows that ν^* is also σ-subadditive.

Let $U \in \mathscr{O}$. Because of the monotonicity of μ^* we see that

$$\sup\{\nu^*(K) : K \subset U,\ K \text{ compact}\} \leqslant \nu^*(U).$$

The converse estimate follows from

$$\begin{aligned} \nu^*(U) = \nu(U) &\overset{21.9(\text{iv})}{=} \sup\left\{\nu(V) : \overline{V} \subset U,\ V \in \widehat{\mathscr{O}}\right\} \\ &\underset{\overline{V}\text{ compact}}{\overset{\nu^*(V) \leqslant \nu^*(\overline{V})}{\leqslant}} \sup\left\{\nu^*(K) : K \subset U,\ K \text{ compact}\right\}. \end{aligned}$$ $\square$

In the proof of Carathéodory's extension theorem – Theorem 6.1, (6.2) – we introduced the ν^*-measurable sets:

$$\mathscr{A}_\nu^* = \{A \subset X : \nu^*(Q) = \nu^*(Q \cap A) + \nu^*(Q \setminus A) \quad \forall Q \subset X\}.$$

Lemma 21.11 *If ν, ν^* satisfy the conditions of Lemma 21.10, then $\mathscr{B}(X) \subset \mathscr{A}_\nu^*$.*

Proof Let us show that the closed sets $\mathscr{C}$ are ν^*-measurable. Fix $F \in \mathscr{C}$ and $Q \subset U$, $U \in \mathscr{O}$. Lemma 21.9(iv) shows that

$$\exists U_n \in \mathscr{O},\ \overline{U}_n \subset U \setminus F : \nu(U_n) \uparrow \nu(U \setminus F).$$

[1] That is, a set function satisfying (OM_1)–(OM_3), page 40.

Since ν is additive and $U_n \uplus (U \setminus \overline{U}_n) = U \setminus \partial U_n$,[2] we have

$$\nu(U) \geqslant \nu(U \setminus \partial U_n) = \nu(U_n) + \nu(U \setminus \overline{U}_n) \geqslant \nu(U_n) + \nu^*(U \cap F).$$

Letting $n \to \infty$ yields

$$\nu(U) \geqslant \nu(U \setminus F) + \nu^*(U \cap F) \overset{U \supset Q}{\geqslant} \nu^*(Q \setminus F) + \nu^*(Q \cap F),$$

and, if we take the infimum over all open sets $U \supset Q$, the definition of ν^* yields $\nu^*(Q) \geqslant \nu^*(Q \setminus F) + \nu^*(Q \cap F)$. Since ν^* is (σ-)subadditive, we have the other inequality, too, and so $F \in \mathscr{A}_\nu^*$.

The proof of Theorem 6.1 (Step 3b) shows that $\mathscr{A}_\nu^*$ is a σ-algebra. Since $\mathscr{C}$ is a generator of the Borel sets $\mathscr{B}(X)$, we conclude that $\mathscr{B}(X) = \sigma(\mathscr{C}) \subset \mathscr{A}_\nu^*$. $\square$

We can finally attend to the proof of Riesz's representation theorem.

Proof of Theorem 21.8 Existence. Starting with the functional I, we construct the set functions ν and ν^* as in Lemmas 21.9–21.11. The proof of Carathéodory's extension theorem (Theorem 6.1, Step 3b on page 44) shows that $\mu := \nu^*|_{\mathscr{B}(X)}$ is a measure which is regular, see Lemma 21.10.

Let us show the representation property $I(u) = \int u \, d\mu$, $u \in C_c(X)$. Since I is linear, we may assume that $0 \leqslant u \leqslant 1$, otherwise we consider $u^\pm / \|u\|_\infty$. Define, as shown in Fig. 21.1,

$$u_k^n(x) := (n \cdot u(x) - k)^+ \wedge 1 \quad \text{and} \quad U_k^n := \{n \cdot u > k\} = \{u_k^n > 0\},$$

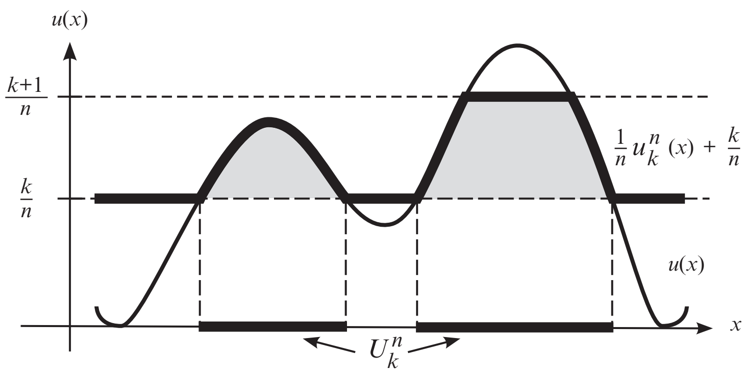


Fig. 21.1. The shaded area is the subgraph of the function $\frac{1}{n} u_k^n(x)$; up to vertical shifts the graph of the function $u(x)$ is partitioned in horizontal strips of width $\frac{1}{n}$ by the functions $\frac{1}{n} u_k^n(x)$, $k = 0, \ldots, n$.

[2] By $\partial U := \overline{U} \setminus U$ we denote the (topological) boundary of the open set U.

and observe that

$$\overline{U}_{k+1}^n = \overline{\{u_{k+1}^n > 0\}} \subset \{u_k^n = 1\}.$$

Therefore,

$$\int u_{k+1}^n \, d\mu \leqslant \int \mathbb{1}_{U_{k+1}^n} \, d\mu = \underbrace{\mu\left(U_{k+1}^n\right) \underset{\mu|_{\mathscr{O}} = \nu|_{\mathscr{O}}}{\overset{(21.11)}{\leqslant}} I(u_k^n) \leqslant \mu\left(\overline{U}_k^n\right)}_{\text{observe that } \mathbb{1}_{U_{k+1}^n} \leqslant u_k^n \leqslant \mathbb{1}_{\overline{U}_k^n}} \leqslant \int u_{k-1}^n \, d\mu.$$

Set $U_0 = U_0^n = \{u > 0\}$. Since $n \cdot u = \sum_{k=0}^n u_k^n$, can sum over the preceding inequalities and use $\int u_k^n \, d\mu \leqslant \mu(U_k^n) \leqslant \mu(U_0)$ to get

$$n \int u \, d\mu - \mu(U_0) \leqslant nI(u) \leqslant n \int u \, d\mu + \mu\left(\overline{U}_0\right).$$

Since $\operatorname{supp} u$ is compact, we see that $\mu\left(\overline{U}_0\right) < \infty$. Dividing by n and letting $n \to \infty$ finally gives $\int u \, d\mu = I(u)$.

Uniqueness. Assume that μ and ρ are two measures such that

$$\int u \, d\mu = I(u) = \int u \, d\rho \quad \forall u \in C_c(X).$$

If we combine (21.6) with monotone convergence, we get $\mu(K) = \rho(K)$ for all compact sets $K \subset X$. Since μ and ρ are regular, (21.9) shows that $\mu(B) = \rho(B)$ for all $B \in \mathscr{B}(X)$. □

We close this section with two interesting applications of the Riesz representation theorem.

The family of bounded linear functionals $I: C_c(X) \to \mathbb{R}$ is called the *dual space* of $C_c(X)$ and denoted by $(C_c(X))^*$. We can combine Theorems 21.4 and 21.8 in order to characterize the dual space. A *signed Radon measure* $\rho \in \mathfrak{M}_{\mathrm{r}}(X)$ on (X, d) is the difference $\rho = \mu - \nu$ of two regular[3] measures μ, ν on $(X, \mathscr{B}(X))$. We define $\int u \, d\rho := \int u \, d\mu - \int u \, d\nu$.

Corollary 21.12 *Let (X, d) be a locally compact metric space. We can identify $(C_c(X))^*$ with the signed Radon measures $\mathfrak{M}_{\mathrm{r}}(X)$ in the sense that for some unique $\rho \in \mathfrak{M}_{\mathrm{r}}(X)$ every $I \in (C_c(X))^*$ is of the form $I_\rho(u) = \int u \, d\rho$.*

Proof Uniqueness. Assume that $\int u \, d\rho = I(u) = \int u \, d\rho'$ for all $u \in C_c(X)$ and $\rho, \rho' \in \mathfrak{M}_{\mathrm{r}}(X)$. Since $\rho = \mu - \nu$ and $\rho' = \mu' - \nu'$, we get

$$\int u \, d(\mu - \nu) = \int u \, d(\mu' - \nu') \iff \int u \, d(\mu + \nu') = \int u \, d(\nu + \mu').$$

[3] See (21.9) for the definition.

As in the last part of the proof of Theorem 21.8 we can use Urysohn's lemma and regularity to get $\mu + \nu' = \nu + \mu'$, hence $\rho = \rho'$.

Existence. If I is a positive linear functional, the assertion follows from Theorem 21.8. If I is a general bounded linear functional, we use Theorem 21.4 and decompose it into $I = I^+ - I^-$, where $I^\pm$ are positive linear functionals. These are of the form $I^+ = I_\mu$ and $I^- = I_\nu$ for uniquely determined regular measures μ, ν. Therefore, $I(u) = \int u\, d(\mu - \nu)$. Although $I^\pm$ need not be unique, the first part of the proof shows that the difference $\mu - \nu \in \mathfrak{M}_{\mathrm{r}}(X)$ is unique. □

Our second application is yet another existence proof of Lebesgue measure – this time based on the Riemann integral.

Corollary 21.13 *An n-dimensional Lebesgue measure λ^n exists and it is the unique regular measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ with $\lambda^n\left(\bigtimes_{i=1}^n [a_i, b_i]\right) = \prod_{i=1}^n (b_i - a_i)$.*

Proof On $C_c(\mathbb{R}^n)$ the Riemann integral defines a positive linear functional:

$$I(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n) dx_1 \cdots dx_n, \quad u \in C_c(\mathbb{R}^n).$$

By Theorem 21.8, there is a unique measure λ^n such that $I(u) = \int u\, d\lambda^n$.

Let $J = \bigtimes_{i=1}^n [a_i, b_i]$ and define $J(h) := \bigtimes_{i=1}^n [a_i - h, b_i + h]$ for any $h \in \mathbb{R}$. Using Urysohn functions (21.6), (21.7) we can construct $u_h, w_h \in C_c(\mathbb{R}^n)$ such that $\mathbb{1}_{J(-h)} \leqslant u_h \leqslant \mathbb{1}_J \leqslant w_h \leqslant \mathbb{1}_{J(h)}$. Therefore,

$$\begin{aligned}
\int_{a_1+h}^{b_1-h} \cdots \int_{a_n+h}^{b_n-h} dx_1 \cdots dx_n &\leqslant \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_h(x_1, \ldots, x_n) dx_1 \cdots dx_n \\
&\leqslant \lambda^n([a_1, b_1] \times \cdots \times [a_n, b_n]) \\
&\leqslant \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} w_h(x_1, \ldots, x_n) dx_1 \cdots dx_n \\
&\leqslant \int_{a_1-h}^{b_1+h} \cdots \int_{a_n-h}^{b_n+h} dx_1 \cdots dx_n.
\end{aligned}$$

Letting $h \to 0$ we see that $\lambda^n(J) = \prod_{i=1}^n (b_i - a_i)$, which means that λ^n is Lebesgue measure. □

Vague and Weak Convergence of Measures

We will finally discuss the convergence of measures. As before, (X, d) is a locally compact metric space and $\mathscr{B}(X)$ are the Borel sets in X. We write $C_c(X)$

and $C_b(X)$ for the compactly supported resp. bounded continuous functions $u\colon X \to \mathbb{R}$, and $\mathfrak{M}_r^+(X)$ denotes the regular measures[4] on $(X, \mathscr{B}(X))$.

Definition 21.14 A sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathfrak{M}_r^+(X)$ converges *vaguely* [*weakly*] to a measure $\mu \in \mathfrak{M}_r^+(X)$ if

$$\lim_{n\to\infty} \int u\, d\mu_n = \int u\, d\mu \quad \forall u \in C_c(X) \quad \big[\forall u \in C_b(X)\big].$$

We write $\mu_n \xrightarrow{\mathrm{v}} \mu$ and $\mu_n \xrightarrow{\mathrm{w}} \mu$ to indicate vague and weak convergence.

If the limit $I(u) = \lim_{n\to\infty} \int u\, d\mu_n$ exists for all $u \in C_c(X)$, it defines a positive linear functional and the Riesz representation theorem, Theorem 21.8, shows that $I(u) = \int u\, d\mu$ for a unique and regular measure μ. In particular, vague limits are unique.

Vague convergence can be characterized by a portmanteau-type theorem. For any set $A \subset X$ we denote by A° the open interior, i.e. the largest open set contained in A, and by $\overline{A}$ the closure, i.e. the smallest closed set containing A; the topological boundary is $\partial A = \overline{A} \setminus A^\circ$.

Theorem 21.15 (portmanteau theorem) *Let (X, d) be a locally compact metric space and $\mu, \mu_n \in \mathfrak{M}_r^+(X)$, $n \in \mathbb{N}$. The following assertions are equivalent:*

(a) $\mu_n \xrightarrow{v} \mu$;
(b) $\limsup_{n\to\infty} \mu_n(K) \leqslant \mu(K)$ *for all compact sets* $K \subset X$ *and* $\liminf_{n\to\infty} \mu_n(U) \geqslant \mu(U)$ *for all relatively compact open sets* $U \subset X$;
(c) $\lim_{n\to\infty} \mu_n(B) = \mu(B)$ *for all relatively compact Borel sets* $B \subset X$ *such that* $\mu(\partial B) = 0$.

Proof (a)$\Rightarrow$(b). Let K be a compact set and use (21.6) to construct a sequence $u_k \in C_c(X)$ with $u_k \downarrow \mathbb{1}_K$. We have for all $k \in \mathbb{N}$

$$\limsup_{n\to\infty} \mu_n(K) = \limsup_{n\to\infty} \int \mathbb{1}_K\, d\mu_n \leqslant \limsup_{n\to\infty} \int u_k\, d\mu_n \overset{\text{(a)}}{=} \int u_k\, d\mu.$$

Now we can use monotone convergence for the limit $k \to \infty$,

$$\limsup_{n\to\infty} \mu_n(K) \leqslant \inf_{k\in\mathbb{N}} \int u_k\, d\mu = \int \inf_{k\in\mathbb{N}} u_k\, d\mu = \int \mathbb{1}_K\, d\mu = \mu(K).$$

Let U be a relatively compact open set. Using (21.7) we can find then a sequence $w_k \in C_c(X)$ such that $w_k \uparrow \mathbb{1}_U$. Therefore,

$$\int w_k\, d\mu = \liminf_{n\to\infty} \int w_k\, d\mu_n \leqslant \liminf_{n\to\infty} \int \mathbb{1}_U\, d\mu_n = \liminf_{n\to\infty} \mu_n(U).$$

[4] See (21.9) or Appendix H for the definition; in particular, regular measures are finite on compact sets.

Again by monotone convergence,

$$\mu(U) = \int \mathbb{1}_U \, d\mu = \int \sup_{k\in\mathbb{N}} w_k \, d\mu = \sup_{k\in\mathbb{N}} \int w_k \, d\mu \leqslant \liminf_{n\to\infty} \mu_n(U).$$

(b) $\Rightarrow$ (c) If $B \in \mathscr{B}(X)$ is relatively compact, then $\overline{B}$ is compact, B° is open and relatively compact, and $B^\circ \subset B \subset \overline{B}$. Therefore, (b) and the monotonicity of measures show for all $B \in \mathscr{B}(X)$ with $\mu(\partial B) = 0$ that

$$\begin{aligned}\limsup_{n\to\infty} \mu_n(B) \leqslant \limsup_{n\to\infty} \mu_n(\overline{B}) \overset{\text{(b)}}{\leqslant} \mu(\overline{B}) \overset{\mu(\partial B)=0}{=} \mu(B^\circ) &\overset{\text{(b)}}{\leqslant} \liminf_{n\to\infty} \mu_n(B^\circ) \\ &\leqslant \liminf_{n\to\infty} \mu_n(B),\end{aligned}$$

which means that $\lim_{n\to\infty} \mu_n(B) = \mu(B)$.

(c) $\Rightarrow$ (a) Let $u \in C_c(X)$, $u \geqslant 0$. For $t > 0$ we have $\partial\{u \geqslant t\} \subset \{u = t\} \subset \operatorname{supp} u$ and $\mu(\operatorname{supp} u) < \infty$; therefore, $\mu(\partial\{u \geqslant t\}) > 0$ for at most countably many t [✍]. Now we can use (14.7), dominated convergence and the fact that countable sets are Lebesgue null sets to get

$$\begin{aligned}\lim_{n\to\infty} \int u \, d\mu_n \overset{(14.7)}{=} \lim_{n\to\infty} \int_0^{\|u\|_\infty} \mu_n\{u \geqslant t\} dt &= \int_0^{\|u\|_\infty} \lim_{n\to\infty} \mu_n\{u \geqslant t\} dt \\ &\overset{\text{(c)}}{=} \int_0^{\|u\|_\infty} \mu\{u \geqslant t\} dt \\ &\overset{(14.7)}{=} \int u \, d\mu.\end{aligned}$$

If u has arbitrary sign, we use $u = u^+ - u^-$ and the linearity of the integral. $\square$

Caution The difference between vague and weak convergence is the fact that weak convergence preserves the total mass – this is clear since $1 = \mathbb{1}_X \in C_b(X)$ – while vague convergence allows for 'dissipation' of mass. Consider, for example, the family $(\delta_n)_{n\in\mathbb{N}}$ on $(0, \infty)$, which converges vaguely to $\mu \equiv 0$ – but not weakly.

Lemma 21.16 *Let (X, d) be a locally compact metric space, $\mu, \mu_n \in \mathfrak{M}_{\mathrm{r}}^+(X)$ and $\mu_n \xrightarrow{\mathrm{v}} \mu$. Then $\mu(X) \leqslant \liminf_{n\to\infty} \mu_n(X)$.*

Proof We have for all $u \in C_c(X)$ satisfying $0 \leqslant u \leqslant 1$

$$\int u \, d\mu \underset{\text{limit}}{\overset{\text{vague}}{=}} \liminf_{n\to\infty} \int u \, d\mu_n \leqslant \liminf_{n\to\infty} \int 1 \, d\mu_n = \liminf_{n\to\infty} \mu_n(X).$$

By Urysohn's lemma (21.6), for any compact set K there is some $u \in C_c(X)$ with $\mathbb{1}_K \leqslant u \leqslant 1$. Because of the regularity of the measure μ,

$$\mu(X) = \sup_{K\subset X,\ \text{compact}} \mu(K) \leqslant \sup_{u\in C_c(X), u\leqslant 1} \int u \, d\mu \leqslant \liminf_{n\to\infty} \mu_n(X). \qquad \square$$

Vague and weak convergence coincide, if the limit preserves the total mass.

Theorem 21.17 *Assume that (X, d) is a locally compact metric space and that $\mu, \mu_n \in \mathfrak{M}_r^+(X)$ are finite measures. The following assertions are equivalent:*

(a) $\mu_n \xrightarrow{\mathrm{w}} \mu$;
(b) $\mu_n \xrightarrow{\mathrm{v}} \mu$ *and mass preservation:* $\lim_{n\to\infty} \mu_n(X) = \mu(X)$;
(c) $\mu_n \xrightarrow{\mathrm{v}} \mu$ *and tightness* $\forall \epsilon > 0\ \exists K \subset X$ *compact* : $\sup_{n\in\mathbb{N}} \mu_n(K^c) < \epsilon$.

In particular, vague and weak convergence of probability measures coincide.

Proof The direction (a) $\Rightarrow$ (b) is obvious.

(b) $\Rightarrow$ (c). Since the measures μ_n and μ are regular, we see that for every $\epsilon > 0$ and $n \in \mathbb{N}$ there are compact sets $K_{0,\epsilon}, K_{n,\epsilon} \subset X$ with

$$\mu_n(K_{n,\epsilon}^c) \leqslant \epsilon \quad \text{and} \quad \mu(K_{0,\epsilon}^c) \leqslant \epsilon.$$

From Urysohn's lemma (21.6) there is a relatively compact, open set $U_\epsilon \supset K_{0,\epsilon}$;[5] set $K_\epsilon := \overline{U}_\epsilon$ and apply the portmanteau theorem (Theorem 21.15(b)):

$$\mu(X) \leqslant \mu(K_{0,\epsilon}) + \epsilon \leqslant \mu(U_\epsilon) + \epsilon \leqslant \liminf_{n\to\infty} \mu_n(U_\epsilon) + \epsilon \leqslant \liminf_{n\to\infty} \mu_n(K_\epsilon) + \epsilon.$$

Since, by assumption, $\mu(X) = \lim_{n\to\infty} \mu_n(X)$, we get

$$\limsup_{n\to\infty} \mu_n(X \setminus K_\epsilon) = \lim_{n\to\infty} \mu_n(X) - \liminf_{n\to\infty} \mu_n(K_\epsilon) \leqslant \epsilon.$$

Thus, there is some $N = N(\epsilon) \in \mathbb{N}$ such that $\sup_{n>N} \mu_n(X \setminus K_\epsilon) \leqslant 2\epsilon$. Since the measures $\mu_1, \dots, \mu_N$ are regular, the compact set $K := K_\epsilon \cup K_{1,\epsilon} \cup \dots \cup K_{N,\epsilon}$ satisfies (c).

(c) $\Rightarrow$ (a). Fix $u \in C_b(X)$ and $\epsilon > 0$. By assumption, there is a compact set $K \subset X$ such that $\sup_n \mu_n(K^c) \leqslant \epsilon$. Since μ is regular we can also assume that $\mu(K^c) \leqslant \epsilon$, otherwise we could enlarge K. With (21.6) we find some $\chi \in C_c(X)$ such that $\mathbb{1}_K \leqslant \chi \leqslant 1$ and

$$\begin{aligned}
\left| \int u\,d\mu_n - \int u\,d\mu \right| &\leqslant \left| \int u\chi\,d\mu_n - \int u\chi\,d\mu \right| + \left| \int u(1-\chi)d\mu_n - \int u(1-\chi)d\mu \right| \\
&\leqslant \left| \int u\chi\,d\mu_n - \int u\chi\,d\mu \right| + \|u\|_\infty \left[\int \mathbb{1}_{K^c}\,d\mu_n + \int \mathbb{1}_{K^c}\,d\mu \right] \\
&\leqslant \left| \int u\chi\,d\mu_n - \int u\chi\,d\mu \right| + 2\|u\|_\infty \epsilon.
\end{aligned}$$

[5] Use, e.g. $\{u_k > 0\}$, $\epsilon = 1/k$, in the notation of (21.7).

Since $u\chi \in C_c(X)$, vague convergence shows

$$\limsup_{n\to\infty} \left| \int u\,d\mu_n - \int u\,d\mu \right| \leqslant 2\|u\|_\infty \epsilon \xrightarrow[\epsilon\to 0]{} 0.$$

The last assertion follows from the fact that probability measures on locally compact spaces are regular, see Appendix H. □

Many existence assertions require compactness. Therefore, we are interested in a compactness criterion for vaguely convergent *sequences*. In general, compactness does not guarantee the existence of convergent (sub)sequences. If, however, the space $C_c(X)$ is separable (i.e. contains a countable dense subset), then we can work with sequences. Typically, $C_c(X)$ is separable if the underlying metric space (X,d) is separable.

In $\mathbb{R}^n$ compact sets are closed and bounded and we can 'catch' every compact set K with a closed ball $\overline{B_N(0)}$ with large enough radius $N = N(K)$. This absorption property is far from obvious in general metric spaces. A metric space is σ-compact if there is a sequence of compact sets $K_n \uparrow X$. If X is locally compact, we even get a sequence of compact sets $L_n \subset L_{n+1}^\circ \subset L_{n+1} \uparrow X$ which can catch any compact K, i.e. $K \subset L_N$ if $N = N(K)$ is large, see Lemma B.4.

Theorem 21.18 *Let (X,d) be a locally compact and σ-compact metric space such that $(C_c(X), \|\cdot\|_\infty)$ is separable. If $\mathfrak{N} \subset \mathfrak{M}_r^+(X)$ is vaguely bounded, i.e.*

$$\sup_{\nu\in\mathfrak{N}} \int |u|\,d\nu = c(u) < \infty \quad \forall u \in C_c(X), \tag{21.13}$$

then there is a sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathfrak{N}$ and a measure $\mu \in \mathfrak{M}_r^+(X)$ such that $\mu_n \xrightarrow{\text{v}} \mu$.

Proof Step 1. Using Urysohn's lemma (21.6) and the absorption property, we can construct a sequence $(\chi_n)_{n\in\mathbb{N}} \subset C_c(X)$ such that

$$0 \leqslant \chi_n \uparrow 1 \quad \text{and} \quad \forall u \in C_c(X) \quad \exists N = N_u : \mathbb{1}_{\operatorname{supp} u} \leqslant \chi_N.$$

Denote by $\mathcal{D} = \{w_1, w_2, \dots\}$ a countable dense subset of $C_c(X)$. Clearly, again $\widetilde{\mathcal{D}} := \{w_k\chi_n : k, n \in \mathbb{N}\}$ is a countable dense subset of $C_c(X)$.

Step 2. Denote by $\{u_1, u_2, u_3, \dots\}$ an enumeration of the set $\widetilde{\mathcal{D}}$ from Step 1. We will construct $(\mu_n)_{n\in\mathbb{N}}$ recursively: set $\langle \nu, u\rangle := \int u\,d\nu$. By assumption,

$$(\langle \nu, u_i\rangle)_{\nu\in\mathfrak{N}} \subset [-c_i, c_i],\ c_i := c(u_i) \quad \forall i \in \mathbb{N}.$$

Since the right-hand side is a compact interval, we can use the Bolzano–Weierstraß theorem to extract convergent subsequences:

$$\begin{aligned}
(\langle \nu, u_1\rangle)_{\nu\in\mathfrak{N}} \subset [-c_1, c_1] \quad &\Longrightarrow \exists (\nu_n^1)_{n\in\mathbb{N}} \subset \mathfrak{N}: & I(u_1) = \lim_{n\to\infty} \langle \nu_n^1, u_1\rangle;\\
\Big(\langle \nu_n^1, u_2\rangle\Big)_{n\in\mathbb{N}} \subset [-c_2, c_2] \quad &\Longrightarrow \exists (\nu_n^2)_{n\in\mathbb{N}} \subset (\nu_n^1)_{n\in\mathbb{N}}: & I(u_2) = \lim_{n\to\infty} \langle \nu_n^2, u_2\rangle;\\
&\vdots\\
\Big(\langle \nu_n^{i-1}, u_i\rangle\Big)_{n\in\mathbb{N}} \subset [-c_i, c_i] \quad &\Longrightarrow \exists (\nu_n^i)_{n\in\mathbb{N}} \subset (\nu_n^{i-1})_{n\in\mathbb{N}}: & I(u_i) = \lim_{n\to\infty} \langle \nu_n^i, u_i\rangle.
\end{aligned}$$

Since we have recursively decimated the original sequence, $I(u_k) = \lim_{n\to\infty}\langle \nu_n^i, u_k\rangle$ exists for all $k = 1, \dots, i$. Therefore, the diagonal sequence $\mu_n := \nu_n^n$ satisfies

$$I(u) = \lim_{n\to\infty} \langle \mu_n, u\rangle = \lim_{n\to\infty} \int u\, d\mu_n \quad \forall u \in \widetilde{\mathcal{D}}.$$

Step 3. Let $u \in C_c(X)$ and $\epsilon > 0$. As in Step 1, there are functions $w_\epsilon \in \mathcal{D}$ and $\chi_N \in C_c(X)$ such that $f_\epsilon := w_\epsilon \chi_N \in \widetilde{\mathcal{D}}$ satisfies

$$\|u - f_\epsilon\|_\infty < \epsilon \quad \text{and} \quad |u - f_\epsilon| \leqslant \epsilon\chi_N.$$

This shows

$$\begin{aligned}
&\left|\int u\, d\mu_n - \int u\, d\mu_m\right|\\
&\quad\leqslant \left|\int (u - f_\epsilon) d\mu_n\right| + \left|\int f_\epsilon\, d\mu_n - \int f_\epsilon\, d\mu_m\right| + \left|\int (f_\epsilon - u) d\mu_m\right|\\
&\quad\leqslant \int |u - f_\epsilon| d\mu_n + \left|\int f_\epsilon\, d\mu_n - \int f_\epsilon\, d\mu_m\right| + \int |f_\epsilon - u| d\mu_m\\
&\quad\leqslant \epsilon\left(\int \chi_N\, d\mu_n + \int \chi_N\, d\mu_m\right) + \left|\int f_\epsilon\, d\mu_n - \int f_\epsilon\, d\mu_m\right|. \qquad (21.14)
\end{aligned}$$

Because of (21.13) and $f_\epsilon \in \widetilde{\mathcal{D}}$ we find for all $m, n \to \infty$

$$\limsup_{m,n\to\infty} \left|\int u\, d\mu_n - \int u\, d\mu_m\right| \leqslant 2\epsilon c(\chi_N) \xrightarrow[\epsilon\to 0]{} 0. \qquad (21.15)$$

This means that $I(u) := \lim_{n\to\infty} \int u\, d\mu_n$ exists for all $u \in C_c(X)$ and defines a positive linear functional. The assertion follows now from Riesz's representation theorem, Theorem 21.8. □

If the family $\mathfrak{N}$ satisfies $\sup_{\nu\in\mathfrak{N}} \nu(X) < \infty$, e.g. if $\mathfrak{N}$ are probability measures, then vague boundedness is always satisfied and we can even do away with the topological assumption that X is σ-compact. Indeed, this follows if we take at the steps (21.14) and (21.15) in the above proof $\chi_N \equiv 1$ and $c(1) = \sup_{\nu\in\mathfrak{N}} \nu(X)$, respectively.

Corollary 21.19 *Let (X, d) be a locally compact metric space such that the space $(C_c(X), \|\cdot\|_\infty)$ is separable. Set $\mathfrak{M}^+_{\leqslant 1}(X) := \{\mu \in \mathfrak{M}^+_{\mathrm{r}}(X) : \mu(X) \leqslant 1\}$. Every sequence of measures in $\mathfrak{M}^+_{\leqslant 1}(X)$ has a vaguely convergent subsequence and the limit is again in $\mathfrak{M}^+_{\leqslant 1}(X)$. In other words: $\mathfrak{M}^+_{\leqslant 1}(X)$ is vaguely sequentially compact.*

Proof By Theorem 21.18 there is a subsequential limit μ. Let $(\mu'_n)_{n\in\mathbb{N}} \subset \mathfrak{M}^+_{\leqslant 1}(X)$ be a sequence such that vague-$\lim_{n\to\infty} \mu'_n = \mu$. Then Lemma 21.16 shows that $\mu(X) \leqslant 1$, i.e. $\mu \in \mathfrak{M}^+_{\leqslant 1}(X)$. □

Problems

21.1. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space, $f \in \mathcal{M}(\mathscr{A})$ and $1 \leqslant p < \infty$, $1 < q \leqslant \infty$ conjugate indices.

(i) If $f \in L^p(\mu)$, then $\|f\|_p = \sup\left\{\int fg\,d\mu : g \in L^q(\mu),\ \|g\|_q \leqslant 1\right\}$.

(ii) Show that we can replace in (i) the set $L^q(\mu)$ with a dense subset $\mathcal{D} \subset L^q(\mu)$.

(iii) If $fg \in L^1(\mu)$ for all $g \in L^q(\mu)$, then $f \in L^p(\mu)$.
[*Hint:* consider $g \mapsto I_f(g) = \int |f| g\,d\mu$ and use Theorem 21.5.]

21.2. Weak convergence in L^p. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and let $p \in (1, \infty]$, $q \in [1, \infty)$ be conjugate indices. We assume that $L^q(\mu)$ is separable, containing the countable dense set $\mathcal{D}_q$.

(i) If $(u_n)_{n\in\mathbb{N}}$ is bounded in L^p, i.e. $\sup_n \|u_n\|_p < \infty$, then there is a subsequence such that $\lim_{i\to\infty} \int u_{n(i)} g\,d\mu$ exists for all $g \in \mathcal{D}_q$.
[*Hint:* use the Bolzano–Weierstraß theorem and a diagonal argument.]

(ii) The limit in part (i) exists for all $g \in L^q(\mu)$.

(iii) Conclude from parts (i) and (ii) that there exists a function $u \in L^p(\mu)$ such that $\lim_i \int u_{n(i)} g\,d\mu = \int ug\,d\mu$ ('*weak convergence in L^p*'). This proves the following result: *every L^p-bounded sequence has a weakly convergent subsequence*.
[*Hint:* set $g \mapsto I(g) := \lim_i \int u_{n(i)} g\,d\mu$ and use Theorem 21.5, hence the restriction $p > 1$.]

21.3. Lévy's continuity theorem. The steps below sketch a proof of the following theorem.
Theorem (P. Lévy; continuity theorem). *Let $(\mu_i)_{i\in\mathbb{N}}$ be a sequence of finite measures on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. If the sequence of Fourier transforms $(\widehat{\mu}_n)_{n\in\mathbb{N}}$ converges pointwise to a function ϕ which is continuous at $\xi = 0$, then there is a finite measure μ such that $\mu_i \to \mu$ weakly. In particular, $\widehat{\mu} = \phi$ and the convergence of the functions $\widehat{\mu}_i$ is locally uniform.*

(i) If $\lim_{i\to\infty} \widehat{\mu}_i(\xi) = \phi(\xi)$ for every $\xi \in \mathbb{R}^n$, then $\phi(\xi)$ is a positive semidefinite function (see Problem 19.7 or 21.4). In particular, $|\phi(\xi)| \leqslant \phi(0)$.
[*Hint:* use that $\big(\phi(0)\big)$ and

$$\begin{pmatrix} \phi(0) & \phi(-\xi) \\ \phi(\xi) & \phi(0) \end{pmatrix}$$

are positive semidefinite matrices.]

(ii) Show that $u \mapsto \Lambda u := \lim_i \int u\, d\mu_i$, $u \in C_c^\infty(\mathbb{R}^n)$, is a positive linear functional.
[*Hint:* use Theorem 19.2.]

(iii) Show that Λ from (ii) satisfies $|\Lambda u| \leqslant (2\pi)^n \phi(0) \|u\|_\infty$ and extend Λ to a positive linear functional on $C_c(\mathbb{R}^n)$. This allows you to identify Λ with a measure μ.
[*Hint:* use Problem 15.13.]

(iv) If $\lim_{i\to\infty} \widehat{\mu}_n(\xi) = \phi(\xi)$ for every $\xi \in \mathbb{R}^n$, and if ϕ is continuous at $\xi = 0$, then the sequence $(\mu_i)_{i\in\mathbb{N}}$ is tight, i.e.

$$\forall \epsilon > 0 \quad \exists R = R_\epsilon > 0 : \sup_i \mu_i(\mathbb{R}^n \setminus [-R, R]^n) \leqslant \epsilon.$$

[*Hint:* use Problem 19.6.]

(v) Show that μ is a finite measure and $\mu_i \to \mu$ weakly.

(vi) Assume that $(\mu_i)_{i\in\mathbb{N}}$ is a weakly convergent sequence of finite measures. Show that the sequence $(\widehat{\mu}_i)_{i\in\mathbb{N}}$ is equicontinuous.

(vii) Conclude with (vi) and tightness that the sequence $(\mu_i)_{i\in\mathbb{N}}$ converges to a finite measure μ, and that $\widehat{\mu}_i(\xi) \to \widehat{\mu}(\xi)$ uniformly on compact sets.

21.4. Bochner's theorem. A function $\phi : \mathbb{R}^n \to \mathbb{C}$ is positive semidefinite, if the matrices $\big(\phi(\xi_i - \xi_k)\big)_{i,k=1}^m$ are positive hermitian for all $m \geqslant 1$ and $\xi_1, \ldots, \xi_m \in \mathbb{R}^n$, i.e. $\sum_{i,k=1}^m \phi(\xi_i - \xi_k)\lambda_i \bar{\lambda}_k \geqslant 0$ for all $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$. The steps below sketch the proof of the theorem.
Theorem (Bochner). *A continuous function $\phi : \mathbb{R}^n \to \mathbb{C}$ is the Fourier transform of a finite measure μ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ if, and only if, ϕ is positive semidefinite.*

(i) The Fourier transform $\widehat{\mu}(\xi)$ of a finite measure μ is continuous and positive semidefinite.

(ii) Assume that ϕ is continuous and positive semidefinite. Then $\phi(0) \geqslant 0$, $\phi(\xi) = \overline{\phi(-\xi)}$, and $|\phi(\xi)| \leqslant \phi(0)$.
[*Hint:* use that $\big(\phi(0)\big)$ and

$$\begin{pmatrix} \phi(0) & \phi(-\xi) \\ \phi(\xi) & \phi(0) \end{pmatrix}$$

are positive semidefinite matrices.]

(iii) Prove that $\nu_\epsilon(x) := \iint \phi(\xi - \eta)\big(e^{ix\cdot\xi} e^{-2\epsilon|\xi|^2}\big)\overline{\big(e^{ix\cdot\eta} e^{-2\epsilon|\eta|^2}\big)}\, d\xi\, d\eta$ is positive and that

$$\nu_\epsilon(x) = \frac{1}{c} \int \phi_\epsilon(\eta) e^{ix\cdot\eta}\, d\eta \quad \text{with } \phi_\epsilon(\eta) = e^{-\epsilon|\eta|^2}\phi(\eta).$$

(iv) Show that the function ν_ϵ is Lebesgue integrable.

(v) Conclude that ϕ_ϵ is the Fourier transform of $c\nu_\epsilon$. Apply Lévy's continuity theorem (Problem 21.3) to $\lim_{\epsilon\to 0} \phi_\epsilon(\xi) = \phi(\xi)$.

21.5. Let (X, d) be a locally compact metric space. Write $C_\infty(X) := \overline{C_c(X)}$ for the closure of $C_c(X)$ with respect to the uniform norm. Show that

(i) $C_\infty(X) = \{u \in C(X) : \forall \epsilon > 0\, \exists K_\epsilon \text{ compact}, \sup_{x\notin K_\epsilon} |u(x)| \leqslant \epsilon\}$;

(ii) $C_\infty(X)$ equipped with the norm $\|\cdot\|$ is a Banach space;

(iii) for the measures $\mu, \mu_n \in \mathfrak{M}_r^+(X)$ we have

$$\mu_n \xrightarrow{\text{v}} \mu \quad \text{and} \quad \sup_{n\in\mathbb{N}} \mu_n(X) < \infty \implies \int u\, d\mu_n \xrightarrow[n\to\infty]{} \int u\, d\mu \quad \forall u \in C_\infty(X).$$

21.6. Use Theorem 21.18 for a new proof of Lévy's continuity theorem (Problem 21.3).

(i) Show (as in Problem 21.3) that $\lim_{i\to\infty} \int u\,d\mu_i$ exists for all $u \in C_c(\mathbb{R}^n)$. Since $u \in C_c(\mathbb{R}^n)$ implies $|u| \in C_c(\mathbb{R}^n)$, the sequence is vaguely bounded and there is a measure μ and a vaguely convergent subsequence $\mu_{n(i)} \to \mu$.

(ii) Part (i) can be applied to every subsequence of $(\mu_i)_i$ and it follows that every subsequence contains a vaguely convergent sub-subsequence, all of which have the *same limit* μ. Therefore, $\mu_i \to \mu$ vaguely.

(iii) Use Lévy's *truncation inequality* (Problem 19.6) to show the tightness of the measures $(\mu_i)_i$; Theorem 21.17 proves that $\mu_i \to \mu$ weakly.

21.7. Let (X,d) be a locally compact metric space and $\mu, \mu_n \in \mathfrak{M}_{\mathrm{r}}^+(X)$, $\mu_n \xrightarrow{\mathrm{v}} \mu$. Prove that

$$\lim_n \int_B u\,d\mu_n = \int_B u\,d\mu \quad \forall u \in C_c(X),\ B \in \mathscr{B}(X),\ \mu(\partial B) = 0.$$

[*Hint:* have a look at the proof of Theorem 21.15(iii).]

22

Uniform Integrability and Vitali's Convergence Theorem

Lebesgue's dominated convergence theorem gives a sufficient condition which allows us to interchange limits and integrals. A crucial ingredient is the assumption that $|u_n| \leqslant w$ a.e. for all $n \in \mathbb{N}$ and some positive $w \in \mathcal{L}^1(\mu)$. This condition is not necessary, but a slightly weaker one is indeed necessary and sufficient in order for us to be able to swap limits and integrals. The key idea is to control the size of the sets where the u_n exceed a given reference function. This is the rationale behind the next definition.

Definition 22.1 Let $(X, \mathscr{A}, \mu)$ be a measure space and $\mathcal{F} \subset \mathcal{M}(\mathscr{A})$ be a family of measurable functions. We call $\mathcal{F}$ *uniformly integrable* (also *equi-integrable*) if

$$\forall \epsilon > 0 \quad \exists w_\epsilon \in \mathcal{L}^1(\mu),\ w_\epsilon \geqslant 0 : \sup_{u \in \mathcal{F}} \int_{\{|u| > w_\epsilon\}} |u| d\mu < \epsilon. \tag{22.1}$$

Note that there are other (but for $\mu(X) < \infty$ usually equivalent) definitions of uniform integrability, see Theorem 22.9 below for a discussion. The present formulation is due to G. A. Hunt [23, p. 33].

The other key assumption in the dominated convergence theorem is that the pointwise limit $\lim_{n\to\infty} u_n(x) = u(x)$ exists for (almost) all $x \in X$; we can relax this assumption, too.

Definition 22.2 (convergence in measure) Let $(X, \mathscr{A}, \mu)$ be a measure space. A sequence of measurable functions $u_n : X \to \overline{\mathbb{R}}$ *converges in measure*[1] if

$$\forall \epsilon > 0\, \forall A \in \mathscr{A},\ \mu(A) < \infty : \lim_{n\to\infty} \mu(\{|u_n - u| > \epsilon\} \cap A) = 0 \tag{22.2}$$

holds for some $u \in \mathcal{M}(\mathscr{A})$. We write $\mu\text{-}\lim_{n\to\infty} u_n = u$ or $u_n \xrightarrow{\mu} u$.

[1] If μ is a probability measure one usually speaks of *convergence in probability* or *stochastic convergence*.

Example 22.3 Convergence in measure is strictly weaker than pointwise convergence. To see this, take $(X, \mathscr{A}, \mu) = ([0,1], \mathscr{B}[0,1], \lambda^1|_{[0,1]})$ and set

$$u_n(x) := \mathbb{1}_{[i2^{-k},(i+1)2^{-k}]}(x), \quad n = i + 2^k,\ 0 \leqslant i < 2^k.$$

This is a sequence of rectangular pulses of width 2^{-k} moving in 2^k steps through $[0,1]$, jump back to $x=0$, halve their width and start moving again. Obviously,

$$\lambda^1\{|u_n| > \epsilon\} = 2^{-k} \xrightarrow[n=n(k)\to\infty]{} 0 \qquad \forall \epsilon \in (0,1),$$

so that $u_n \xrightarrow{\lambda^1} 0$ in measure. Since $\liminf_{n\to\infty} u(x) = 0 < 1 = \limsup_{n\to\infty} u(x)$, [✍] the pointwise limit $\lim_{n\to\infty} u_n(x)$ does not exist anywhere.

Lemma 22.4 *Let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$, $p \in [1,\infty)$, and $(w_n)_{n\in\mathbb{N}} \subset \mathcal{M}(\mathscr{A})$. Then*

(i) $\lim\limits_{n\to\infty} \|u_n - u\|_p = 0$ *implies* $u_n \xrightarrow{\mu} u$*;*
(ii) $\lim\limits_{n\to\infty} w_n(x) = w(x)$ *a.e. implies* $w_n \xrightarrow{\mu} w$.

Let $f: \mathbb{R} \to \mathbb{R}$ be a continuous function. If $\mu\{|w| = \infty\} = 0$, then

(iii) $w_n \xrightarrow{\mu} w$ *implies* $f \circ w_n \xrightarrow{\mu} f \circ w$*;*
(iv) $w_n \xrightarrow{\mu} w$ *implies* $w_n \wedge v \xrightarrow{\mu} w \wedge v$ *for any* $v \in \mathcal{M}(\mathscr{A})$.

Proof (i) follows immediately from the Markov inequality (11.4),

$$\mu(\{|u_n - u| > \epsilon\} \cap A) \leqslant \underbrace{\mu\{|u_n - u|^p > \epsilon^p\}}_{=\mu\{|u_n - u| > \epsilon\}} \leqslant \frac{1}{\epsilon^p}\|u_n - u\|_p^p.$$

(ii) Observe that for all $\epsilon > 0$

$$\{|w_n - w| > \epsilon\} \subset \{|w_n - w| \geqslant \epsilon\} = \{\epsilon \wedge |w_n - w| \geqslant \epsilon\}.$$

An application of the Markov inequality (11.3) yields

$$\begin{aligned}\mu(\{|w_n - w| > \epsilon\} \cap A) &\leqslant \mu(\{\epsilon \wedge |w_n - w| \geqslant \epsilon\} \cap A)\\ &\leqslant \frac{1}{\epsilon}\int_A \epsilon \wedge |w_n - w|\,d\mu = \frac{1}{\epsilon}\int \big(\epsilon \wedge |w_n - w|\big)\mathbb{1}_A\,d\mu.\end{aligned}$$

If $\mu(A) < \infty$, the function $\epsilon\mathbb{1}_A \in \mathcal{L}^1(\mu)$ is integrable and dominates the integrand $\big(\epsilon \wedge |w_n - w|\big)\mathbb{1}_A$, and Lebesgue's dominated convergence theorem implies that $\lim_{n\to\infty}\int_A(\epsilon \wedge |w_n - w|)\,d\mu = 0$.

(iii) Let $R>0$ and $\epsilon,\delta\in(0,1)$. Clearly

$$\begin{aligned}&\{|f\circ w-f\circ w_n|>\epsilon\}\\&\qquad\subset\big(\{|f\circ w-f\circ w_n|>\epsilon\}\cap\{|w-w_n|\leqslant\delta\}\cap\{|w|\leqslant R\}\big)\\&\qquad\quad\cup\{|w-w_n|>\delta\}\cup\{|w|>R\}.\end{aligned}$$

Since $\{|w-w_n|\leqslant\delta\}\cap\{|w|\leqslant R\}\subset\{|w_n|\leqslant R+1\}\cap\{|w|\leqslant R\}$, and since f is uniformly continuous on the closed interval $[-R-1,R+1]$, we can choose $\delta=\delta_\epsilon$ so small that the first set on the right-hand side is empty, i.e.

$$\{|f\circ w-f\circ w_n|>\epsilon\}\subset\{|w-w_n|>\delta\}\cup\{|w|>R\}.$$

This shows that for all $A\in\mathscr{A}$ with $\mu(A)<\infty$

$$\mu\big(\{|f\circ w-f\circ w_n|>\epsilon\}\cap A\big)\leqslant\mu\big(\{|w-w_n|>\delta\}\cap A\big)+\mu\big(\{|w|>R\}\cap A\big).$$

Using $w_n\xrightarrow{\mu}w$, we find

$$\limsup_{n\to\infty}\mu\big(\{|f\circ w-f\circ w_n|>\epsilon\}\cap A\big)\leqslant\mu\big(\{|w|>R\}\cap A\big)\xrightarrow[R\to\infty]{}0,$$

since $\mu\{|w|=\infty\}=0$.

(iv) Observe that

$$w_n\wedge v-w\wedge v=\left\{\begin{array}{ll}w_n-w & \text{if } w_n,w\leqslant v\\ 0 & \text{if } w_n,w\geqslant v\\ w_n-v\leqslant 0 & \text{if } w_n\leqslant v\leqslant w\\ v-w\leqslant 0 & \text{if } w\leqslant v\leqslant w_n\end{array}\right\}\leqslant|w_n-w|.$$

This means that $|w_n\wedge v-w\wedge v|\leqslant|w_n-w|\xrightarrow{\mu}0$, and the claim follows. ∎

The following lemma is a first generalization of Lebesgue's dominated convergence theorem (Theorem 12.2, 13.9) for μ-convergent sequences.

Lemma 22.5 *Let $(u_n)_{n\in\mathbb{N}}\subset\mathcal{M}(\mathscr{A})$ be a sequence with $u_n\xrightarrow{\mu}u$. If $|u|,|u_n|\leqslant w$ for some $w\in\mathcal{L}^p(\mu)$, $p\in[1,\infty)$, then*

$$\lim_{n\to\infty}\|u_n-u\|_p=0\quad\textit{and}\quad\lim_{n\to\infty}\int|u_n|^p\,d\mu=\int|u|^p\,d\mu.$$

Proof From $|u|,|u_n|\leqslant w\in\mathcal{L}^p(\mu)$ we see that $u,u_n\in\mathcal{L}^p(\mu)$. As in the proof of the dominated convergence theorem, the convergence of the integrals follows with Minkowski's inequality from convergence in $\mathcal{L}^p(\mu)$. Moreover, $u_n\xrightarrow{\mu}u$ is the same as $u_n-u\xrightarrow{\mu}0$, i.e. we can assume that $u=0$.

For $\epsilon > 0$ and $R > 0$ we find

$$\int |u_n|^p \, d\mu = \int_{\{|u_n| \leqslant \epsilon\}} |u_n|^p \, d\mu + \int_{\substack{\{|u_n| > \epsilon\} \\ \cap \{w > R\}}} |u_n|^p \, d\mu + \int_{\substack{\{|u_n| > \epsilon\} \\ \cap \{w \leqslant R\}}} |u_n|^p \, d\mu$$

and using the fact that $|u_n| \leqslant w$, we infer

$$\begin{aligned} &= \int_{\substack{\{|u_n| \leqslant \epsilon\} \\ \cap \{|u_n| \leqslant w\}}} |u_n|^p \, d\mu + \int_{\substack{\{|u_n| > \epsilon\} \\ \cap \{w > R\}}} |u_n|^p \, d\mu + \int_{\substack{\{|u_n| > \epsilon\} \\ \cap \{\epsilon < w \leqslant R\}}} |u_n|^p \, d\mu \\ &\leqslant \int \epsilon^p \wedge w^p \, d\mu + \int_{\{w > R\}} w^p \, d\mu + R^p \mu\big(\{|u_n| > \epsilon\} \cap \{w > \epsilon\}\big). \end{aligned}$$

By the Markov inequality (11.4), $\mu\{w > \epsilon\} \leqslant \epsilon^{-p} \int w^p \, d\mu < \infty$; since $u_n \xrightarrow{\mu} 0$, we see that

$$\lim_{n \to \infty} R^p \mu\big(\{|u_n| > \epsilon\} \cap \{w > \epsilon\}\big) = 0.$$

Now we can use dominated convergence, letting $R \to \infty$ and $\epsilon \to 0$, and conclude

$$\begin{aligned} \limsup_{n \to \infty} \int |u_n|^p \, d\mu &\leqslant \int \epsilon^p \wedge w^p \, d\mu + \int_{\{w > R\}} w^p \, d\mu \\ &\xrightarrow[R \to \infty]{\text{dom. conv.}} \int \epsilon^p \wedge w^p \, d\mu \xrightarrow[\epsilon \to 0]{\text{dom. conv.}} 0. \end{aligned}$$

□

Lemma 22.6 *Assume that $(X, \mathscr{A}, \mu)$ is σ-finite and $(u_n)_{n \in \mathbb{N}} \subset \mathcal{M}(\mathscr{A})$ converges in measure to u. Then u is a.e. unique.*

Proof Let $(A_k)_{k \in \mathbb{N}} \subset \mathscr{A}$ be a sequence with $A_k \uparrow X$ and $\mu(A_k) < \infty$. Suppose that u and w are two measurable functions such that $u_n \xrightarrow{\mu} u$ and $u_n \xrightarrow{\mu} w$. Because of $|u - w| \leqslant |u - u_n| + |u_n - w|$ we find for $n \in \mathbb{N}$ and $\epsilon > 0$ that

$$\{|u - w| > 2\epsilon\} \subset \{|u - u_n| > \epsilon\} \cup \{|u_n - w| > \epsilon\}.$$

Therefore,

$$\begin{aligned} &\mu\big(A_k \cap \{|u - w| > 2\epsilon\}\big) \\ &\qquad \leqslant \mu\left(A_k \cap \{|u - u_n| > \epsilon\}\right) + \mu\left(A_k \cap \{|u_n - w| > \epsilon\}\right) \xrightarrow[n \to \infty]{} 0 \end{aligned}$$

holds for all $k \in \mathbb{N}$, $\epsilon > 0$, i.e. $A_k \cap \{|u - w| > 2\epsilon\}$ is a null set for all $k \in \mathbb{N}$, $\epsilon > 0$; but then $\{u \neq w\} \subset \bigcup_{\epsilon \in \mathbb{Q}^+} \{|u - w| > \epsilon\} = \bigcup_{k \in \mathbb{N}, \epsilon \in \mathbb{Q}^+} \left(A_k \cap \{|u - w| > \epsilon\}\right)$ is also a null set, and we are done. □

Caution Limits in measure on a non-σ-finite measure space $(X, \mathscr{A}, \mu)$ need not be unique, see Problem 22.6.

We are now ready for the main result of this chapter, which generalizes the dominated convergence theorem (Theorems 12.2 and 13.9).

Theorem 22.7 (Vitali) *Let $(X, \mathscr{A}, \mu)$ be σ-finite and $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$, $p \in [1, \infty)$, be a sequence which converges in measure to some measurable function $u \in \mathcal{M}(\mathscr{A})$. Then the following assertions are equivalent:*

(i) $\lim_{n\to\infty} \|u_n - u\|_p = 0$;

(ii) $(|u_n|^p)_{n\in\mathbb{N}}$ *is a uniformly integrable family;*

(iii) $\lim_{n\to\infty} \int |u_n|^p \, d\mu = \int |u|^p \, d\mu < \infty$.

Proof (iii) $\Rightarrow$ (ii). Fix $\epsilon \in (0,1)$. We have

$$\begin{aligned}\int_{\{\epsilon|u_n|>|u|\}} |u_n|^p \, d\mu &= \int_{\{\epsilon|u_n|>|u|\}} |u|^p \, d\mu + \int_{\{\epsilon|u_n|>|u|\}} \big(|u_n|^p - |u|^p\big) d\mu \\ &= \int_{\{\epsilon|u_n|>|u|\}} |u|^p \, d\mu - \int_{\{\epsilon|u_n|\leqslant|u|\}} \big(|u_n|^p - |u|^p\big) d\mu + \int \big(|u_n|^p - |u|^p\big) d\mu.\end{aligned}$$

Denote the three integrals by I_1, I_2 and I_3, respectively. Because of assumption (iii) we know

$$\exists N_\epsilon \in \mathbb{N} \quad \forall n \geqslant N_\epsilon \,:\, I_3 \leqslant \epsilon^p.$$

Moreover, (iii) shows that $\sup_{n\in\mathbb{N}} \int |u_n|^p \, d\mu \leqslant C < \infty$ which means that

$$I_1 \leqslant \int_{\{\epsilon|u_n|>|u|\}} |u|^p \, d\mu \leqslant \epsilon^p \int_{\{\epsilon|u_n|>|u|\}} |u_n|^p \, d\mu \leqslant \epsilon^p \int |u_n|^p \, d\mu \leqslant C\epsilon^p.$$

Finally, if $\epsilon < 1$ we can rewrite I_2 in the following way:

$$\begin{aligned}I_2 = \int_{\{\epsilon|u_n|\leqslant|u|\}} \big(|u_n|^p - |u|^p\big) d\mu &= \int_{\{\epsilon|u_n|\leqslant|u|\}} \left(|u_n|^p \wedge \frac{|u|^p}{\epsilon^p} - |u|^p \wedge \frac{|u|^p}{\epsilon^p}\right) d\mu \\ &\leqslant \int \left| |u_n|^p \wedge \frac{|u|^p}{\epsilon^p} - |u|^p \wedge \frac{|u|^p}{\epsilon^p} \right| d\mu.\end{aligned}$$

By Lemma 22.4(iii), (iv) $|u_n|^p \wedge |u|^p/\epsilon^p \xrightarrow{\mu} |u|^p \wedge |u|^p/\epsilon^p$, and, since all functions are dominated by an integrable function, we conclude with Lemma 22.5 for $p = 1$ that

$$\exists M_\epsilon \geqslant N_\epsilon \quad \forall n \geqslant M_\epsilon \,:\, I_2 \leqslant \epsilon^p.$$

Setting $w_\epsilon := \epsilon^{-1}\max\{|u_1|,\dots,|u_{M_\epsilon}|,|u|\}$ we have $w_\epsilon \in \mathcal{L}^p(\mu)$ [✍] and

$$\{|u_n| > w_\epsilon\} \left\langle \begin{array}{ll} = \emptyset & \forall n \leqslant M_\epsilon, \\ \subset \{\epsilon|u_n| > |u|\} & \forall n \in \mathbb{N}. \end{array}\right.$$

Consequently,

$$\sup_{n\in\mathbb{N}} \int_{\{|u_n|>w_\epsilon\}} |u_n|^p \, d\mu \leqslant \sup_{n\geqslant M_\epsilon} \int_{\{\epsilon|u_n|>|u|\}} |u_n|^p d\mu \leqslant (C+2)\epsilon^p.$$

Since

$$w_\epsilon \in \mathcal{L}^p(\mu) \Leftrightarrow w_\epsilon^p \in \mathcal{L}^1(\mu) \quad \text{and} \quad \{|u_n| > w_\epsilon\} = \{|u_n|^p > w_\epsilon^p\}, \tag{22.3}$$

we have established the uniform integrability of $(|u_n|^p)_{n\in\mathbb{N}}$.

(ii) $\Rightarrow$ (i). Let us first check that the double sequence $(|u_n - u_k|^p)_{k,n\in\mathbb{N}}$ is again uniformly integrable. In view of (22.3), our assumption reads

$$\int_{\{|u_n|>w\}} |u_n|^p d\mu < \epsilon \quad \forall n \in \mathbb{N} \tag{22.4}$$

for some suitable $w = w_\epsilon \in \mathcal{L}^p_+(\mu)$. From $|a-b| \leqslant |a| + |b| \leqslant 2\max\{|a|,|b|\}$ we deduce

$$\int_{\{|u_n-u_k|>2w\}} |u_n - u_k|^p d\mu \leqslant 2^p \int_{\{|u_n-u_k|>2w\}} (|u_n| \vee |u_k|)^p \, d\mu,$$

and since $|u_n - u_k| \leqslant |u_n| + |u_k|$ we get

$$\{|u_n - u_k| > 2w\} \subset \{|u_n| > w\} \cup \{|u_k| > w\}.$$

Consequently,

$$\begin{aligned} &\int_{\{|u_n-u_k|>2w\}} |u_n - u_k|^p \, d\mu \\ &\qquad \leqslant \ 2^p \left\{ \int_{\{|u_n|>w\}\cap\{|u_k|>w\}} + \int_{\{|u_n|>w\geqslant|u_k|\}} + \int_{\{|u_k|>w\geqslant|u_n|\}} \right\} |u_n|^p \vee |u_k|^p d\mu \\ &\qquad \leqslant \ 2^p \left\{ \int_{\{|u_n|>w\}\cap\{|u_k|>w\}} |u_n|^p \, d\mu + \int_{\{|u_n|>w\}\cap\{|u_k|>w\}} |u_k|^p \, d\mu \right\} \\ &\qquad\qquad + 2^p \int_{\{|u_n|>w\}} |u_n|^p \, d\mu + 2^p \int_{\{|u_k|>w\}} |u_k|^p \, d\mu \\ &\qquad \overset{(22.4)}{\leqslant} \ 4 \cdot 2^p \epsilon \ = \ 2^{p+2}\epsilon. \end{aligned}$$

From this we conclude that for $W = 2w \in \mathcal{L}^p_+(\mu)$

$$\int |u_n - u_k|^p \, d\mu = \int_{\{|u_n - u_k| > W\}} |u_n - u_k|^p \, d\mu + \int_{\{|u_n - u_k| \leqslant W\}} |u_n - u_k|^p \, d\mu$$
$$\leqslant 2^{p+2}\epsilon + \int |u_n - u_k|^p \wedge W^p \, d\mu.$$

Since $(u_n - u_k) \wedge W \xrightarrow{\mu} 0$ [✍], we can use Lemma 22.5 and get

$$\limsup_{k,n\to\infty} \int |u_n - u_k|^p \, d\mu \leqslant 2^{p+2}\epsilon \xrightarrow[\epsilon\to 0]{} 0.$$

Since $\mathcal{L}^p(\mu)$ is complete (see Theorem 13.7), $(u_n)_{n\in\mathbb{N}}$ converges in $\mathcal{L}^p(\mu)$ to a limit $\tilde{u} \in \mathcal{L}^p(\mu)$. Because of Lemma 22.4, $\mathcal{L}^p$-convergence also implies $u_n \xrightarrow{\mu} \tilde{u}$ and, by Lemma 22.6, we have $u = \tilde{u}$ a.e., hence $\mathcal{L}^p\text{-}\lim_{n\to\infty} u_n = u$.

(i) $\Rightarrow$ (iii) is a consequence of the lower triangle inequality for the $\mathcal{L}^p$-norm, see the first two lines of the proof of Theorem 13.10 ☐

Remark 22.8 Vitali's theorem still holds for measure spaces $(X, \mathscr{A}, \mu)$ which are not σ-finite. In this case, however, we can no longer identify the $\mathcal{L}^p(\mu)$-limit and the theorem reads as follows. *If $(u_n)_{n\in\mathbb{N}}$ converges in measure, then the following are equivalent:*

(i) $(u_n)_{n\in\mathbb{N}}$ *converges in* $\mathcal{L}^p(\mu)$;

(ii) $(|u_n|^p)_{n\in\mathbb{N}}$ *is uniformly integrable*;

(iii) $(\|u_n\|_p)_{n\in\mathbb{N}}$ *converges in* $\mathbb{R}$.

The reason for this is evident from the proof of Theorem 22.7: the last few lines of the step (ii) $\Rightarrow$ (i) require σ-finiteness of $(X, \mathscr{A}, \mu)$.

Different Forms of Uniform Integrability

In view of Vitali's convergence theorem one is led to suspect that uniform integrability is essentially a sufficient (and also necessary, if $(X, \mathscr{A}, \mu)$ is σ-finite) condition for *weak sequential relative compactness* in $\mathcal{L}^1(\mu)$, i.e.

$$(22.1) \implies \begin{cases} \text{every } (u_n)_{n\in\mathbb{N}} \subset \mathcal{F} \text{ has a subsequence } (u_{n(i)})_{i\in\mathbb{N}} \text{ such} \\ \text{that } \lim\limits_{i\to\infty} \int u_{n(i)} \cdot \phi \, d\mu \text{ exists for all } \phi \in \mathcal{L}^\infty(\mu), \end{cases}$$

see Dunford and Schwartz [15, pp. 291–292, 386–387]. In $\mathcal{L}^p(\mu)$, $1<p<\infty$, uniform boundedness of $\mathcal{F}\subset\mathcal{L}^p(\mu)$ suffices for this:

$$\sup_{u\in\mathcal{F}}\|u\|_p<\infty \iff \begin{cases}\text{every } (u_n)_{n\in\mathbb{N}}\subset\mathcal{F} \text{ has a subsequence } (u_{n(i)})_{i\in\mathbb{N}} \text{ such}\\ \text{that } \lim\limits_{i\to\infty}\int u_{n(i)}\cdot\phi\,d\mu \text{ exists for all } \phi\in\mathcal{L}^q(\mu), \text{ where}\\ p,q\in(1,\infty) \text{ are conjugate: } \frac{1}{p}+\frac{1}{q}=1.\end{cases}$$

This is a consequence of the reflexivity of the spaces $\mathcal{L}^p(\mu)$, $p>1$.

Let us discuss various equivalent conditions for uniform integrability.

Theorem 22.9 *Let $(X,\mathscr{A},\mu)$ be some measure space and $\mathcal{F}\subset\mathcal{L}^1(\mathscr{A})$. Then the following statements* (i)–(v) *are equivalent:*

(i) *$\mathcal{F}$ is uniformly integrable, i.e.* (22.1) *holds;*

(ii) (a) $\sup\limits_{u\in\mathcal{F}}\int|u|d\mu<\infty$;

(b) $\forall\epsilon>0\quad\exists w_\epsilon\in\mathcal{L}^1_+(\mathscr{A}),\ \delta>0\quad\forall B\in\mathscr{A},$
$$\int_B w_\epsilon\,d\mu<\delta\ :\ \sup_{u\in\mathcal{F}}\int_B|u|\,d\mu<\epsilon;$$

(iii) (a) $\sup\limits_{u\in\mathcal{F}}\int|u|d\mu<\infty$;

(b) $\forall\epsilon>0\quad\exists K_\epsilon\in\mathscr{A},\ \mu(K_\epsilon)<\infty\ :\ \sup\limits_{u\in\mathcal{F}}\int_{K_\epsilon^c}|u|d\mu<\epsilon$;

(c) $\forall\epsilon>0\quad\exists\delta>0\quad\forall B\in\mathscr{A},\ \mu(B)<\delta\ :\ \sup\limits_{u\in\mathcal{F}}\left|\int_B u\,d\mu\right|<\epsilon$;

(iv) (a) $\sup\limits_{u\in\mathcal{F}}\int|u|d\mu<\infty$;

(b) $\forall\epsilon>0\quad\exists K_\epsilon\in\mathscr{A},\ \mu(K_\epsilon)<\infty\ :\ \sup\limits_{u\in\mathcal{F}}\int_{K_\epsilon^c}|u|d\mu<\epsilon$;

(c) $\forall\epsilon>0\quad\exists\delta>0\quad\forall B\in\mathscr{A},\ \mu(B)<\delta\ :\ \sup\limits_{u\in\mathcal{F}}\int_B|u|d\mu<\epsilon$;

(v) (a) $\forall\epsilon>0\quad\exists K_\epsilon\in\mathscr{A},\ \mu(K_\epsilon)<\infty\ :\ \sup\limits_{u\in\mathcal{F}}\int_{K_\epsilon^c}|u|d\mu<\epsilon$;

(b) $\lim\limits_{R\to\infty}\sup\limits_{u\in\mathcal{F}}\int_{\{|u|>R\}}|u|d\mu=0.$

*If $(X,\mathscr{A},\mu)$ is a **σ-finite** measure space,* (i)–(v) *are also equivalent to*

(vi) (a) $\sup\limits_{u\in\mathcal{F}}\int|u|d\mu<\infty$;

(b) $\lim_{n\to\infty}\sup_{u\in\mathcal{F}}\int_{A_n}|u|\,d\mu=0$ *for all decreasing sequences* $(A_n)_{n\in\mathbb{N}}\subset\mathscr{A}$, $A_n\downarrow\emptyset$.
[*Note: it is **not** assumed that* $\mu(A_n)<\infty$.]

(vii) (a) $\sup_{u\in\mathcal{F}}\int|u|\,d\mu<\infty$;

(b) $\lim_{n\to\infty}\sup_{u\in\mathcal{F}}\int_{A_n}u\,d\mu=0$ *for all decreasing sequences* $(A_n)_{n\in\mathbb{N}}\subset\mathscr{A}$, $A_n\downarrow\emptyset$.
[*Note: it is **not** assumed that* $\mu(A_n)<\infty$.]

(viii) (a) $\sup_{u\in\mathcal{F}}\int|u|\,d\mu<\infty$;

(b) $\lim_{n\to\infty}\sup_{u\in\mathcal{F}}\int_{A_n}u\,d\mu=0$ *for all decreasing sequences* $(A_n)_{n\in\mathbb{N}}\subset\mathscr{A}$, $A_n\downarrow\emptyset$;
[*Note: it is **not** assumed that* $\mu(A_n)<\infty$.]

(c) $\lim_{n\to\infty}\sup_{u\in\mathcal{F}}\left|\int_{B_n}u\,d\mu\right|=0$ *for all sequences* $(B_n)_{n\in\mathbb{N}}\subset\mathscr{A}$ *with* $\mu(B_n)\to 0$.

If $(X,\mathscr{A},\mu)$ *is a **finite** measure space,* (i)–(viii) *are also equivalent to*

(ix) $\lim_{R\to\infty}\sup_{u\in\mathcal{F}}\int_{\{|u|>R\}}|u|\,d\mu=0$;

(x) $\sup_{u\in\mathcal{F}}\int\Phi(|u|)\,d\mu<\infty$ *for some increasing, convex function*
$\Phi:[0,\infty)\to[0,\infty)$ *such that* $\lim_{t\to\infty}\dfrac{\Phi(t)}{t}=\infty$.

Remark 22.10 Almost any combination of the above criteria appears in the literature as uniform integrability or under different names. Here is a short list:

(ii)(a)	uniform boundedness
(iii)(b)	tightness
(iii)(c), (iv)(c)	uniform absolute continuity
(vi)(b), (vii)(b)	uniform σ-additivity
(iii)	Dieudonné's condition (weak sequential relative compactness)
(vii), (viii)	Dunford–Pettis condition (weak sequential relative compactness)
(ix)	uniform integrability (mainly in probability theory)
(x)	de la Vallée Poussin's condition

Note that condition (iii)(c) is often written as '$\lim_{\mu(B)\to 0}\left|\int_B u\,d\mu\right|=0$ uniformly in $u\in\mathcal{F}$'; a similar formulation is used instead of (iv)(c).

Proof of Theorem 22.9 First we show (v) $\Rightarrow$ (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) $\Rightarrow$ (v) for general measure spaces, then (ii) $\Rightarrow$ (vi) $\Rightarrow$ (vii) $\Rightarrow$ (viii) $\Rightarrow$ (iii) for σ-finite measure spaces and, finally, for finite measure spaces (v) $\Rightarrow$ (ix) $\Rightarrow$ (x) $\Rightarrow$ (i).

(v) $\Rightarrow$ (iv). Condition (iii)(b) is clear. Given $\epsilon > 0$ we can pick $K = K_{\epsilon/2} \in \mathscr{A}$ and $R = R_{\epsilon/2} > 0$ such that

$$\begin{aligned}\int |u| d\mu &= \int_{K \cap \{|u| > R\}} |u| d\mu + \int_{K \cap \{|u| \leqslant R\}} |u| d\mu + \int_{K^c} |u| d\mu \\ &\leqslant \frac{1}{2}\epsilon + R\mu(K) + \frac{1}{2}\epsilon < \infty,\end{aligned}$$

uniformly for all $u \in \mathcal{F}$. On setting $\delta := \epsilon/(2R)$ we see for every $B \in \mathscr{A}$ with $\mu(B) < \delta$ that

$$\begin{aligned}\int_B |u| d\mu &= \int_{B \cap \{|u| > R\}} |u| d\mu + \int_{B \cap \{|u| \leqslant R\}} |u| d\mu \\ &\leqslant \int_{\{|u| > R\}} |u| d\mu + R\mu(B) \leqslant \frac{1}{2}\epsilon + R\delta = \epsilon,\end{aligned}$$

and (iv) follows.

(iv) $\Rightarrow$ (iii) is obvious.

(iii) $\Rightarrow$ (ii). Condition (ii)(a) is clear. Given $\epsilon > 0$, we pick $K = K_\epsilon \in \mathscr{A}$ with $\mu(K) < \infty$ and $\delta = \delta_\epsilon > 0$. We then set $w_\epsilon := \mathbb{1}_{K_\epsilon}$. If $B \in \mathscr{A}$ is such that $\mu(B \cap K_\epsilon) = \int_B w_\epsilon \, d\mu < \delta$, then the sets $B^\pm := B \cap \{\pm u > 0\}$ will satisfy $\mu(B^\pm \cap K_\epsilon) < \delta$. Hence,

$$\left| \int_{B^\pm} u \, d\mu \right| \leqslant \left| \int_{B^\pm \cap K_\epsilon} u \, d\mu \right| + \int_{K_\epsilon^c} |u| d\mu \leqslant 2\epsilon,$$

and we conclude from this that

$$\int_B |u| d\mu = \int_{B^+} u \, d\mu - \int_{B^-} (-u) d\mu \leqslant 4\epsilon.$$

Since δ does not depend on u, we get

$$\sup_u \int_B |u| d\mu \leqslant 4\epsilon \quad \text{for all } B \in \mathscr{A}, \ \mu(B) < \delta.$$

(ii) $\Rightarrow$ (i). Take $w = w_\epsilon$ and $\delta = \delta_\epsilon > 0$ as in (ii). If

$$R > \frac{1}{\delta} \sup_{u \in \mathcal{F}} \int |u| d\mu$$

we see

$$\int |u| d\mu \geqslant \int_{\{|u| > Rw\}} |u| d\mu \geqslant R \int_{\{|u| > Rw\}} w \, d\mu,$$

and so

$$\int_{\{|u| > Rw\}} w \, d\mu \leqslant \frac{1}{R} \sup_{u \in \mathcal{F}} \int |u| d\mu < \delta.$$

From (ii)(b) we infer that $\sup_{u\in\mathcal{F}}\int_{\{|u|>Rw\}}|u|\,d\mu\leqslant\epsilon$.

(i)$\Rightarrow$(v). Let $w=w_\epsilon$ be as in (i), resp. (22.1). Since it then holds that $\{|u|\leqslant w\}\cap\{|u|\geqslant R\}\subset\{w\geqslant R\}$, we have

$$\begin{aligned}\int_{\{|u|>R\}}|u|\,d\mu &= \int_{\{|u|>w\}\cap\{|u|>R\}}|u|\,d\mu+\int_{\{|u|\leqslant w\}\cap\{|u|>R\}}|u|\,d\mu\\ &\leqslant \int_{\{|u|>w\}}|u|\,d\mu+\int_{\{|w|>R\}}w\,d\mu\\ &\leqslant \epsilon+\int w\mathbb{1}_{\{|w|>R\}}\,d\mu. \end{aligned}\tag{22.5}$$

From the dominated convergence theorem (Theorem 12.2) we see that the right-hand side tends (uniformly for all $u\in\mathcal{F}$) to ϵ as $R\to\infty$ and (iv)(b) follows. To see (iv)(a) we choose $r=r_\epsilon>0$ so small that $\int_{\{w\leqslant r\}}w\,d\mu\leqslant\int w\wedge r\,d\mu\leqslant\epsilon$; this is possible since by the dominated convergence theorem $\lim_{r\to 0}\int|w|\wedge r\,d\mu=0$. By the Markov inequality (11.4) we see that $\mu\{w>r\}\leqslant(1/r)\int w\,d\mu<\infty$, and we get for $K:=\{w>r\}$

$$\begin{aligned}\sup_{u\in\mathcal{F}}\int_{K^c}|u|\,d\mu &= \sup_{u\in\mathcal{F}}\left(\int_{\{w\leqslant r\}\cap\{|u|>w\}}|u|\,d\mu+\int_{\{w\leqslant r\}\cap\{|u|\leqslant w\}}|u|\,d\mu\right)\\ &\overset{(22.1)}{\leqslant} \epsilon+\sup_{u\in\mathcal{F}}\int_{\{w\leqslant r\}\cap\{|u|\leqslant w\}}|u|\,d\mu\\ &\leqslant \epsilon+\int_{\{w\leqslant r\}}w\,d\mu\leqslant 2\epsilon.\end{aligned}$$

This proves (v).

Assume for the rest of the proof that μ is σ-finite

(ii)$\Rightarrow$(vi). (v)(a) is clear. If $A_n\downarrow\emptyset$ we see from the monotone convergence theorem, Theorem 12.1, that $\lim_{n\to\infty}\int_{A_n}w\,d\mu=0$, so that we have because of (ii)(b) $\sup_{u\in\mathcal{F}}\int_{A_n}|u|\,d\mu<\epsilon$ for sufficiently large $n\in\mathbb{N}$.

(vi)$\Rightarrow$(vii) is obvious.

(vii)$\Rightarrow$(viii).[2] (viii)(a) and (viii)(b) are the same as (vii)(a) and (vii)(b). If (viii)(c) fails to hold, then there exists some $\epsilon>0$ such that

$$\forall\delta>0\quad\exists B\in A,\ \mu(B)<\delta,\ u\in\mathcal{F}:\left|\int_B u\,d\mu\right|\geqslant\epsilon.\tag{22.6}$$

[2] I owe this argument to Ms. Franziska Kühn.

Set $\epsilon_k := \epsilon 4^{-k}$. Recursively we construct sets $(B_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ and $(u_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ such that

$$\max_{k\leqslant n-1} \int_{B_n} |u_k|\,d\mu \leqslant \epsilon_n \quad \text{and} \quad \left|\int_{B_n} u_n\,d\mu\right| \geqslant \epsilon.$$

For $n=1$, it is clear how to choose B_1 and u_1. Suppose that we have already constructed $B_1, \dots, B_{n-1}$ and $u_1, \dots, u_{n-1}$; there exists some $\delta \in (0, 1/n)$ such that

$$\max_{k\leqslant n-1} \int_B |u_k|\,d\mu \leqslant \sum_{k=1}^{n-1} \int_B |u_k|\,d\mu \leqslant \epsilon_n \quad \text{for all } B \in \mathscr{A},\ \mu(B) \leqslant \delta.$$

Because of (22.6) there is some $B_n \in \mathscr{A}$, $\mu(B_n) \leqslant \delta$, such that $\left|\int_{B_n} u_n\,d\mu\right| \geqslant \epsilon$.

Set $B := \bigcap_{n=1}^\infty B_n$ and $A_n := \bigcup_{k=n}^\infty (B_k \setminus B) \downarrow \emptyset$. As $\mu(B) \leqslant \mu(B_n) \leqslant n^{-1} \to 0$, we conclude that

$$\begin{aligned}
\left|\int_{A_n} u_n\,d\mu\right| &= \left|\int_{B_n\setminus B} u_n\,d\mu + \int_{A_{n+1}\cap(B_n\setminus B)^c} u_n\,d\mu\right| \\
&\geqslant \left|\int_{B_n\setminus B} u_n\,d\mu\right| - \int_{A_{n+1}} |u_n|\,d\mu \\
&\geqslant \left|\int_{B_n} u_n\,d\mu\right| - \sum_{k=n+1}^\infty \int_{B_k} |u_n|\,d\mu \\
&\geqslant \epsilon - \epsilon \sum_{k=n+1}^\infty 4^{-k} \geqslant \frac{\epsilon}{4},
\end{aligned}$$

contradicting our assumption (vii)(b).

(viii) $\Rightarrow$ (iii).[3] Since $(X, \mathscr{A}, \mu)$ is σ-finite, there is a sequence $(K_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ such that $K_n \uparrow X$ and $\mu(K_n) < \infty$. The condition (iii)(a) is the same as (viii)(a).

Suppose that (iii)(b) does not hold. This means that there is some $\epsilon > 0$ such that for any set $K \in \mathscr{A}$, $\mu(K) < \infty$, there exists a function $u \in \mathcal{F}$ such that $\int_{K^c} |u|\,d\mu \geqslant \epsilon$. In particular, we may use $K = K_1$ and find some $u_1 \in \mathcal{F}$ satisfying

$$\int_{K_1^c} |u_1|\,d\mu \geqslant \epsilon.$$

[3] I owe this argument to Ms. Franziska Kühn.

Choose $n(1) \in \mathbb{N}$ so large, that $\int_{K^c_{n(1)}} |u_1| d\mu \leqslant \epsilon/8$. Recursively we find $u_{k+1} \in \mathcal{F}$ and $n(k+1) > n(k)$ such that

$$\int_{K^c_{n(k)}} |u_{k+1}| d\mu \geqslant \epsilon \quad \text{and} \quad \int_{K^c_{n(k+1)}} |u_{k+1}| d\mu \leqslant \frac{1}{8}\epsilon.$$

Since $|u_{k+1}| = u^+_{k+1} + u^-_{k+1}$, this means that

$$\int_{K^c_{n(k)}} u^+_{k+1}\, d\mu \geqslant \frac{1}{2}\epsilon \quad \text{or} \quad \int_{K^c_{n(k)}} (-u^-_{k+1}) d\mu \geqslant \frac{1}{2}\epsilon$$

for all $k \in \mathbb{N}$. At least one of these two inequalities must hold for infinitely many $k \in \mathbb{N}$; without loss of generality we may assume that it is the first relation. Define $A_k := K^c_{n(k)} \cup \big(\{u_k > 0\} \cap (K_{n(k)} \setminus K_{n(k-1)})\big)$ and observe that we have $A_k \subset K^c_{n(k-1)} \subset A_{k-1}$, i.e. A_k is a decreasing sequence. Moreover,

$$\begin{aligned}
\int_{A_k} u_k\, d\mu &= \int_{\{u_k>0\} \cap (K_{n(k)} \setminus K_{n(k-1)})} u_k\, d\mu + \int_{K^c_{n(k)}} u_k\, d\mu \\
&\geqslant \int_{\{u_k>0\} \cap K^c_{n(k-1)}} u_k - 2\int_{K^c_{n(k)}} |u_k| d\mu \\
&= \int_{K^c_{n(k-1)}} u^+_k\, d\mu - 2\int_{K^c_{n(k)}} |u_k| d\mu \\
&\geqslant \frac{1}{2}\epsilon - 2\frac{1}{8}\epsilon = \frac{1}{4}\epsilon
\end{aligned}$$

for infinitely many $k \in \mathbb{N}$. Hence, $\liminf_{n\to\infty} \left| \int_{A_k} u_k\, d\mu \right| \geqslant \epsilon/8$, contradicting (viii)(b).

Finally, (iii)(c) is an immediate consequence of (viii)(c).

Assume for the rest of the proof that μ is finite

(v) $\Rightarrow$ (ix) is trivial.

(ix) $\Rightarrow$ (x). For $u \in \mathcal{F}$ we set $\alpha_n := \alpha_n(u) := \mu\{|u| > n\}$ and define

$$\Phi(t) := \int_{[0,t)} \phi(s)\lambda(ds), \qquad \phi(s) := \sum_{n=1}^{\infty} \gamma_n \mathbb{1}_{[n,n+1)}(s).$$

We will now determine the numbers $\gamma_1, \gamma_2, \gamma_3, \ldots$. Clearly,

$$\Phi(t) = \sum_{n=1}^{\infty} \gamma_n \int_{[0,t)} \mathbb{1}_{[n,n+1)}(s)\lambda(ds) = \sum_{n=1}^{\infty} \gamma_n[(t-n)^+ \wedge 1]$$

and

$$\int \Phi(|u|)d\mu = \sum_{n=1}^{\infty} \gamma_n \int \left[(|u|-n)^+ \wedge 1\right] d\mu \leqslant \sum_{n=1}^{\infty} \gamma_n \mu\{|u|>n\}. \tag{22.7}$$

If we can construct $(\gamma_n)_{n\in\mathbb{N}} \subset [0,\infty)$ such that it increases to ∞ and (22.7) is finite (uniformly for all $u \in \mathcal{F}$), then we are done: $\phi(s)$ will increase to ∞, $\Phi(t)$ will be convex[4] and satisfy

$$\frac{\Phi(t)}{t} = \frac{1}{t}\int_{[0,t)} \phi(s)\lambda(ds) \geqslant \frac{1}{t}\int_{[t/2,t)} \phi(s)\lambda(ds) \geqslant \tfrac{1}{2}\phi\left(\tfrac{t}{2}\right) \uparrow \infty.$$

By assumption there is an increasing sequence $(r_i)_{i\in\mathbb{N}} \subset \mathbb{R}$ with $r_i \to \infty$ and $\int_{\{r_i<|u|\}} |u| d\mu \leqslant 2^{-i}$. Now

$$\sum_{n=r_i}^{\infty} \mathbb{1}_{\{n<|u|\}} \leqslant \sum_{n=r_i}^{|u|} \mathbb{1}_{\{n<|u|\}} \leqslant \sum_{n=r_i}^{|u|} \mathbb{1}_{\{r_i<|u|\}} \leqslant |u|\mathbb{1}_{\{r_i<|u|\}}.$$

If we integrate both sides of this inequality and use Beppo Levi's theorem in the form of Corollary 9.9 on the left-hand side, we get

$$\sum_{n=r_i}^{\infty} \mu\{|u|>n\} = \sum_{n=r_i}^{\infty} \int \mathbb{1}_{\{n<|u|\}}\, d\mu \leqslant \int |u|\mathbb{1}_{\{r_i<|u|\}}\, d\mu \leqslant 2^{-i}.$$

Summing over $i \in \mathbb{N}$ finally yields

$$\sum_{i=1}^{\infty}\sum_{n=r_i}^{\infty} \mu\{|u|>n\} \leqslant \sum_{i=1}^{\infty} 2^{-i} = 1,$$

[4] Usually one argues that $\Phi'' \geqslant 0$ a.e., but for this we need to know that the monotone function $\phi = \Phi'$ is almost everywhere differentiable – and this requires Lebesgue's differentiation theorem (Theorem 25.20). Here is an alternative argument: it is not hard to see that $\Phi : (a,b) \to \mathbb{R}$ is convex if, and only, if

$$\frac{\Phi(y)-\Phi(x)}{y-x} \leqslant \frac{\Phi(z)-\Phi(x)}{z-x}$$

holds for all $a<x<y<z<b$; use e.g. the technique of the proof of Lemma 13.12. Since $\Phi(x) = \int_0^x \phi(s)ds$ (by Lemma 14.14 and Theorem 12.9), this is the same as

$$\frac{1}{y-x}\int_x^y \phi(s)ds \leqslant \frac{1}{z-x}\int_x^z \phi(s)ds \iff \frac{1}{y-x}\int_x^y \phi(s)ds \leqslant \frac{1}{z-y}\int_y^z \phi(s)ds$$
$$\iff \int_0^1 \phi(s(y-x)+x)\, ds \leqslant \int_0^1 \phi(s(z-y)+y)\, ds.$$

The latter inequality follows from the fact that ϕ is increasing and $s(y-x)+x \in [x,y]$ while $s(z-y)+y \in [y,z]$ for $0 \leqslant s \leqslant 1$.

and a simple interchange in the order of the summation gives

$$\sum_{i=1}^{\infty}\sum_{n=r_i}^{\infty}\mu\{|u|>n\}=\sum_{n=1}^{\infty}\underbrace{\left(\sum_{i=1}^{\infty}\mathbb{1}_{[1,n]}(r_i)\right)}_{=:\,\gamma_n}\mu\{|u|>n\}\leqslant 1.$$

This finishes the construction of the sequence $(\gamma_n)_{n\in\mathbb{N}}$.

(x) $\Rightarrow$ (i). Since $\mu(X)<\infty$, the constants are integrable and we may take $w_\epsilon(x)\equiv r_\epsilon$. Fix $\epsilon>0$ and choose r_ϵ so big that $t^{-1}\Phi(t)>1/\epsilon$ for all $t>r_\epsilon$. Thus,

$$\int_{\{|u|>r_\epsilon\}}|u|d\mu\leqslant\int_{\{|u|>r_\epsilon\}}\epsilon\Phi(|u|)d\mu\leqslant\epsilon\int\Phi(|u|)d\mu,$$

and (i) follows. □

Problems

22.1. Let $(X,\mathscr{A},\mu)$ be a finite measure space and $(u_n)_{n\in\mathbb{N}}\subset\mathcal{M}(\mathscr{A})$. Prove that

$$\lim_{k\to\infty}\mu\Big\{\sup_{n\geqslant k}|u_n|>\epsilon\Big\}=0\quad\forall\epsilon>0\implies u_n\xrightarrow[n\to\infty]{}0\ \text{ a.e.}$$

[Hint: $|u_n|\to 0$ a.e. if, and only if, $\mu\big(\bigcup_{n\geqslant k}\{|u_n|>\epsilon\}\big)$ is small for all $\epsilon>0$ and big $k\geqslant k_\epsilon$.]

22.2. Show that for a sequence $(u_n)_{n\in\mathbb{N}}$ of measurable functions on a finite measure space

$$\lim_{k\to\infty}\mu\Big\{\sup_{n\geqslant k}|u_n|>\epsilon\Big\}=\mu\Big(\limsup_{n\to\infty}\Big\{|u_n|>\epsilon\Big\}\Big)\quad\forall\epsilon>0,$$

and combine this with Problem 22.1 to give a new criterion for a.e. convergence.

22.3. Let $(X,\mathscr{A},\mu)$ be a σ-finite measure space and $(u_n)_{n\in\mathbb{N}}\subset\mathcal{M}(\mathscr{A})$. Show that $u_n\to u$ $(n\to\infty)$ in measure if, and only if, $u_n-u_k\to 0$ $(k,n\to\infty)$ in measure.

22.4. Consider one-dimensional Lebesgue measure λ on $([0,1],\mathscr{B}[0,1])$. Compare the convergence behaviour (a.e., $\mathcal{L}^p$, in measure) of the following sequences:

(i) $f_{i,n}:=n\mathbb{1}_{[(i-1)/n,i/n]}$, $n\in\mathbb{N}$, $1\leqslant i\leqslant n$, run through in lexicographical order;
(ii) $g_n:=n\mathbb{1}_{(0,1/n)}$, $n\in\mathbb{N}$;
(iii) $h_n:=a_n(1-nx)^+$, $n\in\mathbb{N}$, $x\in[0,1]$ and a sequence $(a_n)_{n\in\mathbb{N}}\subset\mathbb{R}^+$.

22.5. Let $(u_n)_{n\in\mathbb{N}}$, $(w_n)_{n\in\mathbb{N}}$ be two sequences of measurable functions on $(X,\mathscr{A},\mu)$. Suppose that $u_n\xrightarrow{\mu}u$ and $w_n\xrightarrow{\mu}w$. Show that au_n+bw_n, $a,b\in\mathbb{R}$, $\max\{u_n,w_n\}$, $\min\{u_n,w_n\}$ and $|u_n|$ converge in measure and find their limits.

22.6. Let $(X,\mathscr{A},\mu)$ be a measure space which is not σ-finite. Construct an example of a sequence $(u_n)_{n\in\mathbb{N}}\subset\mathcal{M}(\mathscr{A})$ which converges in measure but whose limit is not unique. Can this happen in a σ-finite measure space?
[Hint: let $X_{\sigma f}:=\bigcup\{F:\mu(F)<\infty\}$ be the σ-finite part of X. Show that $X\setminus X_{\sigma f}\neq\emptyset$, that every measurable $E\subset X\setminus X_{\sigma f}$ satisfies $\mu(E)=\infty$ and that we can change every limit of $(u_n)_{n\in\mathbb{N}}$ outside $X_{\sigma f}$.]

22.7. (i) Prove, without using Vitali's convergence theorem, the following theorem.
Theorem (bounded convergence). *Let $(X, \mathscr{A}, \mu)$ be a measure space, let $A \in \mathscr{A}$ be a set with $\mu(A) < \infty$ and let $(u_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions. Suppose that all u_n vanish on A^c, that $|u_n| \leqslant C$ for all $n \in \mathbb{N}$ and some constant $C > 0$ and that $u_n \xrightarrow{\mu} u$. Then L^1-$\lim_n u_n = u$.*

(ii) Use one-dimensional Lebesgue measure and the sequence $u_n = \mathbb{1}_{[n,n+1)}$ to show that the assumption $\mu(A) < \infty$ is really needed in (i).

(iii) As L^1-limit the function u is unique but, as we have seen in Problem 22.6, this is not the case for limits in measure. Why does the uniqueness of the limit in (i) not contradict Problem 22.6?

22.8. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. Define for two random variables ξ, η

$$\rho_{\mathbb{P}}(\xi, \eta) := \inf\{\epsilon > 0 : \mathbb{P}\{|\xi - \eta| \geqslant \epsilon\} \leqslant \epsilon\}.$$

(i) $\rho_{\mathbb{P}}$ is a pseudo-metric on the space of random variables $\mathcal{M}(\mathscr{A})$, i.e. it holds that $\rho_{\mathbb{P}}(\xi, \eta) = \rho_{\mathbb{P}}(\eta, \xi)$ and $\rho_{\mathbb{P}}(\xi, \eta) \leqslant \rho_{\mathbb{P}}(\xi, \zeta) + \rho_{\mathbb{P}}(\zeta, \eta)$ (but definiteness $\rho_{\mathbb{P}}(\xi, \eta) = 0 \iff \xi = \eta$ need not hold).

(ii) A sequence $(\xi_n)_{n\in\mathbb{N}} \subset \mathcal{M}(\mathscr{A})$ converges in probability to a random variable ξ if, and only if, $\lim_{n\to\infty} \rho_{\mathbb{P}}(\xi_n, \xi) = 0$.

(iii) $\rho_{\mathbb{P}}$ is a complete pseudo-metric on $\mathcal{M}(\mathscr{A})$, i.e. every $\rho_{\mathbb{P}}$-Cauchy sequence converges in probability to some limit in $\mathcal{M}(\mathscr{A})$.

(iv) Show that

$$g_{\mathbb{P}}(\xi, \eta) := \int \frac{|\xi - \eta|}{1 + |\xi - \eta|}\, d\mathbb{P} \quad \text{and} \quad \delta_{\mathbb{P}}(\xi, \eta) := \int [|\xi - \eta| \wedge 1]\, d\mathbb{P}$$

are pseudo-metrics on $\mathcal{M}(\mathscr{A})$ which have the same Cauchy sequences as $\rho_{\mathbb{P}}$.

22.9. Let $(u_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions on a σ-finite measure space $(X, \mathscr{A}, \mu)$ and assume that $u_n \xrightarrow{\mu} u$.

(i) Show that $\lim_{n\to\infty} \int |u_n - u| \wedge \mathbb{1}_A \, d\mu = 0$ for every $A \in \mathscr{A}$ with finite measure $\mu(A) < \infty$.

(ii) Show that every subsequence $(u'_n)_{n\in\mathbb{N}} \subset (u_n)_{n\in\mathbb{N}}$ contains a further subsequence $(u''_n)_{n\in\mathbb{N}} \subset (u'_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} u''_n = u$ almost everywhere.
[Hint: apply (i) with $A = A_i$ from a sequence $A_i \uparrow X$, $\mu(A_i) < \infty$ and use Corollary 13.8.]

(iii) Use (ii) to give an alternative proof of Lemma 22.5.

22.10. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space. Suppose that $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ satisfies $\lim_{n\to\infty} \mu(A_n) = 0$. Show that $\lim_{n\to\infty} \int_{A_n} u \, d\mu = 0$ for all $u \in \mathcal{L}^1(\mu)$.
[Hint: use Vitali's convergence theorem.]

22.11. Let $(X, \mathscr{A}, \mu)$ be a measure space and $(u_n)_{n\in\mathbb{N}} \subset \mathcal{M}(\mathscr{A})$.

(i) Let $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}$. Show that $\lim_{n\to\infty} x_n = 0$ if, and only if, every subsequence $(x_{n_k})_{k\in\mathbb{N}}$ satisfies $\lim_{k\to\infty} x_{n_k} = 0$.

(ii) Show that $u_n \xrightarrow{\mu} u$ if, and only if, every subsequence $(u_{n_k})_{k\in\mathbb{N}}$ has a sub-subsequence $(\tilde{u}_{n_k})_{k\in\mathbb{N}}$ which converges a.e. to u on every set $A \in \mathscr{A}$ of finite μ-measure.
[Hint: use Lemma 22.4 for necessity. For sufficiency show that $\tilde{u}_{n_k} \to u$ in measure, hence the sequence of reals $\mu(A \cap \{|u_{n_k} - u| > \epsilon\})$ has a subsequence converging to 0; use (i) to conclude that $\mu(A \cap \{|u_n - u| > \epsilon\}) \to 0$.]

(iii) Use part (ii) to show that $u_n \xrightarrow{\mu} u$ entails that $\Phi \circ u_n \xrightarrow{\mu} \Phi \circ u$ for every continuous function $\Phi : \mathbb{R} \to \mathbb{R}$.

22.12. Let $\mathcal{F}$ and $\mathcal{G}$ be two families of uniformly integrable functions on an arbitrary measure space $(X, \mathscr{A}, \mu)$. Show that the following statements hold.

(i) Every finite family $\{f_1, \dots, f_n\} \subset \mathcal{L}^1(\mu)$ is uniformly integrable.
(ii) $\mathcal{F} \cup \{f_1, \dots, f_n\}$, $f_1, \dots, f_n \in \mathcal{L}^1(\mu)$, is uniformly integrable.
(iii) $\mathcal{F} + \mathcal{G} := \{f + g : f \in \mathcal{F},\ g \in \mathcal{G}\}$ is uniformly integrable.
(iv) $\text{c.h.}(\mathcal{F}) := \{tf + (1 - t)\phi : f, \phi \in \mathcal{F},\ 0 \leqslant t \leqslant 1\}$ ('c.h.' stands for convex hull) is uniformly integrable.
(v) The closure of $\text{c.h.}(\mathcal{F})$ in the space $\mathcal{L}^1(\mu)$ is uniformly integrable.

22.13. Assume that $(u_n)_{n\in\mathbb{N}}$ is uniformly integrable. Show that

$$\lim_{k\to\infty} \frac{1}{k} \int \max_{n\leqslant k} u_n \, d\mu = 0.$$

22.14. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. Adapt the proof of Theorem 22.9 to show that a sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mu)$ is uniformly integrable if it is bounded in some space $\mathcal{L}^p(\mathbb{P})$ with $p > 1$, i.e. if $\sup_{n\in\mathbb{N}} \|u_n\|_p < \infty$.

Use Vitali's convergence theorem to construct an example illustrating that $\mathcal{L}^1$-boundedness of $(u_n)_{n\in\mathbb{N}}$ does not guarantee uniform integrability.

22.15. Let $(X, \mathscr{A}, \mu)$ be a finite measure space. Show that $\mathcal{F} \subset \mathcal{L}^1(\mu)$ is uniformly integrable if, and only if, the series $\sum_{n=1}^{\infty} n\mu\{n < |f| \leqslant n + 1\}$ converges uniformly for all $f \in \mathcal{F}$. [Hint: compare (vi) $\Rightarrow$ (vii) of the proof of Theorem 22.9.]

22.16. Let $(f_i)_{i\in I}$ be a family of uniformly integrable functions and let $(u_i)_{i\in I} \subset \mathcal{L}^1(\mu)$ be some further family such that $|u_i| \leqslant |f_i|$ for every $i \in I$. Then $(u_i)_{i\in I}$ is uniformly integrable. In particular, every family of functions $(u_i)_{i\in I}$ with $|u_i| \leqslant g$ for some $g \in \mathcal{L}^1_+(\mu)$ is uniformly integrable.

22.17. Let $(X, \mathscr{A}, \mu)$ be an arbitrary measure space. Show that a family $\mathcal{F} \subset \mathcal{L}^1(\mu)$ is uniformly integrable if, and only if, the following condition holds:

$$\forall \epsilon > 0 \quad \exists g_\epsilon \in \mathcal{L}^1_+(\mu) : \sup_{u\in\mathcal{F}} \int (|u| - g_\epsilon \wedge |u|) d\mu < \epsilon.$$

Give a simplified version of this equivalence for finite measure spaces.

23

Martingales

Martingales are a key tool of modern probability theory, in particular, when it comes to a.e. convergence assertions and related limit theorems. The origins of martingale techniques can be traced back to analysis papers by Kac, Marcinkiewicz, Paley, Steinhaus, Wiener and Zygmund from the early 1930s on independent (or orthogonal) functions and the convergence of certain series of functions, see e.g. the paper by Marcinkiewicz and Zygmund [29] which contains many references. The theory of martingales as we know it now goes back to Doob and most of the material of this and the following chapter can be found in his seminal monograph [13] from 1953.

We want to understand martingales as an analysis tool which will be useful for the study of $\mathcal{L}^p$- and almost everywhere convergence and, in particular, for the further development of measure and integration theory. Our presentation differs somewhat from the standard way to introduce martingales – conditional expectations will be defined later in Chapter 27 – but the results and their proofs are pretty much the usual ones. The other difference is that *we develop the theory for σ-finite measure spaces rather than just for probability spaces*. Those readers who are familiar with martingales and the language of conditional expectations are asked to be patient until Chapter 27, in particular Theorem 27.12, when we catch up with those notions.

Throughout this chapter $(X, \mathscr{A}, \mu)$ is a measure space which admits a *filtration*, i.e. an increasing sequence

$$\mathscr{A}_0 \subset \mathscr{A}_1 \subset \cdots \subset \mathscr{A}_n \subset \cdots \subset \mathscr{A}$$

of sub-σ-algebras of $\mathscr{A}$. If $(X, \mathscr{A}_0, \mu)$ is σ-finite[1] we call $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ a *σ-finite filtered measure space*. This will always be the case from now on. Finally,

[1] That is, there is a sequence $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}_0$ with $A_n \uparrow X$ and $\mu(A_n) < \infty$.

we write $\mathscr{A}_\infty := \sigma(\mathscr{A}_n, n = 0, 1, 2, \dots)$ for the smallest σ-algebra generated by all $\mathscr{A}_n$.

Definition 23.1 Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space. A sequence of $\mathscr{A}$-measurable functions $(u_n)_{n\in\mathbb{N}_0}$ is called a *martingale* (with respect to the filtration $(\mathscr{A}_n)_{n\in\mathbb{N}_0}$), if $u_n \in \mathcal{L}^1(\mathscr{A}_n)$ for each $n \in \mathbb{N}_0$ and

$$\int_A u_{n+1}\, d\mu = \int_A u_n\, d\mu \quad \forall A \in \mathscr{A}_n. \tag{23.1}$$

We say that $(u_n)_{n\in\mathbb{N}_0}$ is a *submartingale* (w.r.t. $(\mathscr{A}_n)_{n\in\mathbb{N}_0}$) if $u_n \in \mathcal{L}^1(\mathscr{A}_n)$ and

$$\int_A u_{n+1}\, d\mu \geqslant \int_A u_n\, d\mu \quad \forall A \in \mathscr{A}_n, \tag{23.2}$$

and a *supermartingale* (w.r.t. $(\mathscr{A}_n)_{n\in\mathbb{N}_0}$) if $u_n \in \mathcal{L}^1(\mathscr{A}_n)$ and

$$\int_A u_{n+1}\, d\mu \leqslant \int_A u_n\, d\mu \quad \forall A \in \mathscr{A}_n. \tag{23.3}$$

If we want to emphasize the underlying filtration, we write $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}_0}$.

Remark 23.2 (i) We can replace the σ-algebra $\mathscr{A}_n$ in (23.1) with any $\cap$-stable generator $\mathscr{G}_n$ of $\mathscr{A}_n$ containing a sequence $(G_k)_{k\in\mathbb{N}} \subset \mathscr{G}_n$ with $G_k \uparrow X$. This follows from the fact that

$$\int_A u_{n+1}\, d\mu = \int_A u_n d\mu \iff \underbrace{\int_A (u_{n+1}^+ + u_n^-) d\mu}_{=:\nu(A)} = \underbrace{\int_A (u_{n+1}^- + u_n^+)\, d\mu}_{=:\rho(A)},$$

where ν, ρ are finite measures on $\mathscr{A}_n$ and the uniqueness theorem for measures (Theorem 5.7), $\nu|_{\mathscr{G}_n} = \rho|_{\mathscr{G}_n}$, implies – under our assumptions on $\mathscr{G}_n$ – that $\nu = \rho$ on $\mathscr{A}_n$.

(For sub- or supermartingales we need, in addition, that $\mathscr{G}_n$ is a semi-ring, see Lemma 16.6.)

(ii) Set $\mathcal{S}_n := \{A \in \mathscr{A}_n : \mu(A) < \infty\}$. It is not hard to see that $\mathcal{S}_n$ is a semi-ring and that, because of σ-finiteness, $\sigma(\mathcal{S}_n) = \mathscr{A}_n$. Therefore (i) means that it suffices to assume (23.1)–(23.3) for all sets in $\mathcal{S}_n$, i.e. for all sets with *finite* μ-measure.

(iii) Condition (23.2) in Definition 23.1 is equivalent to

$$\int \phi u_{n+1}\, d\mu \geqslant \int \phi u_n\, d\mu \quad \forall \phi \in \mathcal{L}^\infty(\mathscr{A}_n),\ \phi \geqslant 0. \tag{23.2$'$}$$

Indeed: with $\phi := \mathbb{1}_A$ and $A \in \mathscr{A}_n$ we see that (23.2′) implies (23.2). Conversely, if $\phi \in \mathcal{E}(\mathscr{A}_n)$ is a positive simple function, (23.2′) follows from (23.2) by linearity. For general $\phi \in \mathcal{L}^\infty(\mathscr{A}_n)$, $\phi \geqslant 0$, the sombrero lemma (Theorem 8.8) furnishes a sequence of $\mathscr{A}_n$-measurable simple functions $\phi_k \geqslant 0$ with $\phi_k \uparrow \phi$. Since $\phi u_n, \phi u_{n+1} \in \mathcal{L}^1(\mathscr{A})$, we can use Lebesgue's dominated convergence theorem (Theorem 12.2) and get

$$\int \phi u_{n+1}\, d\mu = \lim_{k\to\infty} \int \phi_k u_{n+1}\, d\mu \overset{(23.2)}{\geqslant} \lim_{k\to\infty} \int \phi_k u_n\, d\mu = \int \phi u_n\, d\mu.$$

Similar statements hold for martingales (23.1) and supermartingales (23.3).

(iv) With some obvious (notational) changes in Definiton 23.1 we can also consider other index sets such as $\mathbb{N}$, $\mathbb{Z}$ or $-\mathbb{N}$.

Example 23.3 Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space.

(i) $(u_n)_{n\in\mathbb{N}_0}$ is a martingale if, and only if, it is both a sub- and a supermartingale.

(ii) $(u_n)_{n\in\mathbb{N}_0}$ is a supermartingale if, and only if, $(-u_n)_{n\in\mathbb{N}_0}$ is a submartingale.

(iii) Let $(u_n)_{n\in\mathbb{N}_0}$ and $(w_n)_{n\in\mathbb{N}_0}$ be [sub]martingales and let α, β be [positive] real numbers. Then $(\alpha u_n + \beta w_n)_{n\in\mathbb{N}_0}$ is a [sub]martingale.

(iv) Let $(u_n)_{n\in\mathbb{N}_0}$ be a submartingale. Then $(u_n^+)_{n\in\mathbb{N}_0}$ is a submartingale.
Indeed: take $A \in \mathscr{A}_n$ and observe that $\{u_n \geqslant 0\} \in \mathscr{A}_n$. Then

$$\begin{aligned}\int_A u_{n+1}^+\, d\mu \geqslant \int_{A\cap\{u_n\geqslant 0\}} u_{n+1}^+\, d\mu \ &\geqslant \int_{A\cap\{u_n\geqslant 0\}} u_{n+1}\, d\mu \\ &\overset{(23.2)}{\geqslant} \int_{A\cap\{u_n\geqslant 0\}} u_n\, d\mu = \int_A u_n^+\, d\mu.\end{aligned}$$

(v) Let $(u_n)_{n\in\mathbb{N}_0}$ be a martingale. Then $(|u_n|)_{n\in\mathbb{N}_0}$ is a submartingale. This follows from $|u_n| = 2u_n^+ - u_n$, (iii) and (iv).

(vi) Let $(u_n)_{n\in\mathbb{N}_0}$ be a martingale. If $u_n \in \mathcal{L}^p(\mathscr{A}_n)$ for some $p \in (1,\infty)$, then $(|u_n|^p)_{n\in\mathbb{N}_0}$ is a submartingale.
Indeed: note that $|y|^p - |x|^p = \int_{|x|}^{|y|} pt^{p-1}\, dt \geqslant p|x|^{p-1}(|y| - |x|)$ for all $x, y \in \mathbb{R}$, where we set, as usual, $\int_{|x|}^{|y|} = -\int_{|y|}^{|x|}$ if $|x| > |y|$. If we take $y = u_{n+1}$ and $x = u_n$ and integrate over $A \in \mathscr{A}_n$, we find by dominated convergence

$$\begin{aligned}\int_A \left(|u_{n+1}|^p - |u_n|^p\right) d\mu \ &\geqslant \ p \int (\mathbb{1}_A |u_n|^{p-1})(|u_{n+1}| - |u_n|)\, d\mu \\ &= \lim_{N\to\infty} p \int \big[\underbrace{(\mathbb{1}_A |u_n|^{p-1}) \wedge N}_{\in \mathcal{L}^\infty_+(\mathscr{A}_n)}\big](|u_{n+1}| - |u_n|)\, d\mu \\ &\underset{(v)}{\overset{(23.2')}{\geqslant}} 0,\end{aligned}$$

since $(|u_n|)_{n\in\mathbb{N}_0}$ is, by (v), a submartingale.

(vii) Let $u_n \in \mathcal{L}^1(\mathscr{A}_n)$, $n \in \mathbb{N}_0$, and $u_0 \leqslant u_1 \leqslant u_2 \leqslant \cdots$. Then $(u_n)_{n\in\mathbb{N}_0}$ is a submartingale.

(viii) Let $(X, \mathscr{A}, \mu) = \big([0,1), \mathscr{B}[0,1), \lambda := \lambda^1|_{[0,1)}\big)$ and then consider the finite (σ-)algebras generated by all *dyadic intervals* of $[0,1)$ of length 2^{-i}, $i \in \mathbb{N}_0$:

$$\mathscr{A}_i^{\Delta} := \sigma\big([0,2^{-i}), \ldots, [k2^{-i}, (k+1)2^{-i}), \ldots, [(2^i-1)2^{-i}, 1)\big).$$

Obviously, $\mathscr{A}_0^{\Delta} \subset \mathscr{A}_1^{\Delta} \subset \cdots \subset \mathscr{B}[0,1)$ and $([0,1), \mathscr{B}[0,1), \mathscr{A}_i^{\Delta}, \lambda)$ is a (σ-)finite filtered measure space. *Then* $(u_i)_{i\in\mathbb{N}_0}$, $u_i := 2^i{}_{[0,2^{-i})}$, *is a martingale.*

Indeed: since the sets $[k2^{-i}, (k+1)2^{-i})$, $k = 0, 1, \ldots, 2^i - 1$ are a disjoint partition of $[0,1)$, every $A \in \mathscr{A}$ consists of a (finite) disjoint union of such sets. If $[0, 2^{-i}) \subset A$, we have

$$\int_A u_{i+1}\, d\lambda = \int 2^{i+1} \mathbb{1}_{A\cap[0,2^{-(i+1)})}\, d\lambda = 1 = \int 2^i \mathbb{1}_{A\cap[0,2^{-i})}\, d\lambda = \int_A u_i\, d\lambda$$

and, otherwise, if $[0,2^{-i}) \cap A = \emptyset$,

$$\int_A u_{i+1}\, d\lambda = \int_A 2^{i+1} \mathbb{1}_{[0,2^{-(i+1)})}\, d\lambda = 0 = \int_A 2^i \mathbb{1}_{[0,2^{-i})}\, d\lambda = \int_A u_i\, d\lambda.$$

(ix) Let $(X, \mathscr{A}, \mu) = \big([0,\infty)^n, \mathscr{B}([0,\infty)^n), \lambda = \lambda^n|_{[0,\infty)^n}\big)$ and consider the σ-algebras $\mathscr{A}_i$ generated by the lattice of half-open *dyadic squares* of side-length 2^{-i}, $i \in \mathbb{N}_0$,

$$\mathscr{A}_i^{\Delta} := \sigma\big(z + [0,2^{-i})^n : z \in 2^{-i}\mathbb{N}_0^n\big), \quad i \in \mathbb{N}_0.$$

Then $\mathscr{A}_0^{\Delta} \subset \mathscr{A}_1^{\Delta} \subset \cdots \subset \mathscr{B}([0,\infty)^n)$, and $([0,\infty)^n, \mathscr{B}([0,\infty)^n), \mathscr{A}_i^{\Delta}, \lambda)$ is a σ-finite filtered measure space.

For every real-valued function $u \in \mathcal{L}^1([0,\infty)^n, \lambda)$ we can define an $\mathscr{A}_i^{\Delta}$-measurable step function u_i on the dyadic squares in $\mathscr{A}_i^{\Delta}$ by

$$\begin{aligned} u_i(x) &:= \sum_{z\in 2^{-i}\mathbb{N}_0^n} \frac{\int_{z+[0,2^{-i})^n} u\, d\lambda}{\lambda\big(z + [0,2^{-i})^n\big)} \mathbb{1}_{z+[0,2^{-i})^n}(x) \\ &= \sum_{z\in 2^{-i}\mathbb{N}_0^n} \left\{ \int u \frac{\mathbb{1}_{z+[0,2^{-i})^n}}{\lambda\big(z+[0,2^{-i})^n\big)}\, d\lambda \right\} \mathbb{1}_{z+[0,2^{-i})^n}(x). \end{aligned} \tag{23.4}$$

We claim that $(u_i, \mathscr{A}_i^{\Delta})_{i\in\mathbb{N}}$ *is a martingale.*

Indeed: since the sets $z + [0,2^{-i})^n$ are disjoint for different $z \in 2^{-i}\mathbb{N}_0^n$, the sums in (23.4) are actually finite sums.

That $u_i \in \mathcal{L}^1(\mathscr{A}_i^\Delta)$ is clear from the construction. To see (23.1), fix $z' \in 2^{-i}\mathbb{N}_0^n$ and $i \in \mathbb{N}_0$ and observe that for all $k = i, i+1, i+2, \ldots$

$$\begin{aligned}
&\int_{z'+[0,2^{-i})^n} u_k(x)\lambda(dx) \\
&\quad = \sum_{z\in 2^{-k}\mathbb{N}_0^n} \left\{ \int u \frac{\mathbb{1}_{z+[0,2^{-k})^n}}{\lambda\big(z+[0,2^{-k})^n\big)}\,d\lambda \right\} \cdot \int \mathbb{1}_{z+[0,2^{-k})^n} \mathbb{1}_{z'+[0,2^{-i})^n}\,d\lambda \\
&\quad = \sum_{\substack{z\in 2^{-k}\mathbb{N}_0^n \\ z+[0,2^{-k})^n \subset z'+[0,2^{-i})^n}} \int u \frac{\mathbb{1}_{z+[0,2^{-k})^n}}{\lambda\big(z+[0,2^{-k})^n\big)}\,d\lambda \cdot \lambda\big(z+[0,2^{-k})^n\big) \\
&\quad = \sum_{\substack{z\in 2^{-k}\mathbb{N}_0^n \\ z+[0,2^{-k})^n \subset z'+[0,2^{-i})^n}} \int_{z+[0,2^{-k})^n} u(x)\lambda(dx) \\
&\quad = \int_{z'+[0,2^{-i})^n} u(x)\lambda(dx).
\end{aligned}$$

The right-hand side is independent of k and, therefore, we get

$$\int_{z'+[0,2^{-i})^n} u_i\,d\lambda = \int_{z'+[0,2^{-i})^n} u\,d\lambda = \int_{z'+[0,2^{-i})^n} u_{i+1}\,d\lambda.$$

Since $\mathscr{A}_i^\Delta$ is generated by (disjoint unions of) squares of form $z' + [0, 2^{-i})^n$ with $z' \in 2^{-i}\mathbb{N}_0^n$, the claim follows from Remark 23.2(i).

(x) Assume that $(X, \mathscr{A}, \mu)$ is a probability space, i.e. a measure space where $\mu(X) = 1$. The functions $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mathscr{A})$ are called *independent*, if

$$\mu\left(\bigcap_{n=1}^{N} u_n^{-1}(B_n)\right) = \prod_{n=1}^{N} \mu\big(u_n^{-1}(B_n)\big) \tag{23.5}$$

holds for all $N \in \mathbb{N}$ and any choice of $B_1, B_2, \ldots, B_N \in \mathscr{B}(\mathbb{R})$. Then, if $\mathscr{A}_n := \sigma(u_1, u_2, \ldots, u_n)$ is the σ-algebra generated by $u_1, u_2, \ldots, u_n$, the sequence of partial sums

$$s_n := u_1 + u_2 + \cdots + u_n, \quad n \in \mathbb{N},$$

is an $(\mathscr{A}_n)_{n\in\mathbb{N}}$-submartingale if, and only if, $\int u_n\,d\mu \geqslant 0$ for all n.

To see this we need an *auxiliary result* which is of some interest on its own: *if $u_1, u_2, \ldots, u_{n+1}$ are independent integrable functions, then*

$$\int_A u_{n+1}\,d\mu = \mu(A) \int u_{n+1}\,d\mu \qquad \forall A \in \sigma(u_1, u_2, \ldots, u_n) \tag{23.6}$$

and

$$\int \phi u_{n+1}\,d\mu = \int \phi\,d\mu \cdot \int u_{k+1}\,d\mu \qquad \forall \phi \in \mathcal{L}^1(\sigma(u_1,\dots,u_n)). \tag{23.7}$$

In particular, integrable independent functions satisfy

$$\int \prod_{i=1}^n u_i\,d\mu = \prod_{i=1}^n \int u_i\,d\mu.$$

The proof of (23.6) and (23.7) will be given in Scholium 23.4 below.

Returning to the original problem, we find for all $A \in \mathscr{A}_n$ that

$$\begin{aligned}\int_A s_{n+1}\,d\mu = \int_A (s_n + u_{n+1})d\mu \;&=\; \int_A s_n\,d\mu + \int_A u_{n+1}\,d\mu \\ &\overset{23.6}{=} \int_A s_n\,d\mu + \mu(A)\int u_{n+1}\,d\mu.\end{aligned}$$

Thus, $\int u_{n+1}\,d\mu \geqslant 0$ is necessary and sufficient for $(s_n)_{n\in\mathbb{N}}$ to be a submartingale.

(xi) Let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mathscr{A}) \cap \mathcal{L}^\infty(\mathscr{A})$, $u_n \geqslant 0$, be independent functions (in the sense of (x)). The products $p_n := u_1 \cdot u_2 \cdot \dots \cdot u_n$, $n \in \mathbb{N}$, form a submartingale w.r.t. the filtration $\mathscr{A}_n := \sigma(u_1, u_2, \dots, u_n)$ if, and only if, $\int u_n\,d\mu \geqslant 1$ for all n. This follows directly from

$$\begin{aligned}\int_A p_{n+1}\,d\mu = \int \mathbb{1}_A p_n u_{n+1}\,d\mu &\overset{(23.7)}{=} \int \mathbb{1}_A p_n\,d\mu \cdot \int u_{n+1}\,d\mu \\ &= \int_A p_n\,d\mu \cdot \int u_{n+1}\,d\mu \quad \forall A \in \mathscr{A}_n.\end{aligned}$$

Scholium 23.4 (on independent functions) (i) For independent integrable functions $u_1, u_2, \dots, u_{n+1}$ on the probability space $(X, \mathscr{A}, \mu)$ the following identities hold:

$$\int_A u_{n+1}\,d\mu = \mu(A)\int u_{n+1}\,d\mu \qquad \forall A \in \sigma(u_1, u_2, \dots, u_n) \tag{23.6}$$

and

$$\int \phi u_{n+1}\,d\mu = \int \phi\,d\mu \cdot \int u_{n+1}\,d\mu \qquad \forall \phi \in \mathcal{L}^1(\sigma(u_1,\dots,u_n)). \tag{23.7}$$

Proof We begin with (23.6). Pick a set $A_N := \bigcap_{i=1}^N u_i^{-1}(B_i)$, $B_1, \dots, B_N \in \mathscr{B}(\mathbb{R})$, $N \leqslant n$, from the generator of $\mathscr{A}_n = \sigma(u_1, u_2, \dots, u_n)$. Because of Corollary 8.9 there exist simple functions $(f_l)_{l\in\mathbb{N}} \subset \mathcal{E}(\sigma(u_{n+1}))$ such that $|f_l| \leqslant |u_{n+1}|$

and $\lim_{l\to\infty} f_l = u_{n+1}$. We get for the standard representations $f_l = \sum_{i=0}^{N(l)} y_i^l \mathbb{1}_{H_i^l}$, $H_i^l \in \sigma(u_{n+1})$, using dominated convergence,

$$\int_{A_N} u_{n+1}\,d\mu \overset{12.2}{=} \lim_{l\to\infty} \int_{A_N} \sum_{i=0}^{N(l)} y_i^l \mathbb{1}_{H_i^l}\,d\mu = \lim_{l\to\infty} \sum_{i=0}^{N(l)} y_i^l \mu(A_N \cap H_i^l)$$
$$\overset{(23.5)}{=} \lim_{l\to\infty} \sum_{i=0}^{N(l)} y_i^l \mu(A_N)\mu(H_i^l) \overset{12.2}{=} \mu(A_N) \int u_{n+1}\,d\mu,$$

where we apply (23.5) for A_N and $H_i^l \in \sigma(u_{n+1})$ ($\iff H_i^l = u_{n+1}^{-1}(C_i^l)$ with some suitable $C_i^l \in \mathscr{B}(\mathbb{R})$). This proves (23.6) for a generator of $\mathscr{A}_n$ which satisfies the conditions stated in Remark 23.2(i); a similar argument to the one in this remark now proves that (23.6) holds for all $A \in \mathscr{A}_n$.

In order to prove (23.7), we observe that $u_{n+1}^+, u_n, \dots, u_1$, resp., $u_{n+1}^-, u_n, \dots, u_1$ are also independent. For u_{n+1}^+ this follows e.g. from the observation that

$$\{u_{n+1}^+ \in B\} = \begin{cases} \{u_{n+1} \in B \cap (0,\infty)\} \cup \{u_{n+1} \leqslant 0\} & \text{if } 0 \in B, \\ \{u_{n+1} \in B \cap (0,\infty)\} & \text{if } 0 \notin B, \end{cases}$$

i.e. every set $(u_{n+1}^+)^{-1}(B)$ can be expressed in the form $u_{n+1}^{-1}(C)$ for a suitable C.

Therefore, we can assume that u_{n+1} is positive. This means that (23.6) is an equality of two finite measures $\nu(A) = \int_A u_{n+1}\,d\mu$ and $\rho(A) = \mu(A)\int u_{n+1}\,d\mu$. With the usual construction of the integral starting with a measure, see Fig. 10.1 on page 82, we get (23.7). □

(ii) In Example 23.3(x) we assumed the existence of *infinitely many independent functions*. As a matter of fact, this matter is not completely trivial. If we want to construct *finitely many* independent functions $u_1, u_2, \dots, u_n$, we can proceed as follows: replace the probability space $(X, \mathscr{A}, \mu)$ by the n-fold product measure space $(X^n, \mathscr{A}^{\otimes n}, \mu^{\times n})$ (again a probability space [✍]) and set $\tilde{u}_i(x_1, \dots, x_n) := u_i(x_i)$ for $i = 1, 2, \dots, n$. Since each of the new functions $\tilde{u}_i$ depends only on the variable x_i, their independence follows from a simple Fubini-type argument. A similar argument can be applied to countably many functions – provided that we know how to construct infinite-dimensional products.

We will not follow this route, but construct instead countably many independent functions $(\xi_n)_{n\in\mathbb{N}}$ on the probability space $([0,1], \mathscr{B}[0,1], \lambda := \lambda^1|_{[0,1]})$, which are *identically distributed*, i.e. the image measures satisfy $\xi_1(\lambda) = \xi_n(\lambda)$ for all $n \in \mathbb{N}$ with a *Bernoulli distribution* $\xi_1(\lambda) = p\delta_1 + (1-p)\delta_0$, $p \in (0,1)$.

Consider the interval map $\beta_p : [0,1) \to [0,1)$

$$\beta_p(x) := \frac{x}{p}\mathbb{1}_{[0,p)}(x) + \frac{x-p}{1-p}\mathbb{1}_{[p,1)}(x),$$

and its iterates $\beta_p^n := \underbrace{\beta_p \circ \ldots \circ \beta_p}_{n \text{ times}}$, see Fig. 23.1 for the graphs of β_p and β_p^2. Define

$$\xi_n(x) := \mathbb{1}_{[0,p)}(\beta_p^{n-1}(x)), \quad n \in \mathbb{N}.$$

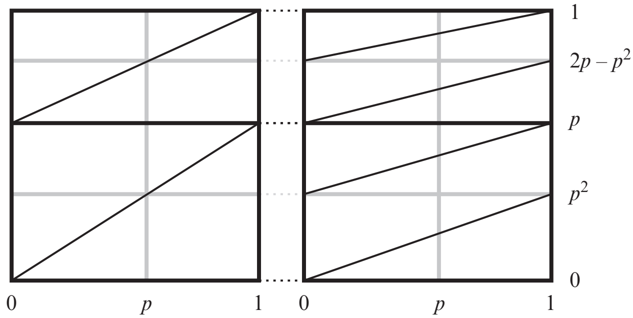


Fig. 23.1. Graph of β_p and β_p^2.

In the first step the interval $[0,1)$ is split according to $p:(1-p)$ into two intervals $[0,p)$ and $[p,1)$, and ξ_1 is 1 on the left segment and 0 on the right. The subsequent iterations split each of the intervals of the previous step – say, step $(n-1)$ – into two new sub-intervals according to the ratio $p:(1-p)$, and we define ξ_n to be 1 on each new left sub-interval and 0 otherwise, see the picture for the first two iterations $n=1,2$. Thus $\lambda\{\xi_n = 1\} = p$ and $\lambda\{\xi_n = 0\} = 1-p$, which means that the ξ_n are identically Bernoulli distributed.

In order to see independence, fix $\epsilon_i \in \{0,1\}$, $i = 1, \ldots, n-1$, and observe that the set $\{\xi_1 = \epsilon_1\} \cap \{\xi_2 = \epsilon_2\} \cap \cdots \cap \{\xi_{n-1} = \epsilon_{n-1}\}$ determines the segment before the nth split. Since each split preserves the proportion between p and $1-p$, we find

$$\begin{aligned}\lambda\big(\{\xi_1 = \epsilon_1\} \cap \cdots \cap \{\xi_{n-1} = \epsilon_{n-1}\} \cap \{\xi_n = 1\}\big)&\\ = \lambda\big(\{\xi_1 = \epsilon_1\} \cap \cdots \cap \{\xi_{n-1} = \epsilon_{n-1}\}\big) \cdot p&,\end{aligned}$$

so that

$$\begin{aligned}\lambda\big(\{\xi_1 = \epsilon_1\} \cap \ldots \cap \{\xi_{n-1} = \epsilon_{n-1}\} \cap \{\xi_n = \epsilon_n\}\big)&\\ = p^{\epsilon_1 + \cdots + \epsilon_n}(1-p)^{n-\epsilon_1-\cdots-\epsilon_n} = \prod_{i=1}^{n} \lambda\{\xi_i = \epsilon_i\}&.\end{aligned}$$

This shows that the ξ_i are independent.

For later reference purposes let us derive some formulae for the arithmetic means $\frac{1}{n}S_n := \frac{1}{n}(\xi_1 + \xi_2 + \cdots + \xi_n)$. The *mean value* is

$$\frac{1}{n}\int S_n\,d\lambda = \frac{1}{n}\int (\xi_1 + \cdots + \xi_n)d\lambda = \int \xi_1\,d\lambda = 1\cdot p + 0\cdot(1-p) = p,$$

while the *variance* is given by

$$\begin{aligned}
\int \left[\frac{1}{n}(S_n - np)\right]^2 d\lambda &= \frac{1}{n^2}\int \Big(\sum_{i=1}^{n}(\xi_i - p)\Big)^2 d\lambda \\
&= \frac{1}{n^2}\sum_{i,k=1}^{n}\int (\xi_i - p)(\xi_k - p)d\lambda \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\int (\xi_i - p)^2 d\lambda && \text{(independence)} \\
&= \frac{1}{n}\int (\xi_1 - p)^2 d\lambda && \text{(identical distr.)} \\
&= \frac{1}{n}\big((1-p)^2 p + p^2(1-p)\big) \\
&= \frac{1}{n}p(1-p). && \square
\end{aligned}$$

In the next chapter we study the convergence behaviour of a martingale $(u_n)_{n\in\mathbb{N}}$; therefore, it is natural to ask questions of the type '*from which index n onwards does $u_n(x)$ exceed a certain threshold*', etc. This means that we must be able to admit indices τ which may depend on the argument x of $u_n(x)$: $u_{\tau(x)}(x)$. The problem is measurability.

Definition 23.5 Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space. A *stopping time* is a map $\tau : X \to \mathbb{N}_0 \cup \{\infty\}$ which satisfies $\{\tau \leqslant n\} \in \mathscr{A}_n$ for all $n \in \mathbb{N}_0$. The associated σ-algebra is given by

$$\mathscr{A}_\tau := \big\{A \in \mathscr{A} : A \cap \{\tau \leqslant n\} \in \mathscr{A}_n \quad \forall n \in \mathbb{N}_0\big\}.$$

As usual, we write $u_\tau(x)$ instead of the more precise $u_{\tau(x)}(x)$.

Lemma 23.6 *Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space and σ, τ stopping times.*

(i) *$\sigma \wedge \tau$, $\sigma \vee \tau$, $\sigma + k$, $k \in \mathbb{N}_0$ are stopping times.*
(ii) *$\{\sigma < \tau\} \in \mathscr{A}_\sigma \cap \mathscr{A}_\tau$ and $\mathscr{A}_\sigma \subset \mathscr{A}_\tau$ if $\sigma \leqslant \tau$.*
(iii) *If $u_n \in \mathcal{M}(\mathscr{A}_n)$ is a sequence of real functions, then u_σ is $\mathscr{A}_\sigma$-measurable.*

Proof (i) follows immediately from the identities

$$\{\sigma \wedge \tau \leqslant n\} = \{\sigma \leqslant n\} \cup \{\tau \leqslant n\} \in \mathscr{A}_n,$$
$$\{\sigma \vee \tau \leqslant n\} = \{\sigma \leqslant n\} \cap \{\tau \leqslant n\} \in \mathscr{A}_n,$$
$$\{\sigma + k \leqslant n\} = \{\sigma \leqslant n-k\} \in \mathscr{A}_{(n-k)\vee 0} \subset \mathscr{A}_n.$$

(ii) Since for all $n \in \mathbb{N}_0$

$$\begin{aligned}\{\sigma < \tau\} \cap \{\sigma \leqslant n\} &= \bigcup_{k=0}^{n} \{\sigma = k\} \cap \{k < \tau\} \\ &= \bigcup_{k=0}^{n} \underbrace{\{\sigma \leqslant k\}}_{\in \mathscr{A}_k} \cap \underbrace{\{\sigma \leqslant k-1\}^c}_{\in \mathscr{A}_k} \cap \underbrace{\{\tau \leqslant k\}^c}_{\in \mathscr{A}_k} \in \mathscr{A}_n,\end{aligned}$$

we find that $\{\sigma < \tau\} \in \mathscr{A}_\sigma$, while a similar calculation for $\{\sigma < \tau\} \cap \{\tau \leqslant n\}$ yields $\{\sigma < \tau\} \in \mathscr{A}_\tau$.

If $\sigma \leqslant \tau$ we find for $A \in \mathscr{A}_\sigma$

$$A \cap \{\tau \leqslant n\} = A \cap \underbrace{\{\sigma \leqslant \tau\}}_{=\Omega} \cap \{\tau \leqslant n\} = \underbrace{A \cap \{\sigma \leqslant n\}}_{\in \mathscr{A}_n} \cap \underbrace{\{\tau \leqslant n\}}_{\in \mathscr{A}_n} \in \mathscr{A}_n,$$

i.e. $A \in \mathscr{A}_\tau$, hence $\mathscr{A}_\sigma \subset \mathscr{A}_\tau$.

(iii) We have for all $B \in \mathscr{B}(\mathbb{R})$ and $n \in \mathbb{N}_0 \cup \{\infty\}$

$$\begin{aligned}\{u_\sigma \in B\} \cap \{\sigma \leqslant n\} &= \bigcup_{k=0}^{n} \{u_k \in B\} \cap \{\sigma = k\} \\ &= \bigcup_{k=0}^{n} \underbrace{\{u_k \in B\}}_{\in \mathscr{A}_k} \cap \underbrace{\{\sigma \leqslant k\}}_{\in \mathscr{A}_k} \cap \underbrace{\{\sigma \leqslant k-1\}^c}_{\in \mathscr{A}_k} \in \mathscr{A}_n. \quad \square\end{aligned}$$

The next result is a very useful characterization of (sub)martingales.

Theorem 23.7 *Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space. For a sequence $(u_n)_{n\in\mathbb{N}_0}$, $u_n \in \mathcal{L}^1(\mathscr{A}_n)$, the following assertions are equivalent:*

(i) *$(u_n)_{n\in\mathbb{N}_0}$ is a submartingale;*
(ii) *$\int u_\sigma \, d\mu \leqslant \int u_\tau \, d\mu$ for all bounded stopping times $\sigma \leqslant \tau$;*
(iii) *$\int_A u_\sigma \, d\mu \leqslant \int_A u_\tau \, d\mu$ for all bounded stopping times $\sigma \leqslant \tau$ and $A \in \mathscr{A}_\sigma$.*

Proof (i)$\Rightarrow$(ii). Let $\sigma \leqslant \tau \leqslant N$ be two stopping times. By Lemma 23.6 u_σ is measurable, and since

$$\int |u_\sigma| d\mu = \sum_{n=1}^{N} \int_{\{\sigma = n\}} |u_n| d\mu \leqslant \sum_{n=1}^{N} \int |u_n| d\mu < \infty,$$

we find that $u_\sigma, u_\tau \in \mathcal{L}^1(X, \mathscr{A}, \mu)$. Therefore, the following integrals exist:

$$\int (u_\sigma - u_\tau) d\mu = \int \sum_{n=\sigma}^{\tau-1} (u_n - u_{n+1}) d\mu = \sum_{n=0}^{N-1} \int (u_n - u_{n+1}) \mathbb{1}_{\{\sigma=n\}\cap\{n<\tau\}} \, d\mu.$$

Note that the set $\{\sigma = n\} \cap \{n < \tau\} = \{\sigma = n\} \cap \{\tau \leqslant n\}^c \in \mathscr{A}_n$. By the definition of a submartingale,

$$\int (u_n - u_{n+1}) \mathbb{1}_{\{\sigma=n\}\cap\{n<\tau\}} \, d\mu = \int_{\{\sigma=n\}\cap\{n<\tau\}} (u_n - u_{n+1}) d\mu \leqslant 0,$$

and this proves (ii).

(ii) $\Rightarrow$ (iii). Note that for any $A \in \mathscr{A}_\sigma$ the function $\rho := \rho_A := \sigma \mathbb{1}_A + \tau \mathbb{1}_{A^c}$ is again a bounded stopping time. This follows from

$$\{\rho \leqslant n\} = (\{\sigma \leqslant n\} \cap A) \cup (\{\tau \leqslant n\} \cap A^c) \in \mathscr{A}_n, \quad n \in \mathbb{N},$$

where we use that $A \in \mathscr{A}_\sigma \subset \mathscr{A}_\tau$, cf. Lemma 23.6. Since $\rho \leqslant \tau$, (ii) shows

$$\int (u_\sigma \mathbb{1}_A + u_\tau \mathbb{1}_{A^c}) \, d\mu = \int u_\rho \, d\mu \overset{\text{(ii)}}{\leqslant} \int u_\tau \, d\mu,$$

which yields $\int_A u_\sigma \, d\mu \leqslant \int_A u_\tau \, d\mu$.

(iii) $\Rightarrow$ (i). Take $\sigma = n$ and $\tau = n + 1$. □

Remark 23.8 One consequence of Theorem 23.7(iii) is the following result. *Let $\tau_1 \leqslant \tau_2 \leqslant \ldots \leqslant \tau_n \leqslant N$ be bounded stopping times. Then*

$$\begin{matrix} (u_i, \mathscr{A}_i)_{i\in\mathbb{N}} \textit{ is a} \\ \textit{submartingale} \end{matrix} \implies \begin{matrix} (u_{\tau_i}, \mathscr{A}_{\tau_i})_{i=1,\ldots,n} \textit{ is a} \\ \textit{submartingale}. \end{matrix}$$

This statement is often called the *optional sampling theorem*. A typical application is the situation where $\tau_i = \sigma_i \wedge N$ and $\sigma_1 \leqslant \sigma_2 \leqslant \cdots$ are increasing stopping times.

The boundedness of the stopping times in Theorem 23.8 essentially guarantees the integrability of the functions u_σ, u_τ. We may replace it by either of the following conditions.

(a) $\tau < \infty$ a.e. and $(u_n)_n$ has an integrable majorant: $|u_n| \leqslant w$ for some $w \in \mathcal{L}^1(\mu)$.
(b) $\int \tau \, d\mu < \infty$ and $|u_n - u_{n-1}| \leqslant K$ for some constant K.

Proof We know from (the proof of) Theorem 23.7 that $u_{\tau\wedge n}, u_{\sigma\wedge n} \in \mathcal{L}^1(\mu)$ and $\int u_{\sigma\wedge n} \, d\mu \leqslant \int u_{\tau\wedge n} \, d\mu$; all we have to do is let $n \to \infty$.

Under (a), $|u_{\tau\wedge n}| \leqslant w$ and $\lim_{n\to\infty} u_{\tau\wedge n} = u_\tau$ a.e. since $\tau < \infty$ a.e. Hence, we can use the dominated convergence theorem to see that $u_\tau \in \mathcal{L}^1(\mu)$ and

$$\int u_\sigma \, d\mu = \lim_{n\to\infty} \int u_{\sigma\wedge n} \leqslant \lim_{n\to\infty} \int u_{\tau\wedge n} = \int u_\tau \, d\mu.$$

If (b) holds, we get

$$|u_{\tau\wedge n} - u_0| = \left| \sum_{i=1}^{\tau\wedge n} (u_i - u_{i-1}) \right| \leqslant \sum_{i=1}^{\tau} |u_i - u_{i-1}| \leqslant K \cdot \tau \in L^1(\mu),$$

and now the argument of part (a) works: $\sigma, \tau < \infty$ a.e. and so $\lim_{n\to\infty} u_{\tau\wedge n} = u_\tau$ a.e. □

Problems

Unless otherwise stated $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ will be a σ-finite filtered measure space.

23.1. Let $(X, \mathscr{A}, \mu)$ be a finite measure space and let $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ be a martingale. Set $\mathscr{A}_0 := \{\emptyset, X\}$ and $u_0 = \mu(X)^{-1} \int u_1 \, d\mu$. Show that this is the only choice which makes $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}_0}$ a martingale.

23.2. Let $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ be a (sub, super)martingale and let $(\mathscr{B}_n)_{n\in\mathbb{N}}$ and $(\mathscr{C}_n)_{n\in\mathbb{N}}$ be filtrations in $\mathscr{A}$ such that $\mathscr{B}_n \subset \mathscr{A}_n \subset \mathscr{C}_n$.

(i) Show that $(u_n, \mathscr{B}_n)_{n\in\mathbb{N}}$ is again a (sub, super)martingale.

(ii) Show that $(u_n, \mathscr{C}_n)_{n\in\mathbb{N}}$ is, in general, no longer a (sub, super)martingale.

23.3. Completion (7). Let $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ be a submartingale and denote by $\overline{\mathscr{A}}_n$ the completion of $\mathscr{A}_n$. Then $(u_n, \overline{\mathscr{A}}_n)_{n\in\mathbb{N}}$ is still a submartingale.

23.4. Show that $(u_n)_{n\in\mathbb{N}}$ is a submartingale if, and only if, $u_n \in \mathcal{L}^1(\mathscr{A}_n)$ for all $n \in \mathbb{N}$ and

$$\int_A u_n \, d\mu \leqslant \int_A u_k \, d\mu \quad \forall n < k, \ \forall A \in \mathscr{A}_n.$$

Find similar statements for martingales and supermartingales.

23.5. Prove the assertion made in Remark 23.2(ii).

23.6. Let $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ be a martingale with $u_n \in \mathcal{L}^2(\mathscr{A}_n)$. Show that

$$\int u_n u_k \, d\mu = \int u_{n\wedge k}^2 \, d\mu.$$

[Hint: assume that $n < k$. Approximate u_n by simple functions from $\mathcal{E}(\mathscr{A}_n)$, then use dominated convergence and (23.1).]

23.7. Martingale transform. Let $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ be a martingale and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of bounded functions such that $f_n \in \mathcal{M}(\mathscr{A}_n)$ for every $n \in \mathbb{N}$. Set $f_0 := 0$ and $u_0 := 0$. Then the so-called *martingale transform*

$$(f \cdot u)_n := \sum_{i=1}^{n} f_{i-1} \cdot (u_i - u_{i-1}), \quad n \in \mathbb{N},$$

is again a martingale w.r.t. $(\mathscr{A}_n)_{n\in\mathbb{N}}$.

23.8. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent identically distributed random variables with $\xi_n \in \mathcal{L}^2(\mathscr{A})$ and $\int \xi_n \, d\mathbb{P} = 0$. Set $\mathscr{A}_n := \sigma(\xi_1, \xi_2, \ldots, \xi_n)$.

(i) Show, without using Example 23.3(vi), that $S_n^2 := (\xi_1 + \xi_2 + \cdots + \xi_n)^2$ is a submartingale w.r.t. $(\mathscr{A}_n)_{n\in\mathbb{N}}$.

(ii) Show that there is a constant κ such that $S_n^2 - \kappa n$ is a martingale w.r.t. $(\mathscr{A}_n)_{n\in\mathbb{N}}$.

23.9. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables with $\xi_n \in \mathcal{L}^2(\mathscr{A})$, $\int \xi_n \, d\mathbb{P} = 0$ and $\int \xi_n^2 \, d\mathbb{P} = \sigma_n^2$. Then set $\mathscr{A}_n := \sigma(\xi_1, \xi_2, \ldots, \xi_n)$ and $A_n := \sigma_1^2 + \cdots + \sigma_n^2$. Show that

$$M_n := S_n^2 - A_n = \left(\sum_{i=1}^n \xi_i \right)^2 - \sum_{i=1}^n \sigma_i^2$$

is a martingale.
[Hint: use formulae (23.6) and (23.7) and Remark 23.2(ii).]

23.10. **Martingale difference sequence.** Let $(d_i)_{i\in\mathbb{N}}$ be a sequence in $\mathcal{L}^2(\mathscr{A}) \cap \mathcal{L}^1(\mathscr{A})$. Define $\mathscr{A}_0 := \{\emptyset, X\}$ and $\mathscr{A}_n := \sigma(d_1, d_2, \ldots, d_n)$. Suppose that for each $n \in \mathbb{N}$

$$\int_A d_n \, d\mu = 0 \quad \forall A \in \mathscr{A}_{n-1}.$$

Show that $(u_n^2)_{n\in\mathbb{N}}$, $u_n := d_1 + \cdots + d_n$, is a submartingale which satisfies

$$\int u_n^2 \, d\mu = \sum_{i=1}^n \int d_i^2 \, d\mu.$$

Show that on $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda^1)$ the sequence $d_i(x) := \operatorname{sgn} \sin(2^i \pi x)$, $x \in \mathbb{R}$, $i \in \mathbb{N}$, is a martingale difference sequence. (See Chapter 28, in particular pages 389 and 391 for further details.)

23.11. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent identically Bernoulli $(p, 1-p)$-distributed random variables with values ± 1, i.e. such that $\mathbb{P}(\xi_n = 1) = p$ and $\mathbb{P}(\xi_n = -1) = 1 - p$ – this can be constructed as in Scholium 23.4. Set $S_n := \xi_1 + \cdots + \xi_n$. Then $((1-p)/p)^{S_n}$ is a martingale w.r.t. the filtration given by $\mathscr{A}_n := \sigma(\xi_1, \ldots, \xi_n)$.

23.12. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space, let ν be a further measure on $\mathscr{A}$ and let $(A_{n,i})_{i\in\mathbb{N}} \subset \mathscr{A}$ be for each $n \in \mathbb{N}$ a sequence of mutually disjoint sets such that $X = \dot\bigcup_{i\in\mathbb{N}} A_{n,i}$. Assume, moreover, that each set $A_{n,i}$ is the union of finitely many sets from the sequence $(A_{n+1,k})_{k\in\mathbb{N}}$. Show that

(i) the σ-algebras $\mathscr{A}_n := \sigma(A_{n,i} : i \in \mathbb{N})$ form a filtration;

(ii) if $\mu(A_{n,i}) > 0$ for all $n, i \in \mathbb{N}$, then

$$u_n := \sum_{i=1}^{\infty} \frac{\nu(A_{n,i})}{\mu(A_{n,i})} \mathbb{1}_{A_{n,i}}$$

is a martingale w.r.t. $(\mathscr{A}_n)_{n\in\mathbb{N}}$.

23.13. Let $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ be a supermartingale and $u_n \geqslant 0$ a.e. Prove that $u_k = 0$ a.e. implies that $u_{k+n} = 0$ a.e. for all $n \in \mathbb{N}$.

23.14. Verify that the family $\mathscr{A}_\tau$ defined in Definition 23.5 is indeed a σ-algebra.

23.15. Show that τ is a stopping time if, and only if, $\{\tau = n\} \in \mathscr{A}_n$ for all $n \in \mathbb{N}_0$.

23.16. Show that, in the notation of Lemma 23.6, $\mathscr{A}_{\sigma\wedge\tau} = \mathscr{A}_\sigma \cap \mathscr{A}_\tau$ for any two stopping times σ, τ.

24

Martingale Convergence Theorems

Throughout this chapter $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ is a σ-finite filtered measure space. Among the foremost applications of martingales are convergence theorems. Let us begin with the following simple observation for a sequence $(u_n)_{n\in\mathbb{N}_0}$ of real numbers. If $(u_n)_{n\in\mathbb{N}_0}$ has a limit $\ell = \lim_{n\to\infty} u_n$ and if we know that $\ell \in (a,b)$, only *finitely many* of the u_n can be outside of (a,b). In particular, if infinitely many u_n are bigger than b and infinitely many smaller than a, then the sequence has no limit at all. We call any occurrence of

$$u_n \leqslant a \quad \text{and} \quad u_{n+k} \geqslant b \quad \text{(for the first possible } k\in\mathbb{N})$$

an *upcrossing of* $[a,b]$ – Fig. 24.1 shows for $n = 0, 1, \ldots, N$ three such upcrossings – and we have just observed that, if for *some* pair $a, b \in \mathbb{R}$, $a < b$,

$$\#\{\text{upcrossings of } [a,b]\} = \infty \implies (u_n)_{n\in\mathbb{N}_0} \text{ has no limit.} \tag{24.1}$$

For a submartingale we can estimate the average number of upcrossings U as follows.

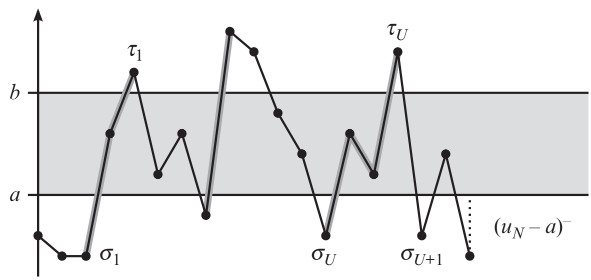


Fig. 24.1. Shown are $U = 3$ upcrossings over the strip $[a,b]$ and one aborted upcrossing attempt which ended at step N below the level a – this is the 'worst case' scenario for the estimate.
Note that $\tau_U < \sigma_{U+1} < \tau_{U+1} = \sigma_{U+2} = \tau_{U+2} = \ldots = \sigma_N = \tau_N = N$.

Lemma 24.1 (Doob's upcrossing estimate) *Let $(u_n)_{n\in\mathbb{N}}$ be a submartingale and denote by $U([a,b];N;x)$ the number of upcrossings of $(u_n(x))_{n\in\mathbb{N}_0}$ across $[a,b]$ which occur for the indices $0\leqslant n\leqslant N$. Then*

$$\int_A U([a,b];N)\,d\mu \leqslant \frac{1}{b-a}\int_A (u_N-a)^+\,d\mu \quad \forall A\in\mathscr{A}_0.$$

Proof In order to keep track of the upcrossings we introduce the following stopping times [✍], see Problem 24.1: define $\tau_0 := 0$, and recursively, for $k=1,2,3,\dots,$

$$\sigma_k := \inf\{n>\tau_{k-1} : u_n\leqslant a\}\wedge N \quad \text{and} \quad \tau_k := \inf\{n>\sigma_k : u_n\geqslant b\}\wedge N;$$

as usual we set $\inf\emptyset = +\infty$. Writing $U=U([a,b];N)$, we find that

$$\tau_0 = 0<\sigma_1<\tau_1<\sigma_2<\cdots<\sigma_U<\tau_U\leqslant\sigma_{U+1}\leqslant\tau_{U+1}=\cdots=\sigma_N=\tau_N=N.$$

By the very definition of an upcrossing we get

$$\begin{aligned}(b-a)\,U([a,b];N) &\leqslant \overbrace{(u_{\tau_1}-a)}^{\geqslant b-a}+\overbrace{(u_{\tau_2}-u_{\sigma_2})}^{\geqslant b-a}+\cdots+\overbrace{(u_{\tau_U}-u_{\sigma_U})}^{\geqslant b-a},\\ -(u_N-a)^- &\leqslant (u_{\tau_{U+1}}-u_{\sigma_{U+1}})+\underbrace{(u_{\tau_{U+2}}-u_{\sigma_{U+2}})}_{=0}+\cdots+\underbrace{(u_{\tau_N}-u_{\sigma_N})}_{=0}.\end{aligned}$$

Adding these two estimates and integrating both sides of the resulting inequality over $A\in\mathscr{A}_0$ yields, after some simple rearrangements,

$$\begin{aligned}&(b-a)\int_A U([a,b];N)\,d\mu - \int_A (u_N-a)^-\,d\mu\\ &\leqslant -\int_A a\,d\mu + \overbrace{\int_A (u_{\tau_1}-u_{\sigma_2})\,d\mu}^{\leqslant 0}+\cdots+\overbrace{\int_A (u_{\tau_{N-1}}-u_{\sigma_N})\,d\mu}^{\leqslant 0}+\int_A u_{\tau_N}\,d\mu\\ &\overset{23.7}{\leqslant} \int_A (u_{\tau_N}-a)\,d\mu.\end{aligned}$$

□

The upcrossing lemma is the basis for all martingale convergence theorems.

Theorem 24.2 (submartingale convergence) *Let $(u_n,\mathscr{A}_n)_{n\in\mathbb{N}_0}$ be a submartingale on the σ-finite filtered measure space $(X,\mathscr{A},\mathscr{A}_n,\mu)$. If $\sup_{n\in\mathbb{N}}\int u_n^+\,d\mu<\infty$, then the limit $u_\infty(x) := \lim_{n\to\infty}u_n(x)$ exists for almost all $x\in\mathbb{R}$ and is an $\mathscr{A}_\infty$-measurable function.*

Before we give the details of the proof, let us note some immediate consequences.

Corollary 24.3 *Under any of the following conditions the limit* $\lim_{n\to\infty} u_n$ *exists a.e. in* $\mathbb{R}$.

(i) $(u_n)_{n\in\mathbb{N}_0}$ *is a supermartingale and* $\sup_{n\in\mathbb{N}} \int u_n^- \, d\mu < \infty$.
(ii) $(u_n)_{n\in\mathbb{N}_0}$ *is a positive supermartingale.*
(iii) $(u_n)_{n\in\mathbb{N}_0}$ *is a martingale and* $\sup_{n\in\mathbb{N}} \int |u_n| d\mu < \infty$.

Proof of Theorem 24.2 In view of (24.1) we have

$$\Big\{x : \lim_{n\to\infty} u_n(x) \text{ does not exist}\Big\} = \Big\{x : \limsup_{n\to\infty} u_n(x) > \liminf_{n\to\infty} u_n(x)\Big\}$$
$$= \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \Big\{x : \sup_{N\in\mathbb{N}} U([a,b];N;x) = \infty\Big\}.$$

Since $\mu|_{\mathscr{A}_0}$ is σ-finite, there exists an exhausting sequence $(A_k)_{k\in\mathbb{N}} \subset \mathscr{A}_0$ such that $A_k \uparrow X$ and $\mu(A_k) < \infty$. From the inequality $(\beta - \alpha)^+ \leqslant \beta^+ + |\alpha|$ we find

$$\begin{aligned}\int_{A_k} \sup_{N\in\mathbb{N}} U([a,b];N) d\mu &\overset{9.6}{=} \sup_{N\in\mathbb{N}} \int_{A_k} U([a,b];N) d\mu \\ &\overset{24.1}{\leqslant} \frac{1}{b-a} \sup_{N\in\mathbb{N}} \int_{A_k} (u_N - a)^+ \, d\mu \\ &\leqslant \frac{1}{b-a} \Big(\sup_{N\in\mathbb{N}} \int_{A_k} u_N^+ \, d\mu + |a|\mu(A_k)\Big) < \infty,\end{aligned}$$

and a routine application of Markov's inequality (Proposition 11.5) yields

$$\mu\Big(\Big\{\sup_{N\in\mathbb{N}} U([a,b];N) = \infty\Big\} \cap A_k\Big) = 0.$$

Since $\bigcup_{k\in\mathbb{N}} \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \big(\{x : \sup_{N\in\mathbb{N}} U([a,b];N;x) = \infty\} \cap A_k\big)$ is a countable union of null sets, it is itself a null set; thus the limit $u_\infty(x) = \lim_{n\to\infty} u_n(x)$ exists for almost all $x \in \mathbb{R}$ and is $\mathscr{A}_\infty$-measurable. An application of Fatou's lemma (Theorem 9.11) shows that

$$\int |u_\infty| d\mu = \int \liminf_{n\to\infty} |u_n| d\mu \leqslant \liminf_{n\to\infty} \int |u_n| d\mu \leqslant \sup_{n\in\mathbb{N}} \int |u_n| d\mu,$$

while the submartingale property gives

$$\sup_{n\in\mathbb{N}_0} \int |u_n| d\mu = \sup_{n\in\mathbb{N}_0} \Big(2\int u_n^+ \, d\mu - \int u_n \, d\mu\Big) \leqslant 2 \sup_{n\in\mathbb{N}_0} \int u_n^+ \, d\mu - \int u_1 \, d\mu.$$

The last expression is, by assumption, finite and we conclude that $u_\infty \in \mathcal{L}^1(\mathscr{A}_\infty)$ and $|u_\infty| < \infty$ a.e. $\square$

We have seen in Example 23.3(v) that for a martingale $(u_n)_{n\in\mathbb{N}_0}$ the sequence $(|u_n|)_{n\in\mathbb{N}_0}$ is a submartingale. Therefore,

$$\int |u_0|d\mu \leqslant \int |u_1|d\mu \leqslant \int |u_2|d\mu \leqslant \cdots .$$

Hence, any martingale with index set *running to the left*, say, $(w_\ell)_{\ell\in-\mathbb{N}_0}$, automatically satisfies condition (iii) of Corollary 24.3: $\int |w_{-n}|d\mu \leqslant \int |w_0|d\mu < \infty$, and the a.e. limit $\lim_{n\to+\infty} w_{-n}$ always exists.

Definition 24.4 Let $(X, \mathscr{A}, \mu)$ be a measure space and $\mathscr{A}_0 \supset \mathscr{A}_{-1} \supset \mathscr{A}_{-2} \supset \cdots$ be a decreasing filtration of sub-σ-algebras of $\mathscr{A}$ such that $\mu|_{\mathscr{A}_{-n}}$ is σ-finite for every $n \in \mathbb{N}$. A (sub, super)martingale $(w_\ell, \mathscr{A}_\ell)_{\ell\in-\mathbb{N}_0}$ is called a *reversed* or *backwards* (*sub, super*)martingale.

Corollary 24.5 (backwards convergence) *Let $(w_\ell, \mathscr{A}_\ell)_{\ell\in-\mathbb{N}_0}$ be a backwards submartingale. If on $\mathscr{A}_{-\infty} = \bigcap_{\ell\in-\mathbb{N}_0} \mathscr{A}_\ell$ the measure $\mu|_{\mathscr{A}_{-\infty}}$ is σ-finite, then the limit $\lim_{n\to+\infty} w_{-n}(x) \in [-\infty, +\infty)$ exists for a.a. x and defines an $\mathscr{A}_{-\infty}/\mathscr{B}[-\infty, \infty)$-measurable function.*

Proof The proof of Doob's upcrossing lemma obviously applies to the (finite) sequence $w_{-N}, w_{-N+1}, \ldots, w_0$ and yields the following variant of the upcrossing inequality:

$$(b-a)\int_A U([a,b]; -N)d\mu \leqslant \int_A (w_0 - a)^+ \, d\mu \quad \forall A \in \mathscr{A}_{-N}.$$

This means that the arguments of the proof of Theorem 24.2 remain valid without further conditions on $(w_\ell)_{\ell\in-\mathbb{N}}$. But $(w_\ell^+)_{\ell\in-\mathbb{N}}$ is again a submartingale, see Example 23.3(iv), thus

$$\int w_{-\infty}^+ \, d\mu = \int \liminf_{n\to+\infty} w_{-n}^+ \, d\mu \overset{9.11}{\leqslant} \liminf_{n\to+\infty} \int w_{-n}^+ \, d\mu \leqslant \int w_0^+ \, d\mu < \infty.$$

By Corollary 11.6, $\mu\{w_{-\infty} = +\infty\} = \mu\{w_{-\infty}^+ = +\infty\} = 0$. □

Theorem 24.2 does not guarantee $\mathcal{L}^1$-convergence of a martingale. An example of a martingale which satisfies all of the conditions of Theorem 24.2, but fails to have a limit in $\mathcal{L}^1$, is given in Example 23.3(viii): here $u_n \to 0$ a.e., while $\int_{[0,1)} u_n \, d\lambda = 1 \not\to 0$. Such phenomena can be avoided if we assume that the submartingale is *uniformly integrable* (*UI*).

Recall from Definition 22.1 that $(u_n)_{n\in\mathbb{N}_0}$ is uniformly integrable if

$$\forall \epsilon > 0 \quad \exists w_\epsilon \in \mathcal{L}^1(\mathscr{A}, \mu),\ w_\epsilon \geqslant 0 : \sup_{n\in\mathbb{N}_0} \int_{\{|u_n| > w_\epsilon\}} |u_n| d\mu < \epsilon.$$

Theorem 24.6 (convergence of UI submartingales) *Let $(u_n)_{n\in\mathbb{N}_0}$ be a submartingale on the σ-finite filtered measure space $(X, \mathscr{A}, \mathscr{A}_n, \mu)$. The following assertions are equivalent.*

(i) $u_\infty(x) = \lim_{n\to\infty} u_n(x)$ *exists a.e.,* $u_\infty \in \mathcal{L}^1(\mathscr{A}_\infty, \mu)$,

$$\lim_{n\to\infty} \int u_n \, d\mu = \int u_\infty \, d\mu$$

and $(u_n)_{n\in\mathbb{N}_0\cup\{\infty\}}$ *is a submartingale.*

(ii) $(u_n)_{n\in\mathbb{N}_0}$ *is uniformly integrable.*

(iii) $(u_n)_{n\in\mathbb{N}_0}$ *converges in* $\mathcal{L}^1(\mathscr{A}_\infty)$.

Proof (i) $\Rightarrow$ (ii). Since $\mu|_{\mathscr{A}_0}$ is σ-finite, there exists a sequence $(A_k)_{k\in\mathbb{N}} \subset \mathscr{A}_0$ which is exhausting, i.e. $A_k \uparrow X$ and $\mu(A_k) < \infty$. It is not hard to see that $w := \sum_{k=1}^{\infty} 2^{-k}(1 + \mu(A_k))^{-1}\mathbb{1}_{A_k}$ is strictly positive and integrable: $w > 0$ and $w \in \mathcal{L}^1(\mathscr{A}_0, \mu)$. Example 23.3(iv) shows that $(u_n^+)_{n\in\mathbb{N}\cup\{\infty\}}$ is still a submartingale, so that for every $L > 0$

$$\begin{aligned}
\int_{\{u_n^+ > Lw\}} u_n^+ \, d\mu &\leqslant \int_{\{u_n^+ > Lw\}} u_\infty^+ \, d\mu \\
&\leqslant \int_{\left\{u_n^+ > Lw\right\}\cap\left\{u_\infty^+ > \frac{1}{2}Lw\right\}} u_\infty^+ \, d\mu + \int_{\left\{u_n^+ > Lw\right\}\cap\left\{u_\infty^+ \leqslant \frac{1}{2}Lw\right\}} u_\infty^+ \, d\mu \\
&\leqslant \int_{\left\{u_\infty^+ > \frac{1}{2}Lw\right\}} u_\infty^+ \, d\mu + \frac{1}{2}\int_{\left\{u_n^+ > Lw\right\}\cap\left\{u_\infty^+ \leqslant \frac{1}{2}Lw\right\}} u_n^+ \, d\mu.
\end{aligned}$$

For the last estimate we use that on the set $\left\{u_n^+ > Lw\right\} \cap \left\{u_\infty^+ \leqslant \frac{1}{2}Lw\right\}$ the integrand of the second integral satisfies $u_\infty^+ \leqslant \frac{1}{2}Lw < \frac{1}{2}u_n^+$. Now we subtract $\frac{1}{2}\int_{\{u_n^+ > Lw\}} u_n^+ \, d\mu$ from both sides and get, uniformly for all $n \in \mathbb{N}_0$,

$$\frac{1}{2}\int_{\{u_n^+ > Lw\}} u_n^+ \, d\mu \leqslant \int_{\left\{u_\infty^+ > \frac{1}{2}Lw\right\}} u_\infty^+ \, d\mu \xrightarrow[L\to\infty]{} 0.$$

This follows from the dominated convergence theorem since $|u_\infty| \in \mathcal{L}^1(\mathscr{A}_\infty, \mu)$ dominates $u_\infty^+ \mathbb{1}_{\left\{u_\infty^+ > \frac{1}{2}Lw\right\}} \xrightarrow[L\to\infty]{} 0$. Thus, $(u_n^+)_{n\in\mathbb{N}}$ is uniformly integrable. From $\lim_{n\to\infty} u_n = u_\infty$ a.e., we conclude that $\lim_{n\to\infty} u_n^+ = u_\infty^+$, and Vitali's convergence theorem (Theorem 22.7) shows that $\lim_{n\to\infty} \int u_n^+ \, d\mu = \int u_\infty^+ \, d\mu$. Thus

$$\int |u_n| d\mu = \int (2u_n^+ - u_n) d\mu \xrightarrow[n\to\infty]{} \int (2u_\infty^+ - u_\infty) d\mu = \int |u_\infty| d\mu,$$

and another application of Vitali's theorem proves that $(u_n)_{n\in\mathbb{N}}$ is uniformly integrable.

(ii) $\Rightarrow$ (iii). Because of uniform integrability we have for some $\epsilon > 0$ and a suitable $w_\epsilon \in \mathcal{L}^1(\mathscr{A})$

$$\int |u_n| d\mu = \int_{\{|u_n| > w_\epsilon\}} |u_n| d\mu + \int_{\{|u_n| \leqslant w_\epsilon\}} |u_n| d\mu$$
$$\leqslant \epsilon + \int w_\epsilon \, d\mu < \infty,$$

and the submartingale convergence theorem (Theorem 24.2) guarantees that the pointwise limit $u_\infty = \lim_{n\to\infty} u_n$ exists a.e.; $\mathcal{L}^1$-convergence follows from Vitali's convergence theorem (Theorem 22.7).

(iii) $\Rightarrow$ (i). Since $\mathcal{L}^1$-$\lim_{n\to\infty} u_n = u$ exists we know that $\sup_{n\in\mathbb{N}_0} \int |u_n| d\mu < \infty$ (this follows as in Theorem 13.10). By the submartingale convergence theorem the pointwise limit $u_\infty = \lim_{n\to\infty} u_n$ exists a.e. On the other hand, by Corollary 13.8, $u = \lim_{k\to\infty} u_{n(k)}$ a.e. for some subsequence. This implies that $u = u_\infty$ a.e. and, in particular, that $u_\infty = \mathcal{L}^1$-$\lim_{n\to\infty} u_n$; this entails $\lim_{n\to\infty} \int_A u_n \, d\mu = \int_A u_\infty \, d\mu$ for all $A \in \mathscr{A}$. [✍] Since $(u_n)_{n\in\mathbb{N}_0}$ is a submartingale, we find for all $n > k$ and $A \in \mathscr{A}_k$

$$\int_A u_k \, d\mu \leqslant \int_A u_n \, d\mu \xrightarrow[n\to\infty]{} \int_A u_\infty \, d\mu,$$

proving that $(u_n)_{n\in\mathbb{N}_0\cup\{\infty\}}$ is also a submartingale. $\square$

Again, $\mathcal{L}^1$-convergence of backwards submartingales holds under much weaker assumptions.

Theorem 24.7 *Let $(w_\ell, \mathscr{A}_\ell)_{\ell\in-\mathbb{N}_0}$ be a backwards submartingale and assume that $\mu|_{\mathscr{A}_{-\infty}}$ is σ-finite. Then we have the following.*

(i) $\lim_{n\to+\infty} w_{-n} = w_{-\infty} \in [-\infty, \infty)$ *exists a.e.*
(ii) $\mathcal{L}^1$-$\lim_{n\to+\infty} w_{-n} = w_{-\infty}$ *if, and only if,* $\inf_{n\in\mathbb{N}_0} \int w_{-n} \, d\mu > -\infty$. *If this is the case, then $(w_\ell, \mathscr{A}_\ell)_{\ell\in-\mathbb{N}_0\cup\{-\infty\}}$ is a submartingale and $w_{-\infty}$ is a.e. real-valued.*

For a backwards martingale, the condition in (ii) *is automatically satisfied.*

Proof Part (i) has already been proved in Corollary 24.5. For part (ii) we start with the observation that for a backwards submartingale

$$\sup_{n\in\mathbb{N}_0} \int |w_{-n}| d\mu < \infty \iff \inf_{n\in\mathbb{N}_0} \int w_{-n} \, d\mu > -\infty \iff \lim_{n\to+\infty} \int w_{-n} \, d\mu \in \mathbb{R}.$$

Indeed: the second equivalence follows from the submartingale property,

$$\int w_{-n-1} \, d\mu \leqslant \int w_{-n} \, d\mu \leqslant \int w_0 \, d\mu,$$

while '$\Leftarrow$' of the first equivalence derives from the fact that $(w_\ell^+)_{\ell\in-\mathbb{N}_0}$ is again a submartingale, see Example 23.3(iv), and

$$\int |w_{-n}|d\mu = \int (2w_{-n}^+ - w_{-n})d\mu \leqslant 2\int w_{-1}^+\,d\mu - \int w_{-n}\,d\mu;$$

the other direction '$\Rightarrow$' is obvious. With exactly the same reasoning as in the proof of Theorem 24.6, (i)$\Rightarrow$(ii), we can show that $(w_\ell^+)_{\ell\in-\mathbb{N}_0}$ and $(w_\ell)_{\ell\in-\mathbb{N}_0}$ are uniformly integrable (of course, the function w_ϵ used as a bound for uniform integrability is now $\mathscr{A}_{-\infty}$-measurable). The submartingale property of $(w_\ell)_{\ell\in-\mathbb{N}_0\cup\{-\infty\}}$ follows literally with the same arguments as the corresponding assertion in (iii)$\Rightarrow$(i) of Theorem 24.6. $\square$

We close this chapter with a simple but far-reaching application of the backwards martingale convergence theorem.

Example 24.8 (Kolmogorov's strong law of large numbers) Let $(\Omega,\mathscr{A},\mathbb{P})$ be any probability space and $(\xi_n)_{n\in\mathbb{N}}$ a sequence of identically distributed independent random variables – that is, the $\xi_n:\Omega\to\mathbb{R}$ are measurable, independent functions (in the sense of Example 23.3(x) and Scholium 23.4) such that $\xi_n(\mathbb{P})=\xi_1(\mathbb{P})$ for all $n\in\mathbb{N}$ – the *strong law of large numbers* holds, i.e. the limit

$$\lim_{n\to\infty}\frac{1}{n}(\xi_1(\omega)+\cdots+\xi_n(\omega))$$

exists and is finite for a.e. $\omega\in\Omega$, if, and only if, the ξ_n are integrable. In this case, the limit is given by $\int\xi_1\,d\mathbb{P}$.

Sufficiency. Suppose the ξ_n are integrable. Then $\xi_n^\circ:=\xi_n-\int\xi_n\,d\mathbb{P}$ are again independent identically distributed random variables with zero mean: $\int\xi_n^\circ\,d\mathbb{P}=0$. If we set

$$S_n:=\xi_1^\circ+\xi_2^\circ+\cdots+\xi_n^\circ \quad\text{and}\quad \mathscr{A}_{-n}:=\sigma(S_n,S_{n+1},S_{n+2},\dots),$$

then $\left(\frac{1}{n}S_n,\mathscr{A}_{-n}\right)_{n\in\mathbb{N}}$ is a backwards martingale. *Indeed*, any function of the variables $(\xi_1^\circ,\xi_2^\circ,\dots,\xi_n^\circ,S_n)$ is independent of $(\xi_{n+1}^\circ,\xi_{n+2}^\circ,\dots)$, and (23.6) yields, for every set of the form $A=\bigcap_{i=1}^N\{\xi_{n+i}^\circ\in B_i\}\cap\{S_n\in B_0\}$, $B_0,\dots,B_N\in\mathscr{B}(\mathbb{R})$,

$N \in \mathbb{N}$, and all $k = 1, 2, \ldots, n$,

$$\begin{aligned}
\int_A \xi_k^\circ \, d\mathbb{P} &= \int_{\bigcap_{i=1}^N \{\xi_{n+i}^\circ \in B_i\}} \mathbb{1}_{\{S_n \in B_0\}} \xi_k^\circ \, d\mathbb{P} \\
&\overset{23.6}{=} \int_{\{S_n \in B_0\}} \xi_k^\circ \, d\mathbb{P} \cdot \mathbb{P}\Big(\bigcap_{i=1}^N \{\xi_{n+i}^\circ \in B_i\} \Big) \\
&= \int_{\{S_n \in B_0\}} \xi_1^\circ \, d\mathbb{P} \cdot \mathbb{P}\Big(\bigcap_{i=1}^N \{\xi_{n+i}^\circ \in B_i\} \Big),
\end{aligned}$$

where we use that the ξ_k° are identically distributed. Summing over $k = 1, \ldots, n$ gives

$$\int_A S_n \, d\mathbb{P} = n \int_{\{S_n \in B_0\}} \xi_1^\circ \, d\mathbb{P} \cdot \mathbb{P}\Big(\bigcap_{i=1}^N \{\xi_{n+i}^\circ \in B_i\} \Big) = n \int_A \xi_1^\circ \, d\mathbb{P}.$$

This means that $\int_A \xi_1^\circ \, d\mathbb{P} = \int_A \frac{1}{n} S_n \, d\mathbb{P}$ for all $n \in \mathbb{N}$ and all sets A from a generator of $\mathscr{A}_{-n}$ which clearly satisfies the conditions of Remark 23.2(i), proving that $\big(\frac{1}{n} S_n, \mathscr{A}_{-n}\big)_{n \in \mathbb{N}}$ is a backwards martingale. Theorem 24.7 now guarantees that

$$L := \lim_{n \to \infty} \frac{1}{n} S_n = \lim_{n \to \infty} \frac{1}{n^2} S_{n^2}$$

exists a.e. and in $\mathcal{L}^1$. It then remains to show that $L = 0$ a.e. Note that we have $\lim_{n \to \infty} S_n / n^2 = 0$ a.e.; since $e^{-|x|} \leqslant 1$ and since constants are integrable, the dominated convergence theorem (Theorem 12.2) and independence (23.7) show that

$$\begin{aligned}
\int \big(e^{-|L|}\big)^2 \, d\mathbb{P} &= \int \lim_{n \to \infty} \Big(\exp\Big[-\Big| \frac{1}{n} S_n \Big| \Big] \exp\Big[-\Big| \frac{1}{n^2} (S_{n^2} - S_n) \Big| \Big] \Big) d\mathbb{P} \\
&= \lim_{n \to \infty} \int \exp\Big[-\Big| \frac{1}{n} S_n \Big| \Big] \exp\Big[-\Big| \frac{1}{n^2} (S_{n^2} - S_n) \Big| \Big] d\mathbb{P} \\
&= \lim_{n \to \infty} \Big(\int \exp\Big[-\Big| \frac{1}{n} S_n \Big| \Big] d\mathbb{P} \cdot \int \exp\Big[-\Big| \frac{1}{n^2} (S_{n^2} - S_n) \Big| \Big] d\mathbb{P} \Big) \\
&= \Big(\int e^{-|L|} \, d\mathbb{P} \Big)^2.
\end{aligned}$$

Thus

$$\int \Big(e^{-|L|} - \textstyle\int e^{-|L|} \, d\mathbb{P} \Big)^2 d\mathbb{P} = \displaystyle\int \big(e^{-|L|}\big)^2 \, d\mathbb{P} - \Big(\int e^{-|L|} \, d\mathbb{P} \Big)^2 = 0,$$

and we conclude with Theorem 11.2(i) that $e^{-|L|} = \int e^{-|L|}\,d\mathbb{P}$ a.e.; as a consequence, L is almost everywhere constant. Using $L = L^1\text{-}\lim_{n\to\infty} S_n/n$, we get

$$L = \int L\,d\mathbb{P} = \lim_{n\to\infty} \underbrace{\int \frac{1}{n} S_n\,d\mathbb{P}}_{=0} = 0 \quad \text{a.e.}$$

Necessity. Suppose that the a.e. limit $L = \lim_{n\to\infty} \frac{1}{n}(\xi_1(\omega) + \cdots + \xi_n(\omega))$ exists and is finite. If all ξ_n were positive, we could argue as follows: the truncated random variables $\xi_n^c := \xi_n \wedge c$ are still independent and identically distributed. Since they are also integrable, the sufficiency direction of Kolmogorov's law shows that, for all $c > 0$,

$$\int \xi_1^c\,d\mathbb{P} = \lim_{n\to\infty} \frac{\xi_1^c + \cdots + \xi_n^c}{n} \leqslant \lim_{n\to\infty} \frac{\xi_1 + \cdots + \xi_n}{n} = L.$$

Letting $c \to \infty$, Beppo Levi's theorem (Theorem 9.6) proves $\int \xi_1\,d\mathbb{P} < \infty$.

Such a simple argument is not available in the general case. For this we need the converse or 'difficult' half of the Borel–Cantelli lemma.

Theorem 24.9 (Borel–Cantelli) *Let* $(\Omega, \mathscr{A}, \mathbb{P})$ *be a probability space and* $(A_n)_{n\in\mathbb{N}}$ *a sequence in* $\mathscr{A}$.

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}(\limsup_{n\to\infty} A_n) = 0;$$

if the sets A_n *are pairwise independent,*[1] *then*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \implies \mathbb{P}(\limsup_{n\to\infty} A_n) = 1.$$

Proof Recall that $\limsup_n A_n = \bigcap_k \bigcup_{n\geqslant k} A_n$. Thus we have $\omega \in \limsup_n A_n$ if, and only if, ω appears in infinitely many of the A_n. This shows that we have $\limsup_n A_n = \{\sum_{n=1}^{\infty} \mathbb{1}_{A_n} = \infty\}$.

The first of the two implications follows thus: by the Beppo Levi theorem for series, Corollary 9.9, we see that

$$\int \sum_{n=1}^{\infty} \mathbb{1}_{A_n}\,d\mathbb{P} = \sum_{n=1}^{\infty} \int \mathbb{1}_{A_n}\,d\mathbb{P} = \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty.$$

Corollary 11.6 then shows that $\sum_{n=1}^{\infty} \mathbb{1}_{A_n} < \infty$ a.e., and $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$ follows.

[1] That is, $\mathbb{P}(A_n \cap A_k) = \mathbb{P}(A_n)\mathbb{P}(A_k)$ for all $n \neq k$.

For the second implication we set $S_n := \sum_{i=1}^n \mathbb{1}_{A_i}$ and $S := \sum_{i=1}^\infty \mathbb{1}_{A_i}$. Then $m_n := \int S_n \, d\mathbb{P} = \sum_{i=1}^n \mathbb{P}(A_n)$ and, by pairwise independence,

$$\begin{aligned}\int (S_n - m_n)^2 \, d\mathbb{P} &= \sum_{i,k=1}^n \int (\mathbb{1}_{A_i} - \mathbb{P}(A_i))(\mathbb{1}_{A_k} - \mathbb{P}(A_k)) d\mathbb{P} \\ &= \sum_{i=1}^n \int (\mathbb{1}_{A_i} - \mathbb{P}(A_i))^2 \, d\mathbb{P} \\ &= \sum_{i=1}^n \mathbb{P}(A_i)(1 - \mathbb{P}(A_i)) \leqslant m_n.\end{aligned}$$

Since $S_n \leqslant S$, we can use Markov's inequality (11.4) to get

$$\begin{aligned}\mathbb{P}\big(S \leqslant \tfrac{1}{2} m_n\big) \leqslant \mathbb{P}\big(S_n \leqslant \tfrac{1}{2} m_n\big) &= \mathbb{P}\big(S_n - m_n \leqslant -\tfrac{1}{2} m_n\big) \\ &\leqslant \mathbb{P}\big(|S_n - m_n| \geqslant \tfrac{1}{2} m_n\big) = \mathbb{P}\big((S_n - m_n)^2 \geqslant \tfrac{1}{4} m_n^2\big) \\ &\leqslant \frac{4}{m_n^2} \int (S_n - m_n)^2 \, d\mathbb{P} \leqslant \frac{4}{m_n}.\end{aligned}$$

By assumption $\lim_{n\to\infty} m_n = \infty$, so $\mathbb{P}(S < \infty) = \lim_{n\to\infty} \mathbb{P}(S \leqslant \frac{1}{2} m_n) = 0$. □

24.8 Example (continued) We can now continue with the proof of the *necessity* part of Kolmogorov's strong law of large numbers. Since the a.e. limit exists, we get

$$\frac{\xi_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1} \xrightarrow[n\to\infty]{} 0,$$

which shows that $\omega \in A_n := \{|\xi_n| \geqslant n\}$ happens only for finitely many n. In other words, $\mathbb{P}(\sum_{j=1}^\infty \mathbb{1}_{A_j} = \infty) = 0$; since the A_n are independent, the Borel–Cantelli lemma shows that $\sum_{j=1}^\infty \mathbb{P}(A_j) < \infty$. Thus

$$|\xi_1| \leqslant 1 + \sum_{n=1}^{|\xi_1|} 1 = 1 + \sum_{n=1}^{\infty} \mathbb{1}_{\{n \leqslant |\xi_1|\}}.$$

If we integrate this inequality w.r.t. $\mathbb{P}$ and use Beppo Levi's theorem, we arrive at

$$\int |\xi_1| d\mathbb{P} \leqslant 1 + \sum_{n=1}^\infty \mathbb{P}\{|\xi_1| \geqslant n\} = 1 + \sum_{n=1}^\infty \mathbb{P}\{|\xi_n| \geqslant n\},$$

since ξ_1 and ξ_n have the same distribution. □

We will see more applications of the martingale convergence theorems in the following chapters.

Problems

Unless stated otherwise, $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ will be a σ-finite filtered measure space.

24.1. Verify that the random times σ_k and τ_k defined in the proof of Lemma 24.1 are stopping times.

24.2. Let $(\mathscr{A}_{-n})_{n\in\mathbb{N}}$ be a decreasing filtration such that $\mu|_{\mathscr{A}_{-\infty}}$ is σ-finite. Assume that $(u_{-n}, \mathscr{A}_{-n})_{n\in\mathbb{N}}$ is a backwards supermartingale which converges a.e. to a real-valued function $u_{-\infty} \in \mathcal{L}^1(\mu)$ which closes the supermartingale to the left, i.e. such that $(u_{-n}, \mathscr{A}_{-n})_{n\in\mathbb{N}\cup\{\infty\}}$ is still a supermartingale. Then

$$\lim_{n\to\infty} \int u_{-n}\, d\mu = \int u_{-\infty}\, d\mu.$$

24.3. Let $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ be a supermartingale such that $u_n \geqslant 0$ and $\lim_{n\to\infty} \int u_n\, d\mu = 0$. Then $u_n \to 0$ pointwise a.e. and in $\mathcal{L}^1$.

Remark: Positive supermartingales with $\lim_{n\to\infty} \int u_n\, d\mu = 0$ are called *potentials*.

24.4. Let $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ be a martingale. If $\mathcal{L}^1$-$\lim_{n\to\infty} u_n$ exists, then the pointwise limit $\lim_{n\to\infty} u_n(x)$ exists for almost every x.

24.5. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. Find a martingale $(u_n)_{n\in\mathbb{N}}$ for which $0 < \mathbb{P}(u_n \text{ converges}) < 1$.

[Hint: take a sequence $(\xi_n)_{n\in\mathbb{N}_0}$ of independent Bernoulli $(\frac{1}{2}, \frac{1}{2})$-distributed random variables with values ± 1; try $u_n := \frac{1}{2}(\xi_0 + 1)(\xi_1 + \xi_2 + \cdots + \xi_n)$.]

24.6. The following exercise furnishes an example of a martingale $(M_n)_{n\in\mathbb{N}}$ on the probability space $\big([0,1], \mathscr{B}[0,1], \lambda = \lambda^1|_{[0,1]}\big)$ such that λ-$\lim_{n\to\infty} M_n$ exists but the pointwise limit $\lim_{n\to\infty} M_n(x)$ doesn't. Compare this with Problem 24.4.

(i) Construct a sequence $(\xi_n)_{n\in\mathbb{N}}$ of independent, identically Bernoulli-distributed random variables with $\lambda\{\xi_1 = 1\} = \lambda\{\xi_1 = -1\} = \frac{1}{2}$.

(ii) Let $\mathscr{A}_n = \sigma(\xi_1, \ldots, \xi_{n^2})$. Show that $A_n := \{\xi_{(n-1)^2+2} + \cdots + \xi_{n^2} = 0\}$ is for each $n \in \mathbb{N}$, $n \geqslant 2$, contained in $\mathscr{A}_n$ and

$$\lim_{n\to\infty} \lambda(A_n) = 0 \quad \text{and} \quad \lambda(\limsup_{n\to\infty} A_n) = 1.$$

Conclude that the set of all x for which $\lim_n \mathbb{1}_{A_n}(x)$ exists is a null set.

[Hint: use the 'difficult' direction of the Borel–Cantelli lemma (Theorem 24.9). Moreover, Stirling's formula $n! \sim \sqrt{2\pi n}(n/e)^n$ might come in handy.]

(iii) The sequence $M_2 := 0$ and $M_{n+1} := M_n(1 + \xi_{n^2+1}) + \mathbb{1}_{A_n}\xi_{n^2+1}$, $n \in \mathbb{N}$, $n \geqslant 2$, defines a martingale $(M_n, \mathscr{A}_n)_{n\geqslant 1}$.

(iv) Show that $\lambda(M_{n+1} \neq 0) \leqslant \frac{1}{2}\lambda(M_n \neq 0) + \lambda(A_n)$.

(v) Show that for every $x \in \{\lim_n M_n \text{ exists}\}$ the limit $\lim_n \mathbb{1}_{A_n}(x)$ exists, too. Conclude that $\lim_n \lambda(M_n = 0) = 1$ and that $\lambda\{\lim_n M_n \text{ exists}\} = 0$.

24.7. Consider the probability space $(\mathbb{N}, \mathscr{P}(\mathbb{N}), \mathbb{P})$ with

$$\mathbb{P}\{n\} := \frac{1}{n} - \frac{1}{n+1}.$$

Set

$$\mathscr{A}_n := \sigma\big(\{1\}, \{2\}, \ldots, \{n\}, [n+1, \infty) \cap \mathbb{N}\big)$$

and show that $\xi_n := (n+1)\mathbb{1}_{[n+1,\infty)\cap\mathbb{N}}$, $n \in \mathbb{N}$, is a positive martingale such that $\int \xi_n\, d\mathbb{P} = 1$, $\lim_{n\to\infty} \xi_n = 0$ but $\sup_{n\in\mathbb{N}} \xi_n = \infty$.

24.8. $\mathcal{L}^2$**-bounded martingales.** A martingale $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ is called $\mathcal{L}^2$-*bounded*, if the $\mathcal{L}^2$-norms are bounded: $\sup_{n\in\mathbb{N}} \int u_n^2 \, d\mu < \infty$. For ease of notation set $u_0 := 0$.

(i) Show that $(u_n)_{n\in\mathbb{N}}$ is $\mathcal{L}^2$-bounded if, and only if,

$$\sum_{n=1}^{\infty} \int (u_n - u_{n-1})^2 \, d\mu < \infty.$$

[Hint: use Problem 23.6.]

Assume from now on that $(u_n)_{n\in\mathbb{N}}$ is $\mathcal{L}^2$-bounded.

(ii) Show that $\lim_{n\to\infty} u_n = u$ exists a.e.
[Hint: $(\mathbb{1}_K u_i)_{i\in\mathbb{N}}$ is an L^1-bounded martingale for any $K \in \mathscr{A}_0$ such that $\mu(K) < \infty$.]

(iii) Show that

$$\lim_{n\to\infty} \int (u - u_n)^2 \, d\mu = 0.$$

[Hint: check that $\int (u_{n+k} - u_n)^2 \, d\mu = \sum_{\ell=n+1}^{n+k} \int (u_\ell - u_{\ell-1})^2 \, d\mu$ and apply Fatou's lemma.]

(iv) Assume now that $\mu(X) < \infty$. Show that $(u_n)_{n\in\mathbb{N}}$ is uniformly integrable, that $u_n \to u$ in $\mathcal{L}^1$ and that $u_\infty := u$ closes the martingale to the right, i.e. that $(u_n)_{n\in\mathbb{N}\cup\{\infty\}}$ is again a martingale.

24.9. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space.

(i) Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of independent identically Bernoulli $(\frac{1}{2}, \frac{1}{2})$-distributed random variables with values ± 1. Show that for any sequence $(y_n)_{n\in\mathbb{N}}$

$$\sum_{n=1}^{\infty} y_n^2 < \infty \iff \sum_{n=1}^{\infty} \varepsilon_n y_n$$

converges a.e.

(ii) Generalize (i) to a sequence of independent random variables $(\xi_n)_{n\in\mathbb{N}}$ with zero mean $\int \xi_n \, d\mathbb{P} = 0$ and finite variances $\int \xi_n^2 \, d\mathbb{P} = \sigma_n^2 < \infty$ and prove that

$$\sum_{n=1}^{\infty} \sigma_n^2 < \infty \implies \sum_{n=1}^{\infty} \xi_n$$

converges a.e.
[Hint: consider the martingale $S_n = \xi_1 + \cdots + \xi_n$ and use Problem 24.8.]

(iii) If $|\xi_n| \leqslant C$ for all $n \in \mathbb{N}$, the converse of (ii) is also true, i.e.

$$\sum_{n=1}^{\infty} \sigma_n^2 < \infty \iff \sum_{n=1}^{\infty} \xi_n$$

converges a.e.
[Hint: show that $M_n := (\xi_1 + \cdots + \xi_n)^2 - (\sigma_1^2 + \cdots + \sigma_n^2) =: S_n^2 - A_n$ is a martingale, use the optional sampling theorem from Remark 23.8 for M_n with $\tau_\kappa := \inf\{n : |M_n| > \kappa\}$, and observe that $|M_{\tau\wedge n_\kappa}| \leqslant C + \kappa$ and that $\int A_{\tau\wedge n_\kappa} \, d\mathbb{P} \leqslant (K + c)^2$.]

25

Martingales in Action

After our venture into the theory of martingales we want to apply martingales to continue the development of measure and integration theory. The central topics of this chapter are martingale proofs of

- the Radon–Nikodým theorem (Theorem 25.2),
- the Hardy–Littlewood maximal theorem (Theorem 25.17),
- Lebesgue's differentiation theorem (Theorem 25.20).

For the last two we need (maximal) inequalities for martingales. These will be treated in a short interlude which is also of independent interest. We will also include the classical proof of Lebesgue's decomposition theorem based on the Radon–Nikodým theorem.

The Radon–Nikodým Theorem

Let $(X, \mathscr{A}, \mu)$ be a measure space. We saw in Lemma 10.8 that for any function $f \in \mathcal{L}^1_+(\mathscr{A})$ – or, indeed, for any $f \in \mathcal{M}^+(\mathscr{A})$ – the set function $\nu := f\mu$ given by $\nu(A) := \int_A f(x)\mu(dx)$ is again a measure. From Theorem 11.2(ii) we know that

$$N \in \mathscr{A},\ \mu(N) = 0 \implies \nu(N) = 0. \tag{25.1}$$

This observation motivates the following definition.

Definition 25.1 Let μ, ν be two measures on the measurable space $(X, \mathscr{A})$. If (25.1) holds, we call ν *absolutely continuous* w.r.t. μ and write $\nu \ll \mu$.

Measures with densities are always absolutely continuous w.r.t. their base measure: $f\mu \ll \mu$. Remarkably, the converse is also true.

Theorem 25.2 (Radon–Nikodým) *Let μ, ν be two measures on the measurable space $(X, \mathscr{A})$. If μ is σ-finite, then the following assertions are equivalent:*

(i) $\nu(A) = \int_A f(x)\,\mu(dx)$ *for some a.e. unique $f \in \mathcal{M}(\mathscr{A})$, $f \geqslant 0$;*

(ii) $\nu \ll \mu$.

The unique function f is called the Radon–Nikodým derivative and it is traditionally denoted by $f = d\nu/d\mu$.

Above we have just verified that (i) $\Rightarrow$ (ii). The converse direction is less obvious and we want to use a martingale argument for its proof. For this we need a few more preparations which extend the notion of a martingale to directed index sets.

Let $(I, \leqslant)$ be any partially ordered index set. We call I *upwards filtering* or *upwards directed* if

$$\forall \alpha, \beta \in I \quad \exists \gamma \in I : \alpha \leqslant \gamma,\ \beta \leqslant \gamma. \tag{25.2}$$

A family $(\mathscr{A}_\alpha)_{\alpha \in I}$ of sub-σ-algebras of $\mathscr{A}$ is called a *filtration* if

$$\forall \alpha, \beta \in I,\ \alpha \leqslant \beta\ :\ \mathscr{A}_\alpha \subset \mathscr{A}_\beta;$$

as before, we set $\mathscr{A}_\infty := \sigma\big(\bigcup_{\alpha \in I} \mathscr{A}_\alpha\big)$, and we treat ∞ as the biggest element of $I \cup \{\infty\}$, i.e. $\alpha < \infty$ for all $\alpha \in I$. If a σ-algebra $\mathscr{A}_0 \subset \mathscr{A}_\alpha$ for all $\alpha \in I$ and if $\mu|_{\mathscr{A}_0}$ is σ-finite, we call $(X, \mathscr{A}, \mathscr{A}_\alpha, \mu)$ a *σ-finite filtered measure space*.

Definition 25.3 Let $(X, \mathscr{A}, \mathscr{A}_\alpha, \mu)$ be a σ-finite filtered measure space. A family of measurable functions $(u_\alpha)_{\alpha \in I}$ is called a *martingale* (w.r.t. the filtration $(\mathscr{A}_\alpha)_{\alpha \in I}$), if $u_\alpha \in \mathcal{L}^1(\mathscr{A}_\alpha)$ for each $\alpha \in I$ and if

$$\int_A u_\beta\, d\mu = \int_A u_\alpha\, d\mu \quad \forall \alpha \leqslant \beta,\ \forall A \in \mathscr{A}_\alpha. \tag{25.3}$$

Uncountable, partially ordered index sets are not at all artificial. This is shown by the following example which will be essential for the proof of Theorem 25.2.

Example 25.4 Let $(X, \mathscr{A}, \mu)$ be a finite measure space and assume that ν is a measure such that $\nu \ll \mu$. Set

$$I := \left\{ \alpha = \{A_1, A_2, \dots, A_n\} : n \in \mathbb{N},\ A_i \in \mathscr{A} \text{ and } \dot{\bigcup_{i=1}^n} A_i = X \right\}$$

and define an order relation '$\leqslant$' on I through

$$\alpha \leqslant \alpha' \iff \forall A \in \alpha : A = A'_1 \,\dot\cup\, \dots \,\dot\cup\, A'_\ell, \text{ where } A'_i \in \alpha',\ \ell \in \mathbb{N}.$$

Since the common refinement β of any two elements $\alpha, \alpha' \in I$,

$$\beta := \{A \cap A' : A \in \alpha, A' \in \alpha'\},$$

is again in I and satisfies $\alpha \leqslant \beta$ and $\alpha' \leqslant \beta$, it is clear that $(I, \leqslant)$ is upwards filtering. In particular,

$$(\mathscr{A}_\alpha)_{\alpha \in I}, \quad \text{where} \quad \mathscr{A}_\alpha := \sigma(A : A \in \alpha),$$

is a filtration as $\mathscr{A}_\alpha \subset \mathscr{A}_{\alpha'}$ whenever $\alpha \leqslant \alpha'$. Moreover, $(f_\alpha, \mathscr{A}_\alpha)_{\alpha \in I}$ defined by

$$f_\alpha := \sum_{A \in \alpha} \frac{\nu(A)}{\mu(A)} \mathbb{1}_A, \qquad \left(\frac{\nu(A)}{\mu(A)} := 0 \quad \text{if} \quad \mu(A) = 0 \right)$$

is a martingale. *Indeed*, if $\alpha \leqslant \beta$, $\alpha, \beta \in I$, then

$$\int_A f_\alpha \, d\mu = \frac{\nu(A)}{\mu(A)} \mu(A) = \left\{ \begin{array}{ll} \nu(A) & \text{if } \mu(A) > 0 \\ 0 & \text{if } \mu(A) = 0 \end{array} \right\} = \nu(A)$$

as $\nu \ll \mu$. Similarly, for $A \in \alpha$ with $A = B_1 \uplus \cdots \uplus B_\ell$ and $B_1, \ldots, B_\ell \in \beta$

$$\begin{aligned} \int_A f_\beta \, d\mu = \sum_{i=1}^{\ell} \int_{B_i} f_\beta \, d\mu &= \sum_{i=1}^{\ell} \frac{\nu(B_i)}{\mu(B_i)} \mu(B_i) \\ &= \sum_{i:\mu(B_i)>0} \nu(B_i) \\ &\overset{(*)}{=} \sum_{i=1}^{\ell} \nu(B_i) \;=\; \nu(A), \end{aligned}$$

where we use in $(*)$ that $\nu \ll \mu$, i.e. $\nu(B_i) = 0$ if $\mu(B_i) = 0$. Thus we have $\int_A f_\alpha \, d\mu = \int_A f_\beta \, d\mu$ for all $A \in \alpha$, hence on $\mathscr{A}_\alpha$, since all $A \in \alpha$ are disjoint and generate $\mathscr{A}_\alpha$[✍], see also Remark 23.2(i).

What Example 25.4 really says is that

$$\nu(A) = \int_A f_\alpha \, d\mu \quad \forall A \in \mathscr{A}_\alpha, \tag{25.4}$$

or $\nu|_{\mathscr{A}_\alpha} \ll \mu|_{\mathscr{A}_\alpha}$ and $d(\nu|_{\mathscr{A}_\alpha})/d(\mu|_{\mathscr{A}_\alpha}) = f_\alpha$. Heuristically we should expect that, if $f_\alpha \xrightarrow[\alpha \to \infty]{} f_\infty$ exists, f_∞ is the Radon–Nikodým derivative $d\nu/d\mu = f_\infty$.

In order to make this idea rigorous, we need the notion of $\mathcal{L}^1$-convergence *along an upwards filtering set*, which is slightly more complicated than for the index set $\mathbb{N}$. We say

$$u = \mathcal{L}^1\text{-}\lim_{\alpha \in I} u_\alpha \iff \forall \epsilon > 0 \; \exists \gamma_\epsilon \in I \; \forall \alpha \geqslant \gamma_\epsilon : \|u - u_\alpha\|_1 < \epsilon.$$

We can now extend Theorem 24.6.

Theorem 25.5 *Let I be an upwards filtering index set, $(X, \mathscr{A}, \mathscr{A}_\alpha, \mu)$ be a σ-finite measure space and $(u_\alpha, \mathscr{A}_\alpha)_{\alpha\in I}$ be a martingale. Then the following assertions are equivalent.*

(i) *There exists a unique $u_\infty \in \mathcal{L}^1(\mathscr{A}_\infty)$ such that $(u_\alpha, \mathscr{A}_\alpha)_{\alpha\in I\cup\{\infty\}}$ is a martingale. In this case $u_\infty = \mathcal{L}^1\text{-}\lim_{\alpha\in I} u_\alpha$.*

(ii) *$(u_\alpha, \mathscr{A}_\alpha)_{\alpha\in I}$ is uniformly integrable.*

Proof (i)$\Rightarrow$(ii) (compare with Theorem 24.6). Denote by $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}_0$ a sequence such that $A_n \uparrow X$ and $\mu(A_n) < \infty$. Since $u_\infty \in \mathcal{L}^1(\mathscr{A}_\infty)$, we find for every $\epsilon > 0$ some $\kappa > 0$ and $N \in \mathbb{N}$ such that

$$\int_{\{|u_\infty|>\kappa\}} |u_\infty|\,d\mu + \int_{A_n^c} |u_\infty|\,d\mu \leqslant \epsilon \quad \forall n \geqslant N.$$

The function $w(x) := \sum_{n\in\mathbb{N}} 2^{-n}(1+\mu(A_n))^{-1}\mathbb{1}_{A_n}(x)$ is in $\mathcal{L}^1(\mathscr{A}_0)$, $w > 0$ and, as $(|u_\alpha|)_{\alpha\in I\cup\{\infty\}}$ is a submartingale (see Example 23.3(v)), we find for every $L > 0$

$$\begin{aligned}
\sup_{\alpha\in I} \int_{\{|u_\alpha|>Lw\}} |u_\alpha|\,d\mu &\leqslant \sup_{\alpha\in I} \int_{\{|u_\alpha|>Lw\}} |u_\infty|\,d\mu \\
&\leqslant \sup_{\alpha\in I} \int_{\{|u_\alpha|>Lw\}\cap A_N\cap\{|u_\infty|\leqslant\kappa\}} |u_\infty|\,d\mu + \int_{A_N^c} |u_\infty|\,d\mu + \int_{\{|u_\infty|>\kappa\}} |u_\infty|\,d\mu \\
&\leqslant \kappa \sup_{\alpha\in I} \mu\big\{|u_\alpha| > L2^{-N}(1+\mu(A_N))^{-1}\big\} + \epsilon.
\end{aligned}$$

For the last step we use that $w(x) \geqslant 2^{-N}(1+\mu(A_N))^{-1}$ for $x \in A_N$. By Markov's inequality (11.3) and the submartingale property we get

$$\begin{aligned}
\sup_{\alpha\in I} \int_{\{|u_\alpha|>Lw\}} |u_\alpha|\,d\mu &\leqslant \kappa \frac{2^N(1+\mu(A_N))}{L} \sup_{\alpha\in I} \int |u_\alpha|\,d\mu + \epsilon \\
&\leqslant \frac{\kappa 2^N(1+\mu(A_N))}{L} \int |u_\infty|\,d\mu + \epsilon,
\end{aligned}$$

and (ii) follows since we can choose $L > 0$ as large as we want.

(ii)$\Rightarrow$(i). *Step 1. Uniqueness.* Assume that $u, w \in \mathcal{L}^1(\mathscr{A}_\infty)$ are two functions which close the martingale $(u_\alpha)_{\alpha\in I}$, i.e. functions satisfying

$$\int_A u\,d\mu = \int_A w\,d\mu = \int_A u_\alpha\,d\mu \quad \forall A \in \mathscr{A}_\alpha,\ \alpha \in I.$$

Since u and w are integrable functions, the family

$$\Sigma := \left\{ A \in \mathscr{A}_\infty : \int_A u\,d\mu = \int_A w\,d\mu \right\}$$

is a σ-algebra which satisfies $\bigcup_{\alpha\in I}\mathscr{A}_\alpha \subset \Sigma \subset \mathscr{A}_\infty$. Since $\mathscr{A}_\infty$ is generated by the $\mathscr{A}_\alpha$, we get $\Sigma = \mathscr{A}_\infty$, which means that $\int_A u\,d\mu = \int_A w\,d\mu$ holds for all $A\in\mathscr{A}_\infty$. Now Corollary 11.7 applies and we get $u = w$ almost everywhere.

Step 2. Existence of the limit. We claim that

$$\forall\epsilon > 0\ \exists\gamma_\epsilon \in I\ \forall\alpha,\beta \geqslant \gamma_\epsilon : \int |u_\alpha - u_\beta|\,d\mu < \epsilon. \tag{25.5}$$

Otherwise there is a sequence $(\alpha_n)_{n\in\mathbb{N}} \subset I$ such that $\int |u_{\alpha_{n+1}} - u_{\alpha_n}|\,d\mu > \epsilon$ for all $n\in\mathbb{N}$. Since I is upwards filtering, we can assume that $(\alpha_n)_{n\in\mathbb{N}}$ is an increasing sequence. [✍] Because of (ii), $(u_{\alpha_n}, \mathscr{A}_{\alpha_n})_{n\in\mathbb{N}}$ is a uniformly integrable martingale with index set $\mathbb{N}$ which is, by construction, not an $\mathcal{L}^1$-Cauchy sequence. This contradicts Theorem 24.6.

We will now prove the existence of the $\mathcal{L}^1$-limit. Pick in (25.5) $\epsilon = \frac{1}{n}$ and choose $\gamma_{1/n}$. Since I is upwards directed, we can assume that $\gamma_{1/n}$ increases as $n\to\infty$; [✍] thus $(u_{\gamma_{1/n}})_{n\in\mathbb{N}} \subset \mathcal{L}^1(\mathscr{A}_\infty)$ is an $\mathcal{L}^1$-Cauchy sequence. By Theorem 24.6 it converges in $\mathcal{L}^1(\mathscr{A}_\infty)$ and a.e. to some $u_\infty := \lim_{n\to\infty} u_{\gamma_{1/n}} \in \mathcal{L}^1(\mathscr{A}_\infty)$. Moreover, for all $A\in\mathscr{A}_\infty$ and $\alpha > \gamma_{1/n}$ we have

$$\int_A |u_\alpha - u_\infty|\,d\mu \leqslant \underbrace{\int_A |u_\alpha - u_{\gamma_{1/n}}|\,d\mu}_{\leqslant 1/n \text{ by } (25.5)} + \int_A |u_{\gamma_{1/n}} - u_\infty|\,d\mu \leqslant \frac{2}{n}.$$

This shows, in particular, that $u_\alpha \mathbb{1}_A \xrightarrow{\mathcal{L}^1} u_\infty \mathbb{1}_A$ for all $A\in\mathscr{A}_\infty$, and, in view of Step 1, u_∞ is the only possible limit. The argument that we used in (iii) $\Rightarrow$ (i) of Theorem 24.6 now yields that $(u_\alpha)_{\alpha\in I\cup\{\infty\}}$ is still a martingale. $\square$

Caution Theorem 25.5 does not claim that $u_\alpha \xrightarrow{\text{a.e. along } I} u_\infty$. This is, in general, *false* for non-linearly ordered index sets I, see e.g. Dieudonné [12].

Proof of Theorem 25.2 (ii) $\Rightarrow$ (i) We assume that μ and ν are finite measures.

Denote by $(f_\alpha, \mathscr{A}_\alpha)_{\alpha\in I}$ the martingale of Example 25.4. It is enough to show that

$$f_\infty = \mathcal{L}^1\text{-}\lim_{\alpha\in I} f_\alpha \text{ exists and } \mathscr{A} = \mathscr{A}_\infty. \tag{25.6}$$

Indeed, (25.6) combined with (25.4) implies

$$\nu(A) = \int_A f_\infty\,d\mu \quad \forall A \in \bigcup_{\alpha\in I}\mathscr{A}_\alpha,$$

and the uniqueness theorem for measures (Theorem 5.7) extends this equality to $\mathscr{A}_\infty = \sigma\big(\bigcup_{\alpha\in I}\mathscr{A}_\alpha\big)$. Since $A\in\mathscr{A}$ is trivially contained in $\mathscr{A}_\alpha$ where $\alpha :=$

$\{A, A^c\}$ – at this point we use the finiteness of the measure μ – we see

$$\mathscr{A} \supset \mathscr{A}_\infty = \sigma\Big(\bigcup_{\alpha\in I}\mathscr{A}_\alpha\Big) \supset \bigcup_{\alpha\in I}\mathscr{A}_\alpha \supset \mathscr{A},$$

and all that remains is to prove the existence of the limit in (25.6). In view of Theorem 25.5 we have to show that $(f_\alpha, \mathscr{A}_\alpha)_{\alpha\in I}$ is uniformly integrable.

We claim that $\sup_{\alpha\in I}\nu\{f_\alpha > R\} \leqslant \epsilon$ for all large enough $R = R_\epsilon > 0$. Otherwise we could find some $\epsilon_0 > 0$ and indices α_n such that $\nu\{f_{\alpha_n} > 2^n\} > \epsilon_0$ for all $n \in \mathbb{N}$. Because of Fatou's lemma (for measures, Problem 9.11) or with the continuity of measures (Proposition 4.3),

$$\nu\Big[\bigcap_{i\in\mathbb{N}}\bigcup_{n\geqslant i}\{f_{\alpha_n} > 2^n\}\Big] = \nu\Big[\limsup_{n\to\infty}\{f_{\alpha_n} > 2^n\}\Big] \geqslant \limsup_{n\to\infty}\nu\{f_{\alpha_n} > 2^n\} > 0.$$

Since ν is a finite measure, we find for every $i \in \mathbb{N}$

$$\mu\Big[\bigcap_{i\in\mathbb{N}}\bigcup_{n\geqslant i}\{f_{\alpha_n} > 2^n\}\Big] \overset{4.3}{\leqslant} \sum_{n\geqslant i}\mu\{f_{\alpha_n} > 2^n\} \overset{11.5}{\leqslant} \sum_{n\geqslant i}\frac{1}{2^n}\underbrace{\int f_{\alpha_n}\,d\mu}_{=\nu(X)} = \frac{2\nu(X)}{2^i}.$$

Letting $i \to \infty$, this tends to zero, contradicting the fact that $\nu \ll \mu$. Finally,

$$\int_{\{|f_\alpha|>R\}} |f_\alpha|\,d\mu = \int_{\{f_\alpha>R\}} f_\alpha\,d\mu = \nu\{f_\alpha > R\} \leqslant \epsilon$$

if $R = R_\epsilon > 0$ is sufficiently large, and uniform integrability follows since the constant function $R \in \mathcal{L}^1(\mu)$. The uniqueness of f_∞ follows also from Theorem 25.5.

If μ is σ-finite and ν arbitrary, we use Steps 6 and 7 of the proof of Theorem 20.2, page 229 *et seq.* □

Corollary 25.6 *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and $\nu = f\mu$. Then*

(i) $\nu(X) < \infty \iff f \in \mathcal{L}^1(\mu)$;
(ii) *ν is σ-finite* $\iff \mu\{f = \infty\} = 0$.

Proof The first assertion (i) is obvious. For (ii) assume first that $\mu\{f = \infty\} = 0$. Since μ is σ-finite, we find an exhausting sequence $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ with $A_n \uparrow X$ and $\mu(A_n) < \infty$. The sets

$$B_k := \{0 \leqslant f \leqslant k\}, \quad B_\infty := \{f = \infty\}$$

obviously satisfy $\bigcup_{k\in\mathbb{N}}(B_k \cup B_\infty) = X$ as well as $\nu(B_\infty) = 0$ and

$$\nu(B_k \cap A_n) = \int_{B_k\cap A_n} f\,d\mu \leqslant k\int_{A_n} d\mu = k\mu(A_n) < \infty.$$

This shows that $(A_n \cap (B_k \cup B_\infty))_{k,n\in\mathbb{N}}$ is an exhausting sequence for ν, proving that ν is σ-finite.

Conversely, let ν be σ-finite and assume that $\mu\{f=\infty\} > 0$. As we can find *one* exhausting sequence $(C_k)_{k\in\mathbb{N}} \subset \mathscr{A}$ for both μ and ν, [✍] we see that

$$\{f=\infty\} = \bigcup_{k\in\mathbb{N}} \big(\{f=\infty\} \cap C_k\big) \supset \{f=\infty\} \cap C_{k_0}$$

for some fixed $k_0 \in \mathbb{N}$ with $\mu(C_{k_0}) > 0$. But then

$$\nu(C_{k_0}) \geqslant \int_{\{f=\infty\}\cap C_{k_0}} f\,d\mu = \infty,$$

which is impossible. □

It is clear that not all measures are absolutely continuous with respect to each other. In some sense, the next notion is the opposite of absolute continuity.

Definition 25.7 Two measures μ, ν on a measurable space $(X, \mathscr{A})$ are called (mutually) *singular* if there is a set $N \in \mathscr{A}$ such that $\nu(N) = 0 = \mu(N^c)$. We write in this case $\mu \perp \nu$ (or $\nu \perp \mu$ as '$\perp$' is symmetric).

Example 25.8 Let $(X, \mathscr{A}) = (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. Then

(i) $\delta_x \perp \lambda^n$ for all $x \in \mathbb{R}^n$;
(ii) $f\mu \perp g\mu$ if $\operatorname{supp} f \cap \operatorname{supp} g = \emptyset$.[1]

The measures μ and ν are singular, if they have disjoint 'supports', that is, if μ lives in a region of X which is not charged by ν and vice versa. In this sense, Example 25.8(ii) is the model case for singular measures. In general, however, two measures are neither purely absolutely continuous nor purely singular, but are a mixture of both.

Theorem 25.9 (Lebesgue decomposition) *Let μ, ν be two σ-finite measures on a measurable space $(X, \mathscr{A})$. Then there exists a unique (up to null sets) decomposition $\nu = \nu^\circ + \nu^\perp$, where $\nu^\circ \ll \mu$ and $\nu^\perp \perp \mu$.*

Proof Obviously $\mu + \nu$ is still a σ-finite measure, [✍] and $\nu \ll (\nu + \mu)$. In this situation Theorem 25.2 applies and shows that

$$\nu = f \cdot (\mu + \nu) = f \cdot \mu + f \cdot \nu. \tag{25.7}$$

[1] Recall that $\operatorname{supp} f := \overline{\{f \neq 0\}}$.

For any $\epsilon > 0$ we conclude, in particular, that

$$\nu\{f \geqslant 1+\epsilon\} = \int_{\{f\geqslant 1+\epsilon\}} f\, d(\mu+\nu)$$
$$\geqslant (1+\epsilon)\mu\{f \geqslant 1+\epsilon\} + (1+\epsilon)\nu\{f \geqslant 1+\epsilon\},$$

i.e. $\mu\{f \geqslant 1+\epsilon\} = \nu\{f \geqslant 1+\epsilon\} = 0$ for all $\epsilon > 0$, and this implies that we have $\mu\{f > 1\} = \nu\{f > 1\} = 0$. Without loss of generality we may therefore assume that $0 \leqslant f \leqslant 1$. In this case (25.7) can be rewritten as

$$(1-f)\nu = f\mu, \tag{25.8}$$

and on the set $N := \{f = 1\}$ we have

$$\mu(N) = \int_{\{f=1\}} d\mu = \int_{\{f=1\}} f\, d\mu \overset{(25.8)}{=} \int_{\{f=1\}} (1-f)\, d\nu = 0.$$

Therefore, $\mu \perp \nu^{\perp}$, where $\nu^{\perp} := \nu(\cdot \cap \{f = 1\})$, and for $\nu^{\circ} := \nu(\cdot \cap \{f < 1\})$ we get from (25.8)

$$\nu^{\circ}(A) = \nu(A \cap \{f < 1\}) = \int_{A\cap\{f<1\}} d\nu = \int_{A\cap\{f<1\}} \frac{f}{1-f}\, d\mu \quad \forall A \in \mathscr{A},$$

showing that $\nu^{\circ} \ll \mu$.

The uniqueness (up to null sets) of this decomposition follows directly from the uniqueness of the Radon–Nikodým derivative $f/(1-f)\mathbb{1}_{\{f<1\}}$. □

Remark 25.10 We have used the martingale convergence theorem to prove the Radon–Nikodým theorem. But the connection between these two theorems is much deeper. For measures with values in a Banach space (‘*vector measures*’) the Radon–Nikodým theorem holds if, and only if, the pointwise martingale convergence theorem is valid. One should add that the Radon–Nikodým theorem for Banach spaces is intimately connected with the geometry of Banach spaces. Note, however, that the techniques required in the theory of vector measures are distinctly different from those in the real case. For more on this see Diestel and Uhl [11, Section V.2], Benyamini and Lindenstrauss [4, Section 5.2] or Métivier [30, Section 11].

Martingale Inequalities

Martingales will allow us to prove maximal inequalities which are useful and important both in analysis and in probability theory. In order to simplify the exposition we introduce the following (quite common) shorthand notation:

$$u_N^*(x) := \max_{1\leqslant n\leqslant N} |u_n(x)| \quad \text{and} \quad u_\infty^*(x) := \lim_{N\to\infty} u_N^*(x) = \sup_{n\in\mathbb{N}} |u_n(x)|.$$

The following simple lemma is the key to all maximal inequalities.

Lemma 25.11 *Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space and $(u_n)_{n\in\mathbb{N}}$ be a submartingale. Then we have for all $s > 0$ and $N \in \mathbb{N}$*

$$\mu\Big\{\max_{1\leqslant n\leqslant N} u_n \geqslant s\Big\} \leqslant \frac{1}{s}\int_{\big\{\max\limits_{1\leqslant n\leqslant N} u_n\geqslant s\big\}} u_N\,d\mu \leqslant \frac{1}{s}\int u_N^+\,d\mu. \tag{25.9}$$

If $u_n \in \mathcal{L}^p(\mu)$, $u \geqslant 0$, or if $(u_n)_{n\in\mathbb{N}} \subset \mathcal{L}^p(\mu)$, $p \in [1,\infty)$, is a martingale, then

$$\mu\{u_N^* \geqslant s\} \leqslant \frac{1}{s^p}\int_{\{u_N^*\geqslant s\}} |u_N|^p\,d\mu \leqslant \frac{1}{s^p}\int |u_N|^p\,d\mu. \tag{25.10}$$

Proof Consider the stopping time when u_n exceeds the level s for the first time:

$$\sigma := \inf\{n \leqslant N : u_n \geqslant s\} \wedge (N+1) \qquad (\inf\emptyset = +\infty),$$

and set $A := \big\{\max_{1\leqslant n\leqslant N} u_n \geqslant s\big\} = \bigcup_{n=1}^N \{u_n \geqslant s\} = \{\sigma \leqslant N\} \in \mathscr{A}_\sigma$, where we use Lemma 23.6. From Theorem 23.7(iii) and the fact that $u_\sigma \geqslant s$ on A, we conclude

$$\mu\Big(\underbrace{\bigcup_{n=1}^N \{u_n \geqslant s\}}_{=A}\Big) \leqslant \int_A \frac{u_\sigma}{s}\,d\mu = \frac{1}{s}\int_A u_\sigma\,d\mu \leqslant \frac{1}{s}\int_A u_N\,d\mu \leqslant \frac{1}{s}\int u_N^+\,d\mu.$$

The second inequality in (25.10) follows along the same lines since, under our assumptions, $(|u_n|^p)_{n\in\mathbb{N}}$ is a submartingale, see Example 23.3(vi). □

The next theorem is commonly referred to as *Doob's maximal inequality*.

Theorem 25.12 (Doob's maximal L^p-inequality) *Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space, $1 < p < \infty$, and let $(u_n)_{n\in\mathbb{N}}$ be a martingale or $(|u_n|^p)_{n\in\mathbb{N}}$ a submartingale. Then we have*

$$\|u_N^*\|_p \leqslant \frac{p}{p-1}\|u_N\|_p \leqslant \frac{p}{p-1}\max_{1\leqslant n\leqslant N}\|u_n\|_p.$$

Proof It is suffices to consider the case where $(u_n)_{n\in\mathbb{N}}$ is a martingale; the situation where $(|u_n|^p)_{n\in\mathbb{N}}$ is a submartingale is similar and simpler.

If $\|u_N\|_p = \infty$, the inequality is trivial; if $u_N \in \mathcal{L}^p(\mu)$, then $u_1, \ldots, u_{N-1}$ are in $\mathcal{L}^p(\mu)$ since $(|u_n|^p)_{n\in\mathbb{N}}$ is a submartingale, see Example 23.3(vi). Thus

$$u_N^* \leqslant |u_1| + |u_2| + \cdots + |u_N| \implies u_N^* \in \mathcal{L}^p(\mu)$$

and using (14.9) of Corollary 14.15 and Tonelli's theorem (Theorem 14.8) we find

$$\begin{aligned}
\int (u_N^*)^p \, d\mu &\overset{14.9}{=} p \int_0^\infty s^{p-1} \mu\{u_N^* \geqslant s\} ds \\
&\overset{(25.10)}{\leqslant} p \int_0^\infty s^{p-2} \left(\int |u_N| \mathbb{1}_{\{u_N^* \geqslant s\}} \, d\mu \right) ds \\
&\overset{14.8}{=} p \int |u_N| \left(\int_0^{u_N^*} s^{p-2} \, ds \right) d\mu \\
&= \frac{p}{p-1} \int |u_N| \, (u_N^*)^{p-1} \, d\mu.
\end{aligned}$$

Hölder's inequality (Theorem 13.2) with $\frac{1}{p} + \frac{1}{q} = 1$, i.e. $q = p/(p-1)$, yields

$$\int (u_N^*)^p \, d\mu \leqslant \frac{p}{p-1} \left(\int |u_N|^p \, d\mu \right)^{1/p} \left(\int (u_N^*)^p \, d\mu \right)^{1-1/p},$$

and the claim follows. □

Using the continuity of measures (Proposition 4.3), resp. Beppo Levi's theorem (Theorem 9.6), we derive from (25.9), resp. Theorem 25.12, the following result.

Corollary 25.13 *Let $(u_n)_{n\in\mathbb{N}}$ be a martingale on the σ-finite filtered measure space $(X, \mathscr{A}, \mathscr{A}_n, \mu)$. Then*

$$\mu\{u_\infty^* \geqslant s\} \leqslant \frac{1}{s} \sup_{n\in\mathbb{N}} \|u_n\|_1; \tag{25.11}$$

$$\|u_\infty^*\|_p \leqslant \frac{p}{p-1} \sup_{n\in\mathbb{N}} \|u_n\|_p, \quad p \in (1, \infty). \tag{25.12}$$

If $(u_n)_{n\in\mathbb{N}\cup\{\infty\}}$ is a martingale, we may replace $\sup_{n\in\mathbb{N}} \|u_n\|_p$, $p \in [1, \infty)$, *in* (25.11) *and* (25.12) *by* $\|u_\infty\|_p$.

An inequality of the form (25.11) is a so-called *weak-type maximal* inequality as opposed to the strong-type (p, p) inequalities of the form (25.12).

If $p = 1$ and $(u_n)_{n\in\mathbb{N}\cup\{\infty\}}$ is a martingale, we cannot expect a $(1, 1)$ strong-type inequality like (25.12) and we have to settle for the weak-type maximal inequality (25.11) instead. Otherwise, the best we can hope for is

$$\|u_\infty^*\|_1 \leqslant \frac{e}{e-1}\left(\mu(X) + \int u_\infty (\log u_\infty)^+ \, d\mu\right)$$

if $\mu(X) < \infty$, or

$$\int_{\{u_\infty^* \geqslant \alpha\}} u_\infty^* \, d\mu \leqslant \frac{e\alpha}{e\alpha - 1}\left(\|u_\infty\|_1 + \int u_\infty (\log u_\infty)^+ \, d\mu\right)$$

otherwise.

Details can be found in Doob [13, pp. 317–8], with a few obvious modifications if $\mu(X) = \infty$.

The Hardy–Littlewood Maximal Theorem

Doob's martingale inequalities Theorem 25.12 and Corollary 25.13 can be seen as abstract versions of the classical Hardy–Littlewood estimates for maximal functions in $\mathbb{R}^n$. To prepare the ground we begin with a dyadic example.

Example 25.14 Consider in $\mathbb{R}^n$ the half-open squares

$$Q_k(z) := z + [0, 2^{-k})^n, \quad k \in \mathbb{Z},\ z \in 2^{-k}\mathbb{Z}^n,$$

with lower left corner z and side-length 2^{-k}. Then

$$\mathscr{A}_k^{[0]} := \sigma\big(Q_k(z) : z \in 2^{-k}\mathbb{Z}^n\big), \quad k \in \mathbb{Z},$$

defines a (two-sided infinite) filtration

$$\ldots \subset \mathscr{A}_{-2}^{[0]} \subset \mathscr{A}_{-1}^{[0]} \subset \mathscr{A}_0^{[0]} \subset \mathscr{A}_1^{[0]} \subset \mathscr{A}_2^{[0]} \subset \ldots$$

of sub-σ-algebras of $\mathscr{B}(\mathbb{R}^n)$. The superscript '[0]' indicates that the square lattice in each $\mathscr{A}_k^{[0]}$ contains some square with the origin $0 \in \mathbb{R}^n$ as lower left corner. Just as in Example 23.3(ix), one sees that for a function $f \in \mathcal{L}^1(\lambda^n)$

$$f_k(x) := \sum_{z \in 2^{-k}\mathbb{Z}^n} \frac{1}{\lambda^n(Q_k(z))} \int_{Q_k(z)} f \, d\lambda^n \, \mathbb{1}_{Q_k(z)}(x), \qquad k \in \mathbb{Z}, \tag{25.13}$$

is a martingale – if you are unhappy about the two-sided infinite index set, then think of $\big(f_k, \mathscr{A}_k^{[0]}\big)_{k \in \mathbb{N}_0}$ as a martingale and of $\big(f_{-k}, \mathscr{A}_{-k}^{[0]}\big)_{k \in \mathbb{N}_0}$ as a backwards martingale.

For the *square maximal function*

$$f_{[0]}^*(x) := \sup_{k \in \mathbb{Z}} |f_k(x)| = \sup_Q \frac{1}{\lambda^n(Q)} \int_Q |f| \, d\lambda^n$$

where the $\sup_Q$ ranges over all cubes $Q = Q_k(z)$ such that $k \in \mathbb{Z}$, $z \in 2^{-k}\mathbb{Z}^n$ and $x \in Q$, and the submartingale $(|f_k|)_{k\in\mathbb{Z}}$, see Example 23.3(v), Doob's inequalities become

$$\lambda^n\{f^*_{[0]} \geqslant s\} \leqslant \frac{1}{s} \sup_{k\in\mathbb{Z}} \int |f_k| d\lambda^n \leqslant \frac{1}{s} \int |f| d\lambda^n.$$

The classical Hardy–Littlewood maximal function is similar to the square maximal function from Example 25.14, the only difference being that one uses balls rather than squares.

Definition 25.15 The *Hardy–Littlewood maximal function* of $u \in \mathcal{L}^p(\lambda^n)$, $1 \leqslant p < \infty$, is defined by

$$u^*(x) := \sup_{B: B\ni x} \frac{1}{\lambda^n(B)} \int_B |u| d\lambda^n,$$

where $B \subset \mathbb{R}^n$ stands for a generic (open or closed) ball of any radius.

From the Hölder inequality we see that for all sets with finite Lebesgue measure

$$\int_A |u| d\lambda^n \leqslant (\lambda^n(A))^{1-1/p} \left(\int_A |u|^p \, d\lambda^n \right)^{1/p}, \quad 1 \leqslant p < \infty,$$

so that u^* is well-defined. However, since u^* is given by a (possibly uncountable) supremum, it is not obvious whether u^* is Borel measurable.

Lemma 25.16 *Let $u \in \mathcal{L}^p(\lambda^n)$, $1 \leqslant p < \infty$. The Hardy–Littlewood maximal function satisfies*

$$u^*(x) = \sup \left\{ \frac{1}{\lambda^n(B_r(c))} \int_{B_r(c)} |u| d\lambda^n : r \in \mathbb{Q}_+,\ c \in \mathbb{Q}^n,\ x \in B_r(c) \right\}.$$

In particular, u^* is Borel measurable.

Proof Since $\mathbb{Q}_+ \times \mathbb{Q}^n$ is countable, the formula shows that u^* arises from a countable supremum of Borel measurable functions and is, by Corollary 8.10, again Borel measurable.

The inequality '$\geqslant$' is clear since every ball with rational centre and radius is admissible in the definition of the maximal function u^*. To see '$\leqslant$', we fix $x \in \mathbb{R}^n$ and pick some generic (open or closed) ball B with $x \in B$. Given some $\epsilon > 0$, we can find $r \in \mathbb{Q}_+$ and $c \in \mathbb{Q}^n$ such that $B' := B_r(c) \subset B$, $\frac{1}{2}\lambda^n(B) \leqslant \lambda^n(B')$ and

$$\lambda^n(B \setminus B')^{1-1/p} \leqslant \frac{\lambda^n(B)}{\|u\|_p} \frac{\epsilon}{2} \leqslant \frac{\lambda^n(B')}{\|u\|_p} \epsilon.$$

Therefore,

$$\begin{aligned}
\frac{1}{\lambda^n(B)}\int_B |u|\,d\lambda^n &\leqslant \frac{1}{\lambda^n(B')}\int_B |u|\,d\lambda^n \\
&= \frac{1}{\lambda^n(B')}\int_{B\setminus B'} |u|\,d\lambda^n + \frac{1}{\lambda^n(B')}\int_{B'} |u|\,d\lambda^n \\
&\leqslant \frac{1}{\lambda^n(B')}\lambda^n(B\setminus B')^{1-1/p}\|u\|_p + \frac{1}{\lambda^n(B')}\int_{B'} |u|\,d\lambda^n \\
&\leqslant \epsilon + \sup_{B':x\in B'} \frac{1}{\lambda^n(B')}\int_{B'} |u|\,d\lambda^n
\end{aligned}$$

(the supremum ranges over all balls B' with rational radius and centre $x\in B'$), where we use Hölder's inequality in the penultimate line. Since ϵ and B are arbitrary, the inequality '$\leqslant$' follows by considering the supremum over all balls with $x\in B$ and then letting $\epsilon\to 0$. □

We will see now that u^* is in $\mathcal{L}^p(\lambda^n)$ if $1<p<\infty$.

Theorem 25.17 (Hardy, Littlewood) *Let $u\in\mathcal{L}^p(\lambda^n)$, $1\leqslant p<\infty$, and write u^* for the maximal function. Then*

$$\lambda^n\{u^*\geqslant s\}\leqslant \frac{c_n}{s}\|u\|_1, \qquad s>0,\ p=1, \tag{25.14}$$

$$\|u^*\|_p\leqslant \frac{pc_n}{p-1}\|u\|_p, \qquad 1<p<\infty, \tag{25.15}$$

with the universal constant

$$c_n=\left(\frac{16}{\sqrt{\pi}}\right)^n\Gamma\left(\frac{n}{2}+1\right).$$

Proof If we could show that the square maximal function $u^*_{[0]}$ satisfies $u^*_{[0]}\geqslant u^*$, then (25.14) and (25.15) would immediately follow from Doob's inequalities (Corollary 25.13), see Example 25.14. The problem, however, is that a ball B_r of radius $r\in\left[\frac14 2^{-k-1},\frac14 2^{-k}\right)$, $k\in\mathbb{Z}$, need not fall entirely within any single square of our lattice $\mathscr{A}_k^{[0]}$, see Fig. 25.1.

But if we move our lattice by $2\cdot\frac14 2^{-k}=\frac12 2^{-k}$ in certain (combinations of) coordinate directions, we can 'catch' B_r inside a single square Q' of the shifted lattice. [✍] More precisely, if

$$\mathbf{e}:=(\epsilon_1,\dots,\epsilon_n), \qquad \epsilon_j\in\left\{0,\tfrac12 2^{-k}\right\},$$

then

$$\mathscr{A}_k^{[\mathbf{e}]}:=\sigma\left(\mathbf{e}+Q_k(z):z\in 2^{-k}\mathbb{Z}^n\right),\ k\in\mathbb{Z}, \quad \left(u_k^{[\mathbf{e}]}\right)_{k\in\mathbb{Z}}, \quad u^*_{[\mathbf{e}]},$$

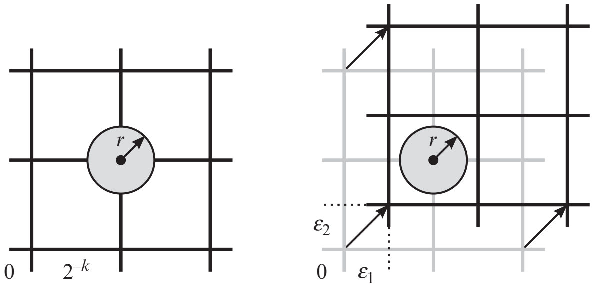


Fig. 25.1. A ball B_r of radius $\frac{1}{4}2^{-k-1} \leqslant r < \frac{1}{4}2^{-k}$, $k \in \mathbb{Z}$, need not fall entirely within any single square of side-length 2^{-k}, but a suitably shifted lattice will do.

are 2^n filtrations with corresponding martingales and square maximal functions. As in Example 25.14 we find that

$$\lambda^n\{u^*_{[\mathbf{e}]} \geqslant s\} \leqslant \frac{1}{s}\|u\|_1, \quad s > 0. \tag{25.16}$$

Combining Corollary 16.21 with the translation invariance and scaling behaviour of Lebesgue measure we see that the volume of a ball B_r of radius $\frac{1}{4}2^{-k-1} \leqslant r < \frac{1}{4}2^{-k}$ and arbitrary centre is

$$\lambda^n(B_r) = r^n\lambda^n(B_1) \overset{16.21}{=} \frac{\pi^{n/2}r^n}{\Gamma\left(\frac{n}{2}+1\right)} \geqslant \frac{\pi^{n/2}\left(\frac{1}{4}2^{-k-1}\right)^n}{\Gamma\left(\frac{n}{2}+1\right)},$$

hence we get from $x \in B_r \subset Q'$ and $\lambda^n(Q') = (2^{-k})^n$ that

$$\begin{aligned}
\frac{1}{\lambda^n(B_r)}\int_{B_r}|u|\,d\lambda^n &\leqslant \frac{\lambda^n(Q')}{\lambda^n(B_r)}\frac{1}{\lambda^n(Q')}\int_{Q'}|u|\,d\lambda^n \\
&\leqslant \frac{\left(2^{-k}\right)^n\Gamma\left(\frac{n}{2}+1\right)}{\pi^{n/2}\left(\frac{1}{8}2^{-k}\right)^n}\frac{1}{\lambda^n(Q')}\int_{Q'}|u|\,d\lambda^n \\
&\leqslant \underbrace{\left(\frac{8}{\sqrt{\pi}}\right)^n\Gamma\left(\frac{n}{2}+1\right)}_{=:\gamma_n}\max_{\mathbf{e}} u^*_{[\mathbf{e}]}(x).
\end{aligned}$$

This shows that $u^* \leqslant \gamma_n \max_{\mathbf{e}} u^*_{[\mathbf{e}]}$ and

$$\begin{aligned}
\lambda^n\{u^* \geqslant s\} \leqslant \lambda^n\left\{\max_{\mathbf{e}} u^*_{[\mathbf{e}]} \geqslant \frac{s}{\gamma_n}\right\} &\leqslant \sum_{\mathbf{e}}\lambda^n\left\{u^*_{[\mathbf{e}]} \geqslant \frac{s}{\gamma_n}\right\} \\
&\overset{(25.16)}{\leqslant} 2^n\gamma_n\frac{1}{s}\int|u|\,\lambda^n(ds).
\end{aligned}$$

A very similar argument yields $\|u^*_{[\mathbf{e}]}\|_p \leqslant (p/(p-1))\|u\|_p$ for all shifts $\mathbf{e}$, and Doob's inequality (25.12) applied to each $u^*_{[\mathbf{e}]}$ finally shows that

$$\|u^*\|_p \leqslant \gamma_n \big\| \max_{\mathbf{e}} u^*_{[\mathbf{e}]} \big\|_p \leqslant \gamma_n \sum_{\mathbf{e}} \|u^*_{[\mathbf{e}]}\|_p \leqslant 2^n \gamma_n \frac{p}{p-1} \|u\|_p.$$

All that remains to be done is to call $c_n := 2^n \gamma_n$. □

The proof of Theorem 25.17 extends with very little effort to maximal functions of finite measures.

Definition 25.18 Let μ be a locally finite[2] measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. The *maximal function* is given by

$$\mu^*(x) := \sup_{B: B \ni x} \frac{\mu(B)}{\lambda^n(B)},$$

where $B \subset \mathbb{R}^n$ stands for a generic open ball of any radius.

If we replace in the proof of Theorem 25.17 the expression $\int_B |u| d\lambda^n$ by $\mu(B)$ and $u^*_{[\mathbf{e}]}(x)$ by

$$\mu^*_{[\mathbf{e}]}(x) := \sup \left\{ \frac{\mu(Q)}{\lambda^n(Q)} : Q = Q_k(z),\ k \in \mathbb{Z},\ z \in 2^{-k}\mathbb{Z}^n,\ x \in Q \right\},$$

we arrive at the following generalization of (25.14)

Corollary 25.19 *Let μ be a finite measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and denote by μ^* the maximal function. For $c_n = (16/\sqrt{\pi})^n \Gamma\left(\frac{n}{2} + 1\right)$ one has*

$$\lambda^n\{\mu^* \geqslant s\} \leqslant \frac{c_n}{s} \mu(\mathbb{R}^n), \quad s > 0. \tag{25.17}$$

Lebesgue's Differentiation Theorem

Let us return once again to the Radon–Nikodým theorem. There we have seen that $\nu \ll \mu$ implies $\nu = f\mu$. The martingale proof, however, shows even more, namely

$$\nu|_{\mathscr{A}_\alpha} = f_\alpha \mu|_{\mathscr{A}_\alpha} \quad \text{and} \quad \mathcal{L}^1\text{-}\lim_{\alpha \in I} f_\alpha = f$$

(notation as in Theorem 25.2). Consider $(X, \mathscr{A}, \mu) = (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \lambda^n)$. In this case we can reduce our consideration to a *countable sequence* of σ-algebras (instead of $(\mathscr{A}_\alpha)_{\alpha \in I}$) – see Problem 25.1 – and use Theorem 24.6 instead of

[2] That is, every point $x \in \mathbb{R}^n$ has a neighbourhood $U = U(x)$ such that $\mu(U) < \infty$. In $\mathbb{R}^n$ this is clearly equivalent to saying that $\mu(B) < \infty$ for every open ball B.

Theorem 25.5. In fact, this would even allow us to get $f(x)$ as a *pointwise* limit. This is one way to prove *Lebesgue's differentiation theorem*.

Theorem 25.20 (Lebesgue) *Let $u \in \mathcal{L}^1(\lambda^n)$. The limit*

$$\lim_{r\to 0} \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} |u(y) - u(x)| \lambda^n(dy) = 0 \tag{25.18}$$

exists for (*Lebesgue*) *almost all $x \in \mathbb{R}^n$. In particular,*

$$u(x) = \lim_{r\to 0} \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} u(y) \lambda^n(dy) \quad a.e. \tag{25.19}$$

We will not follow the route laid out above, but use instead the Hardy–Littlewood maximal theorem (Theorem 25.17) to prove Theorem 25.20. The reason is mainly a didactic one, since this is a beautiful example of how weak-type maximal inequalities (i.e. inequalities like (25.14) and (25.11)) can be used to get a.e. convergence. More on this theme can be found in Krantz [26, pp. 27–30] and Garsia [18, pp. 1–4]. Our proof will also show that the limits in (25.18) and (25.19) can be strengthened to $B \downarrow \{x\}$, where B is any ball containing x and, in the limit, shrinking to $\{x\}$.

Proof of Theorem 25.20 We know from Theorem 17.8 that continuous functions with compact support $C_c(\mathbb{R}^n)$ are dense in $\mathcal{L}^1(\lambda^n)$. Since $\phi \in C_c(\mathbb{R}^n)$ is uniformly continuous, we find for every $\epsilon > 0$ some $\delta > 0$ such that

$$|\phi(x) - \phi(y)| < \epsilon \quad \forall |x-y| \leqslant r,\ r < \delta.$$

Thus,

$$\lim_{r\to 0} \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} |\phi(y) - \phi(x)| \lambda^n(dy) \leqslant \epsilon \quad \forall \phi \in C_c(\mathbb{R}^n) \tag{25.20}$$

and (25.18) is true for $C_c(\mathbb{R}^n)$. For a general $u \in \mathcal{L}^1(\lambda^n)$ we pick a sequence $(\phi_n)_{n\in\mathbb{N}} \subset C_c(\mathbb{R}^n)$ with $\lim_{n\to\infty} \|u - \phi_n\|_1 = 0$. Denote by

$$w_\delta^\natural(x) := \sup_{0<r<\delta} \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} |w| d\lambda^n$$

the restricted maximal function. Since $(u-\phi_n)^\natural_\delta \leqslant (u-\phi_n)^*$ and the supremum is subadditive, we get

$$\begin{aligned}
\lambda^n\Big\{x : \limsup_{r\to 0} \frac{1}{\lambda^n(B_r(x))} \int_{B_r(x)} |u(y)-u(x)|\lambda^n(dy) > 3\epsilon\Big\} \\
&= \lambda^n\Big\{x : \inf_{\delta>0}(u-u(x))^\natural_\delta(x) > 3\epsilon\Big\} \\
&\leqslant \lambda^n\big\{x : (u-u(x))^\natural_\delta(x) > 3\epsilon\big\} \\
&= \lambda^n\big\{x : \big[(u-\phi_n)+(\phi_n-\phi_n(x))+(\phi_n(x)-u(x))\big]^\natural_\delta(x) > 3\epsilon\big\} \\
&\leqslant \lambda^n\big\{(u-\phi_n)^* > \epsilon\big\} + \lambda^n\big\{x : (\phi_n-\phi_n(x))^\natural_\delta(x) > \epsilon\big\} \\
&\quad + \lambda^n\big\{x : |\phi_n(x)-u(x)| > \epsilon\big\} \\
&\leqslant \frac{c_n}{\epsilon}\|u-\phi_n\|_1 + 0 + \frac{1}{\epsilon}\|\phi_n-u\|_1,
\end{aligned}$$

where we use Theorem 25.17, resp., (25.20) with $\delta\to 0$, resp., the Markov inequality 11.5 to deal with each of the above three terms, respectively. The assertion now follows by letting first $n\to\infty$ and then $\epsilon\to 0$. $\square$

Let us now investigate the connection between the Radon–Nikodým derivative and ordinary derivatives. For this the following auxiliary notation will be useful. If μ is a measure on $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ that assigns finite volume to any ball, we set

$$\bar{D}\mu(x) := \limsup_{r\to 0}\frac{\mu(B_r(x))}{\lambda^n(B_r(x))} = \lim_{k\to\infty}\sup_{0<r<1/k}\frac{\mu(B_r(x))}{\lambda^n(B_r(x))}$$

and, whenever the limit exists and is finite,

$$D\mu(x) := \lim_{r\to 0}\frac{\mu(B_r(x))}{\lambda^n(B_r(x))}.$$

Note that, by Lemma 25.16, both $\bar{D}\mu$ and $D\mu$ are Borel measurable functions.

Corollary 25.21 *Let μ be a locally finite*[3] *measure on $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ which is absolutely continuous w.r.t. Lebesgue measure: $\mu\ll\lambda^n$. Then $D\mu$ exists Lebesgue a.e. and coincides a.e. with the Radon–Nikodým derivative $d\mu/d\lambda^n$, that is, $\mu = D\mu\lambda^n$.*

Proof Assume first that μ is a finite measure. By the Radon–Nikodým theorem we know that there is a unique function $f\in\mathcal{L}^1(\lambda^n)$ with

$$\frac{\mu(B_r(x))}{\lambda^n(B_r(x))} = \frac{1}{\lambda^n(B_r(x))}\int_{B_r(x)} f(y)\lambda^n(dy).$$

[3] See the footnote on page 309. Note that μ is automatically σ-finite. [✍]

By Lebesgue's differentiation theorem the right-hand side of the above equality tends for λ^n-almost all x to $f(x)$ as $r \to 0$, so that $D\mu = f$ almost everywhere.

If μ is not finite, we choose an exhausting sequence of open balls $B_k(0)$, $k \in \mathbb{N}$, and set $\mu_k(\cdot) := \mu(B_k(0) \cap \cdot\ )$ and $\lambda_k^n(\cdot) := \lambda^n(B_k(0) \cap \cdot\ )$. Since the measures μ_k and λ_k^n are finite, the previous argument applies and shows that $D\mu_k = d\mu_k / d\lambda_k^n$ a.e. for every $k \in \mathbb{N}$. By the very definition of the Radon–Nikodým derivative, we find that $d\mu_k / d\lambda_k^n = d\mu_i / d\lambda_i^n$ on $B_i(0)$ whenever $i < k$, and the same is true for $D\mu_k$, resp. $D\mu_i$. Thus

$$D\mu(x) := D\mu_k(x) \quad \text{and} \quad \frac{d\mu}{d\lambda^n}(x) := \frac{d\mu_k}{d\lambda_k^n}(x) \quad \forall x \in B_k(0)$$

are well-defined functions which satisfy $D\mu = d\mu / d\lambda^n$ λ^n-almost everywhere. □

Corollary 25.22 *Let ν be a locally finite*[4] *measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ which is singular w.r.t. λ^n, i.e. $\nu \perp \lambda^n$. Then $D\nu = 0$ λ^n-almost everywhere.*

Proof Assume first that ν is a finite measure. Write $\|\nu\|$ for the total mass of ν. It suffices to show that $\bar{D}\nu = 0$ a.e. Since $\nu \perp \lambda^n$, there is some λ^n-null set N with $\nu(N) = \|\nu\|$. From Theorem H.2 we know that ν is inner regular. Thus, for every $\epsilon > 0$, there is some compact set $K = K_\epsilon \subset N$ such that $\nu(K) > \|\nu\| - \epsilon$. Setting

$$\nu_1(\cdot) := \nu(K \cap \cdot\ ) \quad \text{and} \quad \nu_2 := \nu - \nu_1$$

we obtain two measures ν_1, ν_2 with $\nu = \nu_1 + \nu_2$ and $\|\nu_2\| \leqslant \epsilon$. Since K^c is open, we conclude from the definition of the derivative that $\bar{D}\nu_1(x) = 0$ for all $x \in K^c$, so that

$$\bar{D}\nu(x) = \bar{D}\nu_1(x) + \bar{D}\nu_2(x) = \bar{D}\nu_2(x) \leqslant \nu_2^*(x) \qquad \forall x \in K^c,$$

where ν_2^* denotes the maximal function for the measure ν_2. This shows that

$$\{\bar{D}\nu > s\} \subset K \cup \{\nu_2^* > s\} \qquad \forall s > 0.$$

Using that $\lambda^n(K) \leqslant \lambda^n(N) = 0$ and the maximal inequality (Corollary 25.19) we get

$$\lambda^n\{\bar{D}\nu > s\} \leqslant \lambda^n\{\nu_2^* > s\} \leqslant \frac{c_n}{s}\|\nu_2\| \leqslant \frac{c_n}{s}\epsilon.$$

Since $\epsilon > 0$ and $s > 0$ are arbitrary, we conclude that $\bar{D}\nu = 0$ Lebesgue a.e.

If ν is not finite, we choose a sequence of open balls $B_k(0)$, $k \in \mathbb{N}$, and set $\nu_k(\cdot) := \nu(B_k(0) \cap \cdot\ )$. Obviously, $\bar{D}\nu = \bar{D}\nu_k$ on $B_k(0)$, and therefore the first

[4] See the footnote on page 309.

part of the proof shows that $\bar{D}\nu(x)=0$ for Lebesgue almost all $x\in B_k(0)$. Denoting the exceptional set by M_k we see that $\bar{D}\nu(x)=0$ for all $x\notin M:=\bigcup_{k\in\mathbb{N}} M_k$; the latter, however, is an λ^n-null set, and the theorem follows. □

The Calderón–Zygmund Lemma

Our last topic is the famous *Calderón–Zygmund decomposition*, which is at the heart of many further developments in the theory of singular integral operators. We take the proof from Stein's book [50, p. 17] and rephrase it a little to bring out the martingale connection.

Lemma 25.23 (Calderón–Zygmund decomposition) *Let $u\in\mathcal{L}^1(\lambda^n)$, $u\geqslant 0$ and $\alpha>0$. There exists a decomposition of $\mathbb{R}^n$ such that*

(i) $\mathbb{R}^n=F\cup\Omega$ *and* $F\cap\Omega=\emptyset$;

(ii) $u\leqslant\alpha$ *almost everywhere on F;*

(iii) $\Omega=\dot{\bigcup}_{k\in\mathbb{N}}Q_k$ *with mutually disjoint half-open axis-parallel squares Q_k such that for each Q_k*

$$\alpha<\frac{1}{\lambda^n(Q_k)}\int_{Q_k}u\,d\lambda^n\leqslant 2^n\alpha.$$

Proof Let $\mathscr{A}_k:=\mathscr{A}_k^{[0]}$, $k\in\mathbb{Z}$, be the dyadic filtration of Example 25.14 and let $(u_k)_{k\in\mathbb{Z}}$ be the corresponding martingale (25.13). Introduce a stopping time

$$\tau:=\inf\{k\in\mathbb{Z}:u_k>\alpha\},\quad \inf\emptyset:=+\infty,$$

and set $F:=\{\tau=+\infty\}\cup\{\tau=-\infty\}$ and $\Omega:=\{-\infty<\tau<+\infty\}$. By the very definition of the martingale $(u_k)_{k\in\mathbb{N}}$ we see

$$u_k(x)\leqslant\frac{1}{\lambda^n(Q_k)}\int u\,d\lambda^n=2^{nk}\|u\|_1\xrightarrow[k\to-\infty]{}0,$$

so that $\lim_{k\to-\infty}u_k(x)=0$ and $\{\tau=-\infty\}=\emptyset$. If $x\in\{\tau=+\infty\}$, we have $u_k(x)\leqslant\alpha$ and so $u(x)=\lim_{k\to\infty}u_k(x)\leqslant\alpha$ a.e., as the almost everywhere pointwise limit exists by Corollary 24.3 (note that $\int u_k\,d\lambda^n=\int u\,d\lambda^n<\infty$). This settles (i) and (ii).

Since τ is a stopping time, $\{\tau=k\}=\{\tau\leqslant k\}\setminus\{\tau\leqslant k-1\}\in\mathscr{A}_k$, hence $\{\tau=k\}$ as well as $\Omega=\dot{\bigcup}_{k\in\mathbb{Z}}\{\tau=k\}$ are unions of disjoint half-open squares. The estimate in (iii) can be written as

$$\alpha<u_\tau(x)\leqslant 2^n\alpha\quad\forall x\in\Omega.$$

From its definition, $u_\tau > \alpha$ is clear. For the upper estimate we note that every square $Q_{k-1} \in \mathscr{A}_{k-1}$ is made up of 2^n squares $Q_k \in \mathscr{A}_k$, so that

$$\frac{u_k}{u_{k-1}} = \frac{\displaystyle\sum_{z\in 2^{-k+1}\mathbb{Z}^n} \left(\sum_{Q_k(y)\subset Q_{k-1}(z)} \frac{1}{\lambda^n(Q_k(y))} \int_{Q_k(y)} u\, d\lambda^n \, \mathbb{1}_{Q_k(y)} \right)}{\displaystyle\sum_{z\in 2^{-k+1}\mathbb{Z}^n} \frac{1}{\lambda^n(Q_{k-1}(z))} \int_{Q_{k-1}(z)} u\, d\lambda^n \, \mathbb{1}_{Q_{k-1}(z)}}$$

$$\leqslant 2^n \frac{\displaystyle\sum_{z\in 2^{-k+1}\mathbb{Z}^n} \left(\sum_{Q_k(y)\subset Q_{k-1}(z)} \frac{2^{-n}}{\lambda^n(Q_k(y))} \int_{Q_k(y)} u\, d\lambda^n \right) \mathbb{1}_{Q_{k-1}(z)}}{\displaystyle\sum_{z\in 2^{-k+1}\mathbb{Z}^n} \frac{1}{\lambda^n(Q_{k-1}(z))} \int_{Q_{k-1}(z)} u\, d\lambda^n \, \mathbb{1}_{Q_{k-1}(z)}} = 2^n.$$

Finally, by the definition of τ,

$$\begin{aligned} \mathbb{1}_{\{-\infty<\tau<+\infty\}} u_\tau = \sum_{k\in\mathbb{Z}} u_k \mathbb{1}_{\{\tau=k\}} &\leqslant \sum_{k\in\mathbb{Z}} 2^n u_{k-1} \mathbb{1}_{\{\tau=k\}} \\ &\leqslant \sum_{k\in\mathbb{Z}} 2^n \alpha \mathbb{1}_{\{\tau=k\}} \\ &= 2^n \alpha \mathbb{1}_{\{-\infty<\tau<+\infty\}}, \end{aligned}$$

and the proof is complete. □

Note that all results that we have proved here for λ^n and $\mathbb{R}^n$ can be extended to *spaces of homogeneous type*, i.e. metric spaces X with a measure μ that is finite and strictly positive on balls and has the following *volume doubling* property: for some positive constant $\kappa > 0$ we have

$$\mu(B_{2r}(x)) \leqslant \kappa \mu(B_r(x)) \quad \forall x \in X,\ r > 0,$$

see Krantz [26, Section 6.1, pp. 235–61].

Problems

25.1. Show that Theorem 24.6 is enough to prove the Radon–Nikodým theorem (Theorem 25.2) for a countably generated $\mathscr{A}$, i.e. $\mathscr{A} = \sigma(\{A_n\}_{n\in\mathbb{N}})$.
[Hint: set $\mathscr{A}_n := \sigma(A_1, A_2, \dots, A_n)$ and observe that the atoms of $\mathscr{A}_n$ are of the form $C_1 \cap \cdots \cap C_n$, where $C_i \in \{A_i, A_i^c\}$, $1 \leqslant i \leqslant n$.]

25.2. Let μ and ν be measures on the measurable space $(X, \mathscr{A})$. Show that the absolute continuity condition (25.1) is equivalent to

$$\mu(A \triangle B) = 0 \implies \nu(A) = \nu(B) \qquad \text{for all } A, B \in \mathscr{A}.$$

25.3. A theorem of Doob. *Let $(\mu_t)_{t\geqslant 0}$ and $(\nu_t)_{t\geqslant 0}$ be two families of measures on the σ-finite measure space $(X,\mathscr{A})$ such that $\nu_t \ll \mu_t$ for all $t\geqslant 0$ and $t\mapsto \mu_t(A), \nu_t(A)$ are measurable for all $A\in\mathscr{A}$. Then there exists a measurable function $(t,x)\mapsto p(t,x)$, $(t,x)\in[0,\infty)\times X$, such that $\nu_t = p(t,\cdot)\mu_t$ for all $t\geqslant 0$.*

[Hint: argue as in the proof of Theorem 25.2:

$$p_\alpha(t,x) := \sum_{A\in\alpha} \frac{\nu_t(A)}{\mu_t(A)} \mathbb{1}_A(x),$$

and check that this function is jointly measurable in t and x.]

25.4. A measure ν on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ is called *quasi-invariant* if for all $N\in\mathscr{B}(\mathbb{R}^n)$ with $\nu(N)=0$ it holds that $\nu(N+x)=0$ for all $x\in\mathbb{R}^n$. Show that $\nu\ll\lambda$ and $\lambda\ll\nu$, where λ is n-dimensional Lebesgue measure.

[Hint: consider the convolution $\nu\star\lambda(N)$.]

25.5. Do any of the problems at the end of Chapter 20.

25.6. Conditional expectations. Let $(X,\mathscr{A},\mu)$ be a σ-finite measure space and $\mathscr{F}\subset\mathscr{A}$ be a sub-σ-algebra. Then use the Radon–Nikodým theorem to show that for every $u\in\mathcal{L}^1(\mathscr{A})$ there exists a – unique up to null sets – $\mathscr{F}$-measurable function $u^{\mathscr{F}}\in\mathcal{L}^1(\mathscr{F})$ such that

$$\int_F u^{\mathscr{F}}\,d\mu = \int_F u\,d\mu \quad \forall F\in\mathscr{F}. \tag{25.21}$$

Use this result to rephrase the (sub, super)martingale property 23.1.

Remark. Since $u^{\mathscr{F}}$ is unique (modulo null sets) one often writes $u^{\mathscr{F}} = E^{\mathscr{F}}u$, where $E^{\mathscr{F}}$ is an operator which is called *conditional expectation*. We will introduce this operator in a different way in Chapter 27 and show in Theorem 27.12 that we could have defined $E^{\mathscr{F}}$ by (25.21).

25.7. Let μ,ν be σ-finite measures on the measurable space $(X,\mathscr{A})$. Let $(\mathscr{A}_n)_{n\in\mathbb{N}}$ be a filtration of sub-σ-algebras of $\mathscr{A}$ such that $\mathscr{A}=\sigma\left(\bigcup_{n\in\mathbb{N}}\mathscr{A}_n\right)$ and then denote $\mu_n := \mu|_{\mathscr{A}_n}$ and $\nu_n := \nu|_{\mathscr{A}_n}$. If $\mu_n\ll\nu_n$ for all $n\in\mathbb{N}$, then $\mu\ll\nu$. Find an expression for the density $d\mu/d\nu$.

25.8. Kolmogorov's inequality. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables on a probability space $(\Omega,\mathscr{A},\mathbb{P})$. Then we have the following generalization of Chebyshev's inequality, see Problem 11.3(vi),

$$\mathbb{P}\left(\max_{1\leqslant k\leqslant n}\left|\sum_{k=1}^n(\xi_k-\mathbb{E}\xi_k)\right|\geqslant t\right) \leqslant \frac{1}{t^2}\sum_{k=1}^n \mathbb{V}\xi_k,$$

where $\mathbb{E}\xi=\int\xi\,d\mathbb{P}$ is the expectation or mean value and $\mathbb{V}\xi=\int(\xi-\mathbb{E}\xi)^2\,d\mathbb{P}$ the variance of the random variable (i.e. measurable function) $\xi:\Omega\to\mathbb{R}$.

25.9. Let $u,w\geqslant 0$ be measurable functions on a σ-finite measure space $(X,\mathscr{A},\mu)$.

(i) Show that $t\mu\{u\geqslant t\}\leqslant\int_{\{u\geqslant t\}} w\,d\mu$ for all $t>0$ implies that

$$\int u^p\,d\mu \leqslant \frac{p}{p-1}\int u^{p-1}w\,d\mu \quad \forall p>1.$$

(ii) Assume that $u,w\in L^p$. Conclude from (i) that $\|u\|_p\leqslant (p/(p-1))\|w\|_p$ for $p>1$.

[Hint: use the technique of the proof of Theorem 25.12; for (ii) use Hölder's inequality.]

25.10. Show the following improvement of Doob's maximal inequality Theorem 25.12. Let $(u_n)_{n\in\mathbb{N}}$ be a martingale or $(|u_n|^p)_{n\in\mathbb{N}}$, $1<p<\infty$, be a submartingale on a σ-finite filtered measure space. Then

$$\max_{n\leqslant N}\|u_n\|_p \leqslant \|u_N^*\|_p \leqslant \frac{p}{p-1}\|u_N\|_p \leqslant \frac{p}{p-1}\max_{1\leqslant n\leqslant N}\|u_n\|_p.$$

25.11. **$\mathcal{L}^p$-bounded martingales.** A martingale $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ on a σ-finite filtered measure space is called *$\mathcal{L}^p$-bounded*, if $\sup_{n\in\mathbb{N}} \int |u_n|^p \, d\mu < \infty$ for some $p > 1$. Show that the sequence $(u_n)_{n\in\mathbb{N}}$ converges a.e. and in $\mathcal{L}^p$-sense to a function $u \in \mathcal{L}^p(\mu)$.
[Hint: compare this with Problem 24.8.]

25.12. Use Theorem 24.6 to show that the martingale of Example 25.14 is uniformly integrable.

25.13. Let $u: [a, b] \to \mathbb{R}$ be a continuous function. Show that $x \mapsto \int_{[a,x]} u(t)dt$ is everywhere differentiable and find its derivative. What happens if we assume only that $u \in \mathcal{L}^1(dt)$?
[Hint: consider Theorem 25.20.]

25.14. Let $f: \mathbb{R} \to \mathbb{R}$ be a bounded increasing function. Show that f' exists Lebesgue a.e. and $f(b) - -f(a) \geqslant \int_{(a,b)} f'(x)dx$. When do we have equality?
[Hint: assume first that f is left- or right-continuous. Then you can interpret f as the distribution function of a Stieltjes measure μ. Use Lebesgue's decomposition theorem to write $\mu = \mu^\circ + \mu^\perp$ and use Corollaries 25.21 and 25.22 to find f'. If f is not one-sided continuous in the first place, use Lemma 14.14 to find a version ϕ of f which is left- or right-continuous such that $\{\phi \neq f\}$ is at most countable, and hence a Lebesgue null set.]

25.15. **Fubini's 'other' theorem.** *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of monotone increasing functions f_n : $[a, b] \to \mathbb{R}$. If the series $s(x) := \sum_{n=1}^\infty f_n(x)$ converges, then $s'(x)$ exists a.e. and is given by $s'(x) = \sum_{n=1}^\infty f_n'(x)$ a.e.*
[Hint: the partial sums $s_n(x)$ and $s(x)$ are again increasing functions and, by Problem 25.14, $s'(x)$ and $s_n'(x)$ exist a.e.; the latter can be calculated through term-by-term differentiation. Since the f_i are increasing functions, the limits of the difference quotients show that $0 \leqslant s_n' \leqslant s_{n+1}' \leqslant s'$ a.e., hence $\sum_i f_i'$ converges a.e. To identify this series with s', show that $\sum_k (s(x) - s_{n_k}(x))$ converges on $[a, b]$ for some suitable subsequence. The first part of the proof applied to this series implies that $\sum_k (s'(x) - s_{n_k}'(x))$ converges, thus $s' - s_{n_k}' \to 0$.]

26

Abstract Hilbert Spaces

Up to now we have considered only functions with values in $\mathbb{R}$ or $\overline{\mathbb{R}}$. Often it is necessary to admit complex-valued functions, too. In what follows $\mathbb{K}$ will stand for $\mathbb{R}$ or $\mathbb{C}$.

Recall that a $\mathbb{K}$-vector space is a set V with a vector addition '$+$' : $V \times V \to V$, $(v, w) \mapsto v + w$ and a multiplication of a vector with a scalar '$\cdot$' : $\mathbb{K} \times V \to V$, $(\alpha, v) \mapsto \alpha \cdot v$, which are defined in such a way that $(V, +)$ is an Abelian group and that for all $\alpha, \beta \in \mathbb{K}$ and $v, w \in V$ the relations

$$\begin{aligned} (\alpha + \beta)v &= \alpha v + \beta v, & \alpha(v + w) &= \alpha v + \alpha w, \\ (\alpha\beta)v &= \alpha(\beta v), & 1 \cdot v &= v \end{aligned}$$

hold. Typical examples of $\mathbb{R}$-vector spaces are the spaces $\mathcal{L}^p$ or L^p and, in particular, the sequence spaces ℓ^p from Example 13.11. For the $\mathbb{C}$-versions we first need to know how to integrate complex functions.

Remark 26.1 We want to extend the integral by linearity to $\mathbb{C}$-valued integrands $f: X \to \mathbb{C}$. If we write $f = u + iv$, where $u = \operatorname{Re} f$ and $v = \operatorname{Im} f$, then

$$\int f \, d\mu := \int u \, d\mu + i \int v \, d\mu = \int \operatorname{Re} f \, d\mu + i \int \operatorname{Im} f \, d\mu$$

defines a $\mathbb{C}$-linear integral [✍] which enjoys essentially the same properties as the integral for real-valued integrands. In addition, we have

$$\operatorname{Re} \int f \, d\mu = \int \operatorname{Re} f \, d\mu, \quad \operatorname{Im} \int f \, d\mu = \int \operatorname{Im} f \, d\mu, \quad \overline{\int f \, d\mu} = \int \bar{f} \, d\mu,$$

$$f \in \mathcal{L}^p_{\mathbb{C}}(\mu) \iff f \in \mathcal{M}(\mathscr{B}(\mathbb{C})) \quad \text{and} \quad |f| \in \mathcal{L}^p_{\mathbb{R}}(\mu).$$

As for real-valued functions we define the spaces $L^p_{\mathbb{C}}(\mu)$ as equivalence classes of $\mathcal{L}^p(\mu)$-functions which are μ-a.e. equal.

Most of the results for $\mathbb{R}$-valued integrands carry over to $\mathbb{C}$-valued functions by considering real and imaginary parts separately.

The key issue is measurability: since $\mathbb{C}$ is a normed space, there is a natural Borel σ-algebra and we need to know whether $\mathscr{A}/\mathscr{B}(\mathbb{C})$-measurability of f entails $\mathscr{A}/\mathscr{B}(\mathbb{R})$-measurability of $\operatorname{Re} f$ and $\operatorname{Im} f$. Since the maps from $\mathbb{R}^2 \to \mathbb{C}$ and $\mathbb{C} \to \mathbb{R}^2$ defined by

$$(x, y) \mapsto z = x + iy \quad \text{and} \quad z \mapsto (\operatorname{Re} z, \operatorname{Im} z) := \left(\frac{1}{2}(z + \bar{z}),\ \frac{1}{2i}(z - \bar{z})\right)$$

are continuous and inverses of each other, we can identify $\mathscr{B}(\mathbb{C})$ and $\mathscr{B}(\mathbb{R}^2)$; in particular f is measurable if, and only if, $\operatorname{Re} f$ and $\operatorname{Im} f$ are measurable. Proofs and details can be found in Appendix D.

As we have seen in Chapter 13, see Remark 13.5, the spaces $V = \mathcal{L}^p_{\mathbb{K}}(\mu)$, resp. $L^p_{\mathbb{K}}(\mu)$, are semi-normed, resp. normed, vector spaces. Often, V has a richer geometric structure which is due to a scalar product that defines the norm on V.

Definition 26.2 A $\mathbb{K}$-vector space V is an *inner product space* if it supports a *scalar* or *inner product*, i.e. a map $\langle \cdot\,, \cdot \rangle : V \times V \to \mathbb{K}$ with the following properties: for all $u, v, w \in V$ and $\alpha, \beta \in \mathbb{K}$

$$\langle v, v \rangle > 0 \iff v \neq 0, \tag{SP$_1$}$$

$$\langle v, w \rangle = \overline{\langle w, v \rangle}, \tag{SP$_2$}$$

$$\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle. \tag{SP$_3$}$$

If $\mathbb{K} = \mathbb{R}$, (SP$_2$) becomes *symmetry* and (SP$_2$), (SP$_3$) show that both $v \mapsto \langle v, w \rangle$ and $w \mapsto \langle v, w \rangle$ are $\mathbb{R}$-linear; therefore we call $(v, w) \mapsto \langle v, w \rangle$ *bilinear*. If $\mathbb{K} = \mathbb{C}$, (SP$_2$), (SP$_3$) give

$$\begin{aligned} \langle u, \alpha v + \beta w \rangle &\overset{(\mathrm{SP}_2)}{=} \overline{\langle \alpha v + \beta w, u \rangle} \overset{(\mathrm{SP}_3)}{=} \overline{\alpha \langle v, u \rangle + \beta \langle w, u \rangle} \\ &= \bar{\alpha} \overline{\langle v, u \rangle} + \bar{\beta} \overline{\langle w, u \rangle} \overset{(\mathrm{SP}_2)}{=} \bar{\alpha} \langle u, v \rangle + \bar{\beta} \langle u, w \rangle, \end{aligned}$$

i.e. $w \mapsto \langle v, w \rangle$ is *skew-linear*. We call $\langle \cdot\,, \cdot \rangle$ in this case a *sesquilinear form*. Since $\mathbb{K} = \mathbb{C}$ always includes $\mathbb{K} = \mathbb{R}$, we will restrict ourselves to $\mathbb{K} = \mathbb{C}$.

Lemma 26.3 (Cauchy–Schwarz inequality) *Let V be an inner product space. Let*

$$|\langle v, w \rangle|^2 \leqslant \langle v, v \rangle \langle w, w \rangle \quad \forall v, w \in V. \tag{26.1}$$

Equality holds if, and only if, $v = \alpha w$ for some $\alpha \in \mathbb{C}$.

Proof If $v=0$ or $w=0$, there is nothing to show. For all other $v, w \in V$ and $\alpha \in \mathbb{C}$ we have

$$\begin{aligned} 0 \leqslant \langle v-\alpha w, v-\alpha w\rangle &= \langle v, v\rangle - \alpha\langle w, v\rangle - \bar{\alpha}\langle v, w\rangle + \alpha\bar{\alpha}\langle w, w\rangle \\ &= \langle v, v\rangle - 2\,\mathrm{Re}\big(\alpha\langle w, v\rangle\big) + |\alpha|^2\langle w, w\rangle, \end{aligned}$$

where we use that $z+\bar{z}=2\,\mathrm{Re}\,z$. Setting $\alpha=\langle v, v\rangle/\langle w, v\rangle$, we get

$$0 \leqslant \langle v, v\rangle - 2\,\mathrm{Re}\langle v, v\rangle + \frac{\langle v, v\rangle^2\langle w, w\rangle}{|\langle w, v\rangle|^2},$$

which implies (26.1).

Since $\langle v-\alpha w, v-\alpha w\rangle = 0$ only if $v=\alpha w$, this is necessary for equality in (26.1), too. If, indeed, $v=\alpha w$, we see

$$|\langle v, w\rangle|^2 = |\langle \alpha w, w\rangle|^2 = \alpha\bar{\alpha}\langle w, w\rangle\langle w, w\rangle = \langle \alpha w, \alpha w\rangle\langle w, w\rangle = \langle v, v\rangle\langle w, w\rangle,$$

showing that $v=\alpha w$ is also sufficient for equality in (26.1). □

Lemma 26.3 is an abstract version of the Cauchy–Schwarz inequality for integrals, see Corollary 13.3. Just as in Chapter 13 we will use it to show that in an inner product space $(V, \langle\cdot\,,\cdot\rangle)$

$$\|v\| := \sqrt{\langle v, v\rangle}, \quad v \in V, \tag{26.2}$$

defines a norm, i.e. a map $\|\cdot\| : V \to [0,\infty)$ satisfying for all $v, w \in V$ and $\alpha \in \mathbb{C}$

$$\|v\| > 0 \iff v \neq 0 \tag{N_1}$$

$$\|\alpha v\| = |\alpha| \cdot \|v\|, \tag{N_2}$$

$$\|v+w\| \leqslant \|v\| + \|w\|, \tag{N_3}$$

(N_1) is called *definiteness*, (N_2) *positive homogeneity* and (N_3) *triangle inequality*.

Lemma 26.4 *$(V, \langle\cdot\,,\cdot\rangle^{1/2})$ is a normed space.*

Proof Because of (SP_1) the map $\|\cdot\| : V \to [0,\infty)$, $\|v\| := \sqrt{\langle v, v\rangle}$, is well-defined. All we have to do is to check the properties (N_1)–(N_3).

Obviously (SP_1) $\Leftrightarrow$ (N_1), and (N_2) follows from (SP_2), (SP_3):

$$\|\alpha v\|^2 = \langle \alpha v, \alpha v\rangle = \alpha\bar{\alpha}\langle v, v\rangle = |\alpha|^2 \cdot \|v\|^2.$$

The triangle inequality (N_3) is a consequence of the Cauchy–Schwarz inequality:

$$
\begin{aligned}
\|v+w\|^2 = \langle v+w, v+w\rangle &= \langle v,v\rangle + \langle v,w\rangle + \langle w,v\rangle + \langle w,w\rangle \\
&= \|v\|^2 + 2\,\mathrm{Re}\langle v,w\rangle + \|w\|^2 \\
&\leqslant \|v\|^2 + 2|\langle v,w\rangle| + \|w\|^2 \\
&\overset{(26.1)}{\leqslant} \|v\|^2 + 2\|v\|\cdot\|w\| + \|w\|^2 \\
&= (\|v\| + \|w\|)^2.
\end{aligned}
$$

□

Example 26.5 (i) The typical finite-dimensional inner product spaces are

$$
\begin{array}{cc}
\mathbb{R}^n \quad (\mathbb{R}\text{-vector space}) & \mathbb{C}^n \quad (\mathbb{C}\text{-vector space}) \\
\langle x,y\rangle = \sum_{i=1}^n x_i y_i & \langle z,w\rangle = \sum_{i=1}^n z_i \bar{w}_i \\
\|x\| = \left(\sum_{i=1}^n x_i^2\right)^{1/2} & \|z\| = \left(\sum_{i=1}^n |z_i|^2\right)^{1/2}.
\end{array}
$$

(ii) The typical separable[1] infinite-dimensional inner product spaces are

$$
\begin{array}{cc}
\ell^2_{\mathbb{R}}(\mathbb{N}) \quad (\mathbb{R}\text{-vector space}) & \ell^2_{\mathbb{C}}(\mathbb{N}) \quad (\mathbb{C}\text{-vector space}) \\
x = (x_i)_{i\in\mathbb{N}},\ y = (y_i)_{i\in\mathbb{N}} & z = (z_i)_{i\in\mathbb{N}},\ w = (w_i)_{i\in\mathbb{N}} \\
\langle x,y\rangle = (x,y)_{\ell^2} = \sum_{i=1}^\infty x_i y_i & \langle z,w\rangle = (z,w)_{\ell^2} = \sum_{i=1}^\infty z_i \bar{w}_i \\
\|x\| = \|x\|_{\ell^2} = \left(\sum_{i=1}^\infty x_i^2\right)^{1/2} & \|z\| = \|z\|_{\ell^2} = \left(\sum_{i=1}^\infty |z_i|^2\right)^{1/2}.
\end{array}
$$

(iii) Let $(X, \mathscr{A}, \mu)$ be a measure space. The typical general (finite and infinite-dimensional) inner product spaces are

$$
\begin{array}{cc}
L^2_{\mathbb{R}}(\mu) \quad (\mathbb{R}\text{-vector space}) & L^2_{\mathbb{C}}(\mu) \quad (\mathbb{C}\text{-vector space}) \\
\langle u,v\rangle = (u,v)_2 = \int uv\,d\mu & \langle f,g\rangle = (f,g)_2 = \int f\bar{g}\,d\mu \\
\|u\| = \|u\|_2 = \left(\int u^2\,d\mu\right)^{1/2} & \|f\| = \|f\|_2 = \left(\int |f|^2\,d\mu\right)^{1/2}.
\end{array}
$$

[1] *Separable* means that the space contains a countable dense subset, see Definition 26.23 below.

Every inner product space becomes a normed space with norm given by (26.2), but not every normed space is necessarily an inner product space. In fact, $L^p(\mu)$ or $\ell^p(\mathbb{N})$ are for all $1 \leqslant p \leqslant \infty$ normed spaces, but only for $p = 2$ inner product spaces. The reason for this is that in $L^p(\mu), p \neq 2$, the parallelogram law does not hold.

Lemma 26.6 (parallelogram law) *Let* $(V, \langle \cdot\,, \cdot \rangle)$ *be an inner product space.*

$$\left\| \frac{v+w}{2} \right\|^2 + \left\| \frac{v-w}{2} \right\|^2 = \frac{1}{2}\left(\|v\|^2 + \|w\|^2\right) \quad \forall v, w \in V. \tag{26.3}$$

Proof The proof is obvious. [✍] □

Geometrically $v + w$ and $v - w$ are the diagonals of the parallelogram spanned by the vectors v and w. The proof of (26.3) in $\mathbb{R}^n$ would show the *cosine law* for the angle $\sphericalangle(x, y)$ between the vectors $x, y \in \mathbb{R}^n$:

$$\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|} = \cos \sphericalangle(x, y). \tag{26.4}$$

On the other hand, we can use (26.4) to *define* a natural geometry on V which resembles in many respects the Euclidean geometry on $\mathbb{R}^n$.

Definition 26.7 Let $(V, \langle \cdot\,, \cdot \rangle)$ be an inner product space. We call $v, w \in V$ *orthogonal* and write $v \perp w$ if $\langle v, w \rangle = 0$.

Remark 26.8 (i) If $\|\cdot\|$ derives from a scalar product, we can recover $\langle \cdot\,, \cdot \rangle$ from $\|\cdot\|$ with the help of the so-called *polarization identities*: if $\mathbb{K} = \mathbb{R}$,

$$\langle v, w \rangle = \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right) = \frac{1}{2}\left(\|v+w\|^2 - \|v\|^2 - \|w\|^2\right), \tag{26.5}$$

and if $\mathbb{K} = \mathbb{C}$,

$$\langle v, w \rangle = \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2 + i\|v+iw\|^2 - i\|v-iw\|^2\right). \tag{26.6}$$

(ii) One can show that a norm $\|\cdot\|$ derives from a scalar product if, and only if, $\|\cdot\|$ satisfies the parallelogram identity (26.3). For a proof we refer to Yosida [59, p. 39], see also Problem 26.2.

(iii) Let $V = V_{\mathbb{R}}$ be an $\mathbb{R}$-inner product space with scalar product $\langle \cdot\,, \cdot \rangle_{\mathbb{R}}$. Then we can turn V into a $\mathbb{C}$-inner product space using the following *complexification procedure*:

$$V_{\mathbb{C}} := V_{\mathbb{R}} \oplus iV_{\mathbb{R}} = \left\{v + iw : v, w \in V_{\mathbb{R}}\right\},$$

with the following addition

$$(v + iw) + (v' + iw') := (v + v') + i(w + w'), \quad v, v', w, w' \in V_{\mathbb{R}},$$

scalar multiplication

$$(\alpha + i\beta)(v + iw) := (\alpha v - \beta w) + i(\beta v + \alpha w), \quad \alpha, \beta \in \mathbb{R},\ v, w \in V_{\mathbb{R}},$$

inner product

$$\langle v + iw, v' + iw' \rangle_{\mathbb{C}} := \langle v, v' \rangle + i\langle w, v' \rangle - i\langle v, w' \rangle + \langle w, w' \rangle, \quad v, v', w, w' \in V_{\mathbb{R}},$$

and norm $\|\cdot\| := \langle \cdot\,, \cdot \rangle_{\mathbb{C}}^{1/2}$.

Convergence and Completeness

For the analysis on an inner product space the notions of convergence and completeness are essential ingredients. Since $(V, \|\cdot\| := \langle \cdot\,, \cdot \rangle^{1/2})$ is a normed space, we have a natural notion of *convergence*:[2] a sequence $(v_n)_{n\in\mathbb{N}} \subset V$ *converges* to an element $v \in V$ if $(\|v - v_n\|)_{n\in\mathbb{N}}$ converges to 0 in $\mathbb{R}$,

$$\lim_{n\to\infty} v_n = v \iff \lim_{n\to\infty} \|v - v_n\| = 0.$$

But it is *completeness* and the study of *Cauchy sequences* in V,

$$(v_n)_{n\in\mathbb{N}} \subset V \text{ Cauchy sequence} \iff \lim_{k,n\to\infty} \|v_n - v_k\| = 0,$$

that gets the analysis really going. This leads to the following very natural definition.

Definition 26.9 A *Hilbert space* $\mathcal{H}$ is a complete inner product space, i.e. an inner product space where every Cauchy sequence converges.

Example 26.10 The spaces $\mathbb{R}^n$, $\mathbb{C}^n$, $\ell^2_{\mathbb{K}}$ and $L^2_{\mathbb{K}}(\mu)$ over an arbitrary measure space $(X, \mathscr{A}, \mu)$ are Hilbert spaces and, indeed, the 'typical' ones. This follows from Example 26.5 and the Riesz–Fischer theorem, Theorem 13.7.

Since every Hilbert space is an inner product space, we have the notion of orthogonality of $g, h \in \mathcal{H}$, see Definition 26.7:

$$g \perp h \iff \langle g, h \rangle = 0.$$

Definition 26.11 Let $\mathcal{H}$ be a Hilbert space. The *orthogonal complement* $M^\perp$ of a subset $M \subset \mathcal{H}$ is by definition

$$\begin{aligned} M^\perp &:= \{h \in \mathcal{H} : h \perp m \quad \forall m \in M\} \\ &= \{h \in \mathcal{H} : \langle h, m \rangle = 0 \quad \forall m \in M\}. \end{aligned} \tag{26.7}$$

[2] Compare this with Chapter 13, page 118.

Lemma 26.12 *Let $\mathcal{H}$ be a Hilbert space and $M \subset \mathcal{H}$ be any subset. The orthogonal complement $M^\perp$ is a closed linear subspace of $\mathcal{H}$ and $M \subset (M^\perp)^\perp$.*

Proof If $g, h \in M^\perp$ we find for all $\alpha, \beta \in \mathbb{C}$ that

$$\langle \alpha g + \beta h, m\rangle = \alpha\langle g, m\rangle + \beta\langle h, m\rangle = 0 \quad \forall m \in M,$$

i.e. $\alpha g + \beta h \in M^\perp$ and $M^\perp$ is a linear subspace of $\mathcal{H}$. To see the closedness we take a sequence $(h_n)_{n\in\mathbb{N}} \subset M^\perp$ such that $\lim_{n\to\infty} h_n = h$. Then, for all $m \in M$,

$$|\langle h, m\rangle| = |\langle h, m\rangle - \underbrace{\langle h_n, m\rangle}_{=0}| = |\langle h - h_n, m\rangle| \overset{26.3}{\leqslant} \|h - h_n\| \cdot \|m\| \xrightarrow[n\to\infty]{} 0;$$

this shows that $M^\perp$ is closed since $h \in M^\perp$. Finally, if $m \in M$ we get

$$0 = \langle h, m\rangle = \overline{\langle m, h\rangle} \quad \forall h \in M^\perp \implies m \in (M^\perp)^\perp. \qquad \square$$

The next theorem is central for the study of (the geometry of) Hilbert spaces. Recall that a set $C \subset \mathcal{H}$ is *convex* if

$$u, w \in C \implies tu + (1-t)w \in C \quad \forall t \in (0, 1).$$

Theorem 26.13 (projection theorem) *Let $C \neq \emptyset$ be a closed convex subset of the Hilbert space $\mathcal{H}$. For every $h \in \mathcal{H}$ there is a unique minimizer $u \in C$ such that*

$$\|h - u\| = \inf_{w\in C} \|h - w\| =: d(h, C). \tag{26.8}$$

This element $u = P_C h$ is called (orthogonal) projection *of h onto C and is equally characterized by the property*

$$P_C h \in C \quad \textit{and} \quad \operatorname{Re}\langle h - P_C h, w - P_C h\rangle \leqslant 0 \quad \forall w \in C. \tag{26.9}$$

Proof Existence. Let $d := \inf_{w\in C} \|h - w\|$. By the very definition of the infimum, there is a sequence $(w_n)_{n\in\mathbb{N}} \subset C$ such that

$$\lim_{n\to\infty} \|h - w_n\| = d.$$

If we can show that $(w_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, then the limit $u := \lim_{n\to\infty} w_n$ exists because of the completeness of $\mathcal{H}$ and it is in C since C is closed. Applying the parallelogram law (26.3) with $v = h - w_n$ and $w = h - w_k$ gives

$$\left\| h - \frac{w_n + w_k}{2} \right\|^2 + \left\| \frac{w_n - w_k}{2} \right\|^2 = \frac{1}{2}\big(\|h - w_n\|^2 + \|h - w_k\|^2\big).$$

Since C is convex, $\frac{1}{2}w_n + \frac{1}{2}w_k \in C$, thus $d \leqslant \left\|h - \frac{1}{2}(w_n + w_k)\right\|$ and

$$d^2 + \frac{1}{4}\|w_n - w_k\|^2 \leqslant \frac{1}{2}\big(\|h - w_n\|^2 + \|h - w_k\|^2\big) \xrightarrow[k,n\to\infty]{} d^2.$$

This proves that $(w_n)_{n\in\mathbb{N}}$ is a Cauchy sequence.

Uniqueness. Assume that both $u\in C$ and $\tilde{u}\in C$ satisfy (26.8), i.e.

$$\|u-h\| = d = \|\tilde{u}-h\|.$$

Since by convexity $\frac{1}{2}u+\frac{1}{2}\tilde{u}\in C$, the parallelogram law (26.3) gives

$$d^2 \leqslant \underbrace{\left\|h-\left(\frac{1}{2}u+\frac{1}{2}\tilde{u}\right)\right\|^2}_{\geqslant d^2} + \left\|\frac{1}{2}(u-\tilde{u})\right\|^2 = \frac{1}{2}\left(\|h-u\|^2+\|h-\tilde{u}\|^2\right) = d^2,$$

and we conclude that $\|u-\tilde{u}\|^2=0$ or $u=\tilde{u}$.

Equivalence of (26.8)*,* (26.9). Assume that $u\in C$ satisfies (26.8) and let $w\in C$. By convexity, $(1-t)u+tw\in C$ for all $t\in(0,1)$ and by (26.8)

$$\begin{aligned}\|h-u\|^2 &\leqslant \|h-(1-t)u-tw\|^2\\ &= \|(h-u)-t(w-u)\|^2\\ &= \|h-u\|^2 - 2t\,\mathrm{Re}\langle h-u, w-u\rangle + t^2\|w-u\|^2.\end{aligned}$$

Hence, $2\,\mathrm{Re}\langle h-u,w-u\rangle \leqslant t\|w-u\|^2$ and (26.9) follows as $t\to 0$.

Conversely, if (26.9) holds, we have for $u=P_Ch\in C$

$$\|h-u\|^2 - \|h-w\|^2 = 2\,\mathrm{Re}\langle h-u,w-u\rangle - \|u-w\|^2 \leqslant 0 \quad \forall w\in C,$$

which implies (26.8). □

We will now study the properties of the projection operator P_C. If $V, W\subset\mathcal{H}$ are two subspaces with $V\cap W=\{0\}$, we call $V+W=\{v+w : v\in V,\ w\in W\}$ the *direct sum* and write $V\oplus W$.

Corollary 26.14 (i) *Let $\emptyset\neq C\subset\mathcal{H}$ be a closed convex subset. The projection $P_C:\mathcal{H}\to C$ is a contraction, i.e.*

$$\|P_Cg-P_Ch\| \leqslant \|g-h\| \quad \forall g,h\in\mathcal{H}. \tag{26.10}$$

(ii) *If $\emptyset\neq C=F$ is a closed linear subspace of $\mathcal{H}$, P_F is a linear operator and $f=P_Fh$ is the unique element with*

$$f\in F \quad \text{and} \quad h-f\in F^{\perp}. \tag{26.11}$$

In particular, $\mathcal{H}=F\oplus F^{\perp}$.

(iii) *If F is not closed, then $\mathcal{H}=\bar{F}\oplus F^{\perp}$ or, equivalently, $\bar{F}=(F^{\perp})^{\perp}$.*

Proof (i) follows from the inequality

$$\begin{aligned}\|P_C g - P_C h\|^2 &= \operatorname{Re}\big(\langle P_C g - g, P_C g - P_C h\rangle + \langle P_C h - h, P_C h - P_C g\rangle \\ &\qquad + \langle g - h, P_C g - P_C h\rangle\big) \\ &\overset{(26.9)}{\leqslant} \operatorname{Re}\langle g - h, P_C g - P_C h\rangle \\ &\leqslant \|g - h\| \cdot \|P_C g - P_C h\|,\end{aligned}$$

where we use the Cauchy–Schwarz inequality Lemma 26.3 for the last estimate.

(ii) Since F is a linear subspace, $\zeta v \in F$ for all $v \in F, \zeta \in \mathbb{C}$, and (26.9) reads in this case as

$$\operatorname{Re}\langle h - P_F h, \zeta v - P_F h\rangle \leqslant 0 \qquad \forall \zeta \in \mathbb{C},\ v \in F,$$

or, equivalently,

$$\operatorname{Re}\big(\zeta\langle h - P_F h, v\rangle\big) \leqslant \operatorname{Re}\langle h - P_F h, P_F h\rangle \qquad \forall \zeta \in \mathbb{C},\ v \in F,$$

which is possible only if $\langle h - P_F h, v\rangle = 0$ for all $v \in F$; taking, in particular, $v = P_F h$ we get $\langle h - P_F h, P_F h\rangle = 0$, and (26.11) follows.

If, on the other hand, (26.11) is true, we get for all $v \in F$

$$0 = \operatorname{Re}\langle h - f, v\rangle - \operatorname{Re}\langle h - f, f\rangle = \operatorname{Re}\langle h - f, v - f\rangle,$$

and $f = P_F h$ follows by the uniqueness of the projection.

The orthogonal decomposition $\mathcal{H} = F \oplus F^\perp$ follows immediately as $h = P_F h + (h - P_F h)$ and $h \in F \cap F^\perp \iff \langle h, h\rangle = 0 \iff h = 0$. The decomposition also proves the linearity of P_F since for all $g, h \in \mathcal{H}$ and $\alpha, \beta \in K$

$$\Big\langle \underbrace{(\alpha g - \alpha P_F g)}_{\in F^\perp} + \underbrace{(\beta h - \beta P_F h)}_{\in F^\perp},\ \underbrace{\alpha P_F g + \beta P_F h}_{\in F} \Big\rangle = 0$$

as well as

$$\big\langle (\alpha g + \beta h) - P_F(\alpha g + \beta h), P_F(\alpha g + \beta h)\big\rangle = 0,$$

which implies, by the uniqueness of the projection, that $P_F(\alpha g + \beta h) = \alpha P_F g + \beta P_F h$.

(iii) We know from Lemma 26.12 that $F \subset (F^\perp)^\perp$ and that $(F^\perp)^\perp$ is closed; therefore, $\bar{F} \subset (F^\perp)^\perp$. Moreover, $F \subset \bar{F}$ implies $\bar{F}^\perp \subset F^\perp$, [✍] showing that

$$\mathcal{H} \overset{26.14\text{(ii)}}{=} \bar{F} \oplus (\bar{F})^\perp \subset \bar{F} + F^\perp \subset (F^\perp)^\perp \oplus F^\perp \overset{26.14\text{(ii)}}{=} \mathcal{H},$$

and $\mathcal{H} = \bar{F} \oplus F^\perp$ or $\bar{F} = (F^\perp)^\perp$ follows. $\square$

Remark 26.15 (i) It is easy to show that the projection P_F onto a subspace $F \subset \mathcal{H}$ is *symmetric*, i.e. that

$$\langle P_F g, h\rangle = \langle g, P_F h\rangle \quad \forall g, h \in \mathcal{H}, \tag{26.12}$$

and that $P_F^2 = P_F$, i.e.

$$\langle P_F^2 g, h\rangle = \langle P_F g, P_F h\rangle = \langle P_F g, h\rangle \quad \forall g, h \in \mathcal{H}. \tag{26.13}$$

In fact, (26.13) implies (26.12). Since $P_F g \in F$, $P_F(P_F g) = P_F g$ by the uniqueness of the projection and $\langle P_F^2 g, h\rangle = \langle P_F g, h\rangle$ follows. Finally,

$$\langle P_F g, h\rangle = \langle P_F g, P_F h\rangle + \underbrace{\langle P_F g, h - P_F h\rangle}_{=0} = \langle P_F g, P_F h\rangle.$$

(ii) *Pythagoras' theorem* has a particularly nice form for projections:

$$\|h\|^2 = \|P_F h\|^2 + \|h - P_F h\|^2 \quad \forall h \in \mathcal{H}. \tag{26.14}$$

(iii) A very useful interpretation of Corollary 26.14(iii) is the following: a linear subspace $F \subset \mathcal{H}$ is dense in $\mathcal{H}$ if, and only if, $F^\perp = \{0\}$. In other words,

$$F \subset \mathcal{H} \text{ is dense} \iff \langle f, h\rangle = 0 \quad \forall f \in F \text{ entails } h = 0.$$

Let us briefly discuss two important consequences of the projection theorem: F. Riesz's representation theorem on the structure of continuous linear functionals on the Hilbert space $\mathcal{H}$ and the problem of finding a basis in $\mathcal{H}$.

Definition 26.16 A *continuous linear functional* on $\mathcal{H}$ is a map $\Lambda : \mathcal{H} \to \mathbb{K}$, $h \mapsto \Lambda(h)$ which is linear,

$$\Lambda(\alpha g + \beta h) = \alpha\Lambda(g) + \beta\Lambda(h) \quad \forall \alpha, \beta \in \mathbb{K},\ \forall g, h \in \mathcal{H}$$

and satisfies for some constant $c(\Lambda) \in (0, \infty)$ – which depends only on Λ –

$$|\Lambda(g - h)| \leqslant c(\Lambda)\|g - h\| \quad \forall g, h \in \mathcal{H}.$$

It is easy to find examples of continuous linear functionals on $\mathcal{H}$. Just fix some $g \in \mathcal{H}$ and set

$$\mathcal{H} \ni h \mapsto \Lambda_g(h) := \langle h, g\rangle. \tag{26.15}$$

Linearity is clear, and the Cauchy–Schwarz inequality (Lemma 26.3) shows that

$$|\Lambda_g(h - \tilde{h})| = |\langle h - \tilde{h}, g\rangle| \leqslant \underbrace{\|g\|}_{=c(\Lambda)} \|h - \tilde{h}\|.$$

That, in fact, *all* continuous linear functionals of $\mathcal{H}$ arise in this way is the content of the next theorem, due to F. Riesz.

Theorem 26.17 (Riesz representation) *Every continuous linear functional Λ on the Hilbert space $\mathcal{H}$ is of the form* (26.15)*, i.e. there exists a unique $g \in \mathcal{H}$ such that*

$$\Lambda(h) = \Lambda_g(h) = \langle h, g \rangle \quad \forall h \in \mathcal{H}.$$

Proof Set $F := \Lambda^{-1}\{0\}$ which is, due to the continuity and by linearity of Λ, a closed linear subspace of $\mathcal{H}$. [✍] If $F = \mathcal{H}$, $\Lambda \equiv 0$ and $g = 0 \in \mathcal{H}$ does the job. Otherwise we can pick some $g_0 \in \mathcal{H} \setminus F$ and set

$$g := \frac{g_0 - P_F g_0}{\|g_0 - P_F g_0\|} \overset{(26.11)}{\in} F^{\perp} \implies \Lambda(g) \neq 0.$$

Since $\mathcal{H} = F \oplus F^{\perp}$, we can write every $h \in \mathcal{H}$ in the form

$$h = \frac{\Lambda(h)}{\Lambda(g)} g + \Big(h - \frac{\Lambda(h)}{\Lambda(g)} g \Big) \in F^{\perp} \oplus F,$$

hence

$$\begin{aligned} \Big\langle h - \frac{\Lambda(h)}{\Lambda(g)} g, \frac{\Lambda(h)}{\Lambda(g)} g \Big\rangle = 0 \iff \langle h, g \rangle &= \frac{\Lambda(h)}{\Lambda(g)} \underbrace{\langle g, g \rangle}_{=1} \\ \iff \Lambda(h) &= \big\langle h, \overline{\Lambda(g)} g \big\rangle, \end{aligned}$$

and the proof is finished. □

We will finally see how to represent elements of a Hilbert space using an orthonormal base (ONB, for short). We begin with a definition.

Definition 26.18 Let $\mathcal{H}$ be a Hilbert space. (i) The (*linear*) *span* of a family given by $\{e_n : n = 1, 2, \ldots, N\} \subset \mathcal{H}$, $N \in \mathbb{N} \cup \{\infty\}$, is the set of all *finite* linear combinations of the e_n, i.e.

$$\operatorname{span}\{e_1, e_2, \ldots, e_N\} = \Big\{ \sum_{n=1}^{N} \alpha_n e_n : \alpha_1, \ldots, \alpha_N \in \mathbb{R} \Big\}.$$

(ii) A sequence $(e_k)_{k \in \mathbb{N}} \subset \mathcal{H}$ is said to be a (*countable*) *orthonormal system* (ONS, for short) if

$$\langle e_k, e_n \rangle = \begin{cases} 0, & \text{if } k \neq n, \\ 1, & \text{if } k = n, \end{cases}$$

that is, $\|e_n\| = 1$ and $e_k \perp e_n$ whenever $k \neq n$.

Theorem 26.19 *Let $(e_n)_{n \in \mathbb{N}}$ be an orthonormal system in the Hilbert space $\mathcal{H}$ and denote by $E = E(N) = \operatorname{span}\{e_1, \ldots, e_N\}$ the linear span of $e_1, \ldots, e_N$, $N \in \mathbb{N}$.*

(i) *$E = E(N)$ is a closed linear subspace, $P_E g = \sum_{n=1}^{N} \langle g, e_n\rangle e_n$ and*

$$\left\| g - \sum_{n=1}^{N} \langle g, e_n\rangle e_n \right\| < \|g - f\| \quad \forall f \in E, f \neq P_E g,$$

and also $\|P_E g\|^2 = \sum_{n=1}^{N} |\langle g, e_n\rangle|^2$.

(ii) (Pythagoras' theorem). *For $g \in \mathcal{H}$*

$$\|g\|^2 = \|g - P_E g\|^2 + \|P_E g\|^2 = \left\| g - \sum_{n=1}^{N} \langle g, e_n\rangle e_n \right\|^2 + \sum_{n=1}^{N} |\langle g, e_n\rangle|^2.$$

(iii) (Bessel's inequality). *For $g \in \mathcal{H}$*

$$\sum_{n=1}^{\infty} |\langle g, e_n\rangle|^2 \leqslant \|g\|^2.$$

(iv) (Parseval's identity). *The sequence $\left(\sum_{n=1}^{m} c_n e_n\right)_{m\in\mathbb{N}}$, $c_n \in \mathbb{K}$, converges to an element $g \in \mathcal{H}$ if, and only if, $\sum_{n=1}^{\infty} |c_n|^2 < \infty$. In this case, Parseval's identity holds:*

$$\sum_{n=1}^{\infty} |c_n|^2 = \sum_{n=1}^{\infty} |\langle g, e_n\rangle|^2 = \|g\|^2.$$

Proof (i) That $E(N)$ is a linear subspace is due to the very definition of 'span'. The closedness follows from the fact that $E(N)$ is generated by finitely many e_n: if $f \in E(N)$ is of the form $f = \sum_{n=1}^{N} c_n e_n$, $c_n \in \mathbb{K}$, then

$$\langle f, e_k\rangle = \left\langle \sum_{n=1}^{N} c_n e_n, e_k \right\rangle = \sum_{n=1}^{N} c_n \langle e_n, e_k\rangle = c_k.$$

Let $(f^{(i)})_{i\in\mathbb{N}} \subset E(N)$ be a sequence with $f^{(i)} \to f \in \mathcal{H}$. Then

$$\begin{aligned}
\left\| f^{(i)} - \sum_{n=1}^{N} \langle f, e_n\rangle e_n \right\| &= \left\| \sum_{n=1}^{N} \langle f^{(i)} - f, e_n\rangle e_n \right\| \\
&\leqslant \sum_{n=1}^{N} \left|\langle f^{(i)} - f, e_n\rangle\right| \cdot \underbrace{\|e_n\|}_{=1} \\
&\overset{26.3}{\leqslant} \sum_{n=1}^{N} \|f^{(i)} - f\| \\
&= N\|f^{(i)} - f\| \xrightarrow[i\to\infty]{} 0,
\end{aligned}$$

which shows that $\lim_{i\to\infty} f^{(i)} = \sum_{n=1}^{N} \langle f, e_n\rangle e_n \in E(N)$.

If $g \in \mathcal{H}$, we observe that $g - \sum_{n=1}^{N} \langle g, e_n\rangle e_n \perp e_k$ for all $k = 1, 2, \dots, N$, since for these k

$$\Big\langle g - \sum_{n=1}^{N} \langle g, e_n\rangle e_n, e_k \Big\rangle = \langle g, e_k\rangle - \sum_{n=1}^{N} \langle g, e_n\rangle \langle e_n, e_k\rangle$$
$$= \langle g, e_k\rangle - \langle g, e_k\rangle = 0.$$

Since $\mathcal{H} = E(N) \oplus E(N)^{\perp}$, we get $P_{E(N)}g = \sum_{n=1}^{N} \langle g, e_n\rangle e_n$, while (26.8) implies $\big\|g - \sum_{n=1}^{N} \langle g, e_n\rangle e_n\big\| \leqslant \|g - f\|$ for $f \in E(N)$, with equality holding only if $f = P_{E(N)}g$ because of uniqueness of $P_{E(N)}g$. Finally,

$$\|P_{E(N)}g\|^2 = \langle P_{E(N)}g, P_{E(N)}g\rangle = \Big\langle \sum_{n=1}^{N} \langle g, e_n\rangle e_n, \sum_{k=1}^{N} \langle g, e_k\rangle e_k \Big\rangle$$
$$= \sum_{k,n=1}^{N} \langle g, e_n\rangle \overline{\langle g, e_k\rangle} \langle e_n, e_k\rangle = \sum_{n=1}^{N} |\langle g, e_n\rangle|^2,$$

where we use that e_n is an ONS.

(ii) follows from (26.14) and (i).

(iii) From (ii) we get for all $N \in \mathbb{N}$

$$\sum_{n=1}^{N} |\langle g, e_n\rangle|^2 = \|g\|^2 - \|g - P_E g\| \leqslant \|g\|^2.$$

Since the right-hand side is independent of $N \in \mathbb{N}$, we can let $N \to \infty$ and the claim follows.

(iv) Since $\mathcal{H}$ is complete, it sufficies to show that $\big(\sum_{k=1}^{n} c_k e_k\big)_{n\in\mathbb{N}}$ is a Cauchy sequence. Because of the orthogonality of the e_k we see (as in (i))

$$\Big\| \sum_{k=m-1}^{n} c_k e_k \Big\|^2 = \sum_{k=m-1}^{n} |c_k|^2 \|e_k\|^2 = \sum_{k=m-1}^{n} |c_k|^2,$$

which means that $\big(\sum_{k=1}^{n} c_k e_k\big)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}$ if, and only if, the series $\sum_{n=1}^{\infty} |c_n|^2$ converges. In the latter case, Parseval's identity follows from (iii): for $g = \sum_{n=1}^{\infty} c_n e_n$ we have $P_{E(N)}g = \sum_{n=1}^{N} c_n e_n$ and $c_n = \langle g, e_n\rangle$ by (i). Thus,

$$\|g\|^2 \overset{\text{(ii)}}{=} \|g - P_{E(N)}g\|^2 + \sum_{n=1}^{N} |\langle g, e_n\rangle|^2 \xrightarrow[N\to\infty]{} \sum_{n=1}^{\infty} |\langle g, e_n\rangle|^2 = \sum_{n=1}^{\infty} |c_n|^2. \quad \square$$

Two questions remain: can we always find a countable ONS? If so, can we use it to represent all elements of $\mathcal{H}$? The answer to the first question is 'yes', while the second question has to be answered by 'no', unless we are looking at *separable* Hilbert spaces, see Definition 26.23 below. We will restrict ourselves to the latter situation but we will point the reader towards references where the general case is treated.

Definition 26.20 An ONS $(e_n)_{n\in\mathbb{N}}$ in the Hilbert space $\mathcal{H}$ is said to be *maximal* (also *complete*, *total*, *an orthonormal basis*) if for every $g \in \mathcal{H}$

$$\langle g, e_n \rangle = 0 \quad \forall n \in \mathbb{N} \implies g = 0.$$

The idea behind maximality is that we can obtain $\mathcal{H}$ as the limit of finite-dimensional projections, '$\mathcal{H} = \lim_N P_{\operatorname{span}\{e_1,\dots,e_N\}}\mathcal{H}$' or '$\mathcal{H} = \bigoplus_{n\in\mathbb{N}}\{e_n\}$', if the limits and summations are understood in the right way. Here we see that the countability of the ONS entails that $\mathcal{H}$ can be represented as closure of the span of countably many elements – and that this is indeed a restriction should be obvious. Let us make all this more precise.

Theorem 26.21 *Let $(e_n)_{n\in\mathbb{N}}$ be an ONS in the Hilbert space $\mathcal{H}$. Then the following assertions are equivalent:*

(i) $(e_n)_{n\in\mathbb{N}}$ *is maximal;*

(ii) $\displaystyle\bigcup_{N=1}^{\infty} E(N)$ *is dense in $\mathcal{H}$ where $E(N) = \operatorname{span}\{e_1, \dots, e_N\}$;*

(iii) $\displaystyle g = \sum_{n=1}^{\infty} \langle g, e_n \rangle e_n \quad \forall g \in \mathcal{H};$

(iv) $\displaystyle \sum_{n=1}^{\infty} |\langle g, e_n \rangle|^2 = \|g\|^2 \quad \forall g \in \mathcal{H};$

(v) $\displaystyle \sum_{n=1}^{\infty} \langle g, e_n \rangle \overline{\langle h, e_n \rangle} = \langle g, h \rangle \quad \forall g, h \in \mathcal{H}.$

Proof (i) $\Rightarrow$ (ii). Since $F := \bigcup_{N\in\mathbb{N}} \operatorname{span}\{e_1, \dots, e_N\} = \operatorname{span}\{e_n : n \in \mathbb{N}\}$ is a linear subspace of $\mathcal{H}$, the assertion follows from the definition of maximality and Remark 26.15(iii).

(ii) $\Rightarrow$ (iii) is obvious since

$$\sum_{j=1}^{\infty} \langle g, e_j \rangle e_j = \lim_{N\to\infty} \sum_{j=1}^{N} \langle g, e_j \rangle e_j = \lim_{N\to\infty} P_{E(N)} g.$$

(iii) $\Rightarrow$ (iv) follows from Theorem 26.19(iv).

(iv) $\Rightarrow$ (v) follows from the polarization identity (26.6).

(v) $\Rightarrow$ (i). If $\langle u, e_n \rangle = 0$ for some $u \in \mathcal{H}$ and all $n \in \mathbb{N}$, we get from (v) with $g = h = u$ that

$$0 = \sum_{n=1}^{\infty} \langle u, e_n \rangle \overline{\langle u, e_n \rangle} = \langle u, u \rangle = \|u\|^2,$$

and therefore $u = 0$. $\square$

Theorem 26.21 solves the representation issue. To find an orthonormal system, we recall first what we do in a finite-dimensional vector space V to get a basis. If $V = \operatorname{span}\{v_1, \ldots, v_N\}$, we remove recursively all $v_1', \ldots, v_k'$ such that still $V = \operatorname{span}\big(\{v_1, \ldots, v_N\} \setminus \{v_1', \ldots, v_k'\}\big)$. This procedure gives us in at *most N steps* a minimal system $\{w_1, \ldots, w_n\} \subset \{v_1, \ldots, v_N\}$, $N = n + k$, with the property that $V = \operatorname{span}\{w_1, \ldots, w_n\}$. Note that this is, at the same time, a maximally independent system of vectors in V. We can now rebuild $\{w_1, \ldots, w_n\}$ into an ONS by the following procedure.

Gram–Schmidt orthonormalization procedure 26.22

$$\left.\begin{aligned} e_1 &:= \frac{w_1}{\|w_1\|}, \quad \text{and recursively} \\ \tilde{e}_{k+1} &:= w_{k+1} - P_{\operatorname{span}\{e_1, \ldots, e_k\}} w_{k+1} \\ &= w_{k+1} - \sum_{i=1}^{k} \langle w_{k+1}, e_i \rangle e_i, \\ e_{k+1} &:= \frac{\tilde{e}_{k+1}}{\|\tilde{e}_{k+1}\|}. \end{aligned}\right\} \tag{26.16}$$

Another interpretation of (26.16) is this: if we had unleashed the Gram–Schmidt procedure on the set $\{v_1, \ldots, v_N\}$, we would have obtained again n orthonormal vectors, [✍] say, $f_1, \ldots, f_n$ (which are, in general, different from $e_1, \ldots, e_n$ constructed from $w_1, \ldots, w_n$). A close inspection of (26.16) shows that at each step $V = \operatorname{span}\{f_1, \ldots, f_k, v_{k+1}, \ldots, v_N\}$, so that (26.16) extends an partially existing basis $\{f_1, \ldots, f_k\}$ to a full ONB $\{f_1, \ldots, f_n\}$. This means that (26.16) is also a 'basis extension procedure'.

To get (26.16) to work in infinite dimensions we must make sure that $\mathcal{H}$ is the closure of the span of countably many vectors. This motivates the following convenient (but somewhat restrictive) definition.

Definition 26.23 A Hilbert space $\mathcal{H}$ is said to be *separable* if $\mathcal{H}$ contains a countable dense subset $G \subset \mathcal{H}$.

Theorem 26.24 *Every separable Hilbert space $\mathcal{H}$ has a maximal ONS.*

Proof Let $G = \{g_n\}_{n\in\mathbb{N}}$ be an enumeration of some countable dense subset of $\mathcal{H}$. Consider the subspaces $F_k = \operatorname{span}\{g_1, \dots, g_k\}$. Note that $F_k \subset F_{k+1}$, $\dim F_k \leqslant k$ and that $\bigcup_{k\in\mathbb{N}} F_k$ is dense in $\mathcal{H}$. Now construct an ONB in the finite-dimensional space F_k and extend this ONB using (26.16) to an ONB in F_{k+1}, etc. This produces a sequence $(e_n)_{n\in\mathbb{N}}$ of orthonormal elements in $\mathcal{H}$ for which it holds that $\operatorname{span}\{e_n : n \in \mathbb{N}\} = \bigcup_{k\in\mathbb{N}} F_k = G$ is dense in $\mathcal{H}$, and Theorem 26.21 completes the proof. □

Remark 26.25 (i) Assume that $\mathcal{H}$ is *separable*. Then we have the following 'algebraic' interpretation of the results in 26.19–26.24. Consider the maps

coordinate projection	(re-)construction map
$\Phi : \mathcal{H} \to \ell^2_{\mathbb{K}}(\mathbb{N})$	$\Psi : \ell^2_{\mathbb{K}}(\mathbb{N}) \to \mathcal{H}$
$g \mapsto (\langle g, e_n\rangle)_{n\in\mathbb{N}}$	$(c_n)_{n\in\mathbb{N}} \mapsto \sum_{n=1}^{\infty} c_n e_n.$

Because of Theorem 26.19(iv), both Φ and Ψ are well-defined maps, and Theorem 26.19 shows that Diagram 1 in Fig. 26.1 commutes, i.e. $\Phi \circ \Psi = \mathrm{id}_{\ell^2_{\mathbb{K}}(\mathbb{N})}$.

$$\begin{tikzcd} \ell^2_{\mathbb{K}}(\mathbb{N}) \arrow[r, "\Psi"] \arrow[dr, "\mathrm{id}"'] & \mathcal{H} \arrow[d, "\Phi"] \\ & \ell^2_{\mathbb{K}}(\mathbb{N}) \end{tikzcd} \qquad \begin{tikzcd} \mathcal{H} \arrow[r, "\Phi"] \arrow[dr, "\mathrm{id}"'] & \ell^2_{\mathbb{K}}(\mathbb{N}) \arrow[d, "\Psi"] \\ & \mathcal{H} \end{tikzcd}$$

Diagram 1 Diagram 2

Fig. 26.1. Diagrams illustrating Remark 26.25.

This means that, if we start with a square-summable sequence, associate with it an element from $\mathcal{H}$ and project to the coordinates, we get the original sequence back.

The converse operation, if we start with some $h \in \mathcal{H}$, project h down to its coordinates, and then try to reconstruct h from the (square integrable) coordinate sequence, is much more difficult, as we saw in Theorems 26.21 and 26.24.

Nevertheless, it can be done in every *separable* Hilbert space, and Diagram 2 in Fig. 26.1 becomes commutative, i.e. $\Psi \circ \Phi = \mathrm{id}_{\mathcal{H}}$.

This shows that every ***separable*** *Hilbert space* $\mathcal{H}$ *can be isometrically mapped onto* $\ell^2_{\mathbb{K}}(\mathbb{N})$. *The isometry is given by Parseval's identity* 26.19(iv):

$$\sum_{n\in\mathbb{N}} |\langle h, e_n\rangle|^2 = \|\Phi h\|^2_{\ell^2} = \|h\|^2_{\mathcal{H}}.$$

(ii) If $\mathcal{H}$ is ***not*** *separable*, we can still construct an ONB but now we need transfinite induction or Zorn's lemma. A reasonably short account is given in Rudin's book [43, pp. 82–87]. The results 26.19–26.21 carry over to this case if one makes some technical (what is an uncountable sum? etc.) modifications.

Problems

26.1. Show that the examples given in Example 26.5 are indeed inner product spaces.

26.2. This exercise shows the following theorem.

Theorem (Fréchet-von Neumann-Jordan). *An inner product* $\langle\cdot,\cdot\rangle$ *on the* $\mathbb{R}$*-vector space* V *derives from a norm if, and only if, the parallelogram identity* (26.3) *holds.*

(i) Necessity: prove Lemma 26.6.

Assume from now on that $\|\cdot\|$ is a norm satisfying (26.3) and set

$$(v, w) := \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right).$$

(ii) Show that (v, w) satisfies the properties (SP$_1$) and (SP$_2$) of Definition 26.2.

(iii) Prove that $(u+v, w) = (u, w) + (v, w)$.

(iv) Use (iii) to prove that $(qv, w) = q(v, w)$ for all dyadic numbers $q = n2^{-k}$, $n \in \mathbb{Z}$, $k \in \mathbb{N}_0$, and conclude that (SP$_3$) holds for dyadic α, β.

(v) Prove that the maps $t \mapsto \|tv + w\|$ and $t \mapsto \|tv - w\|$ $(t \in \mathbb{R},\ v, w \in V)$ are continuous and conclude that $t \mapsto (tv, w)$ is continuous. Use this and (iv) to show that (SP$_3$) holds for all $\alpha, \beta \in \mathbb{R}$.

26.3. (Continuation of Problem 26.2) Assume now that W is a $\mathbb{C}$-vector space with norm $\|\cdot\|$ satisfying the parallelogram identity (26.3) and let

$$(v, w)_{\mathbb{R}} := \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right).$$

Then $(v, w)_{\mathbb{C}} := (v, w)_{\mathbb{R}} + i(v, iw)_{\mathbb{R}}$ is a complex-valued inner product.

26.4. Does the norm $\|\cdot\|_1$ on $L^1([0,1], \mathscr{B}[0,1], \lambda^1|_{[0,1]})$ derive from an inner product?

26.5. Let $(V, \langle\cdot,\cdot\rangle)$ be a $\mathbb{C}$-inner product space, $n \in \mathbb{N}$ and set $\theta := e^{2\pi i/n}$.

(i) Show that

$$\frac{1}{n}\sum_{j=1}^{n} \theta^{jk} = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } 1 \leqslant k \leqslant n-1. \end{cases}$$

.

(ii) Use (i) to prove for all $n \geqslant 3$ the following generalization of (26.5) and (26.6):

$$\langle v, w\rangle = \frac{1}{n}\sum_{j=1}^{n} \theta^j \|v + \theta^j w\|^2.$$

(iii) Prove the following continuous version of (ii):

$$\langle v, w\rangle = \frac{1}{2\pi}\int_{(-\pi,\pi]} e^{i\phi}\left\|v + e^{i\phi}w\right\|^2 d\phi.$$

26.6. Let V be a real inner product space. Show that $v \perp w$ if, and only if, Pythagoras' theorem $\|v+w\|^2 = \|v\|^2 + \|w\|^2$ holds.

26.7. Show that every convergent sequence in $\mathcal{H}$ is a Cauchy sequence.

26.8. Show that $g \mapsto \langle g, h\rangle$, $h \in \mathcal{H}$, is continuous.

26.9. Show that $|||(g,h)||| := \left(\|g\|^p + \|h\|^p\right)^{1/p}$ is for every $p \geqslant 1$ a norm on $\mathcal{H}\times\mathcal{H}$. For which values of p does $\mathcal{H}\times\mathcal{H}$ become a Hilbert space?

26.10. Show that $(g,h)\mapsto \langle g,h\rangle$ and $(t,h)\mapsto th$ are continuous on $\mathcal{H}\times\mathcal{H}$, resp. $\mathbb{R}\times\mathcal{H}$.

26.11. Show that a Hilbert space $\mathcal{H}$ is separable if, and only if, $\mathcal{H}$ contains a countable maximal orthonormal system.

26.12. Let $w \in \mathcal{H} = L^2(X, \mathscr{A}, \mu)$ and show that $M_w^{\perp} := \left\{u \in L^2 : \int uw\, d\mu = 0\right\}^{\perp}$ is either $\{0\}$ or a one-dimensional subspace of $\mathcal{H}$.

26.13. Let $(e_n)_{n\in\mathbb{N}} \subset \mathcal{H}$ be an orthonormal system.

(i) Show that no subsequence of $(e_n)_{n\in\mathbb{N}}$ converges. However, for every $h \in \mathcal{H}$, $\lim_{n\to\infty}\langle e_n, h\rangle = 0$.
[Hint: show that it can't be a Cauchy sequence. Use Bessel's inequality.]

(ii) The *Hilbert cube*

$$Q := \left\{h \in \mathcal{H} : h = \sum_{n=1}^{\infty} c_n e_n,\ |c_n| \leqslant \frac{1}{n},\ n\in\mathbb{N}\right\}$$

is closed, bounded and compact (i.e. every sequence has a convergent subsequence).

(iii) The set

$$R := \bigcup_{n=1}^{\infty} \overline{B_{1/n}(e_n)}$$

is closed, bounded but not compact (compare this with (ii)).

(iv) The set

$$S := \left\{h \in \mathcal{H} : h = \sum_{n=1}^{\infty} c_n e_n,\ |c_n| \leqslant \delta_n,\ n\in\mathbb{N}\right\}$$

is closed, bounded and compact (cf. (ii)) if, and only if, $\sum_{n=1}^{\infty}\delta_n^2 < \infty$.

26.14. Let $\mathcal{H}$ be a real Hilbert space.

(i) Show that

$$\|h\| = \sup_{g\neq 0}\frac{|\langle g,h\rangle|}{\|g\|} = \sup_{\|g\|\leqslant 1}|\langle g,h\rangle| = \sup_{\|g\|=1}|\langle g,h\rangle|.$$

(ii) Can we replace in (i) $|\langle\cdot\,,\cdot\rangle|$ by $\langle\cdot\,,\cdot\rangle$?

(iii) Is it enough to take g in (i) from a dense subset rather than from $\mathcal{H}$ (resp. $\overline{B_1(0)}$ or $\{k\in\mathcal{H} : \|k\| = 1\}$)?

26.15. Show that the linear span of a sequence $(e_n)_{n\in\mathbb{N}} \subset \mathcal{H}$, $\operatorname{span}\{e_n : e_n \in \mathcal{H},\ n\in\mathbb{N}\}$, is a linear subspace of $\mathcal{H}$.

26.16. A weak form of the **uniform boundedness principle.** Consider the real Hilbert space $\ell^2 = \ell^2_{\mathbb{R}}(\mathbb{N})$ and let $a = (a_n)_{n\in\mathbb{N}}$ and $b = (b_n)_{n\in\mathbb{N}}$ be two sequences of real numbers.

(i) Assume that $\sum_{n=1}^{\infty} a_n^2 = \infty$. Construct a sequence $(n_k)_{k\in\mathbb{N}}$ such that $n_1 = 0$ and $\sum_{n_k < n \leqslant n_{k+1}} a_n^2 > 1$ for all $k \in \mathbb{N}$.

(ii) Define $b_n := \gamma_k a_n$ for all $n_k < n \leqslant n_{k+1}$, $k \in \mathbb{N}$ and show that one can determine the γ_k in such a way that $\sum_{n=1}^{\infty} a_n b_n = \infty$ while $\sum_{n=1}^{\infty} b_n^2 < \infty$.

(iii) Conclude that, if $\langle a, b\rangle < \infty$ for all $b \in \ell^2$, we have necessarily $a \in \ell^2$.

(iv) State and prove the analogue of (iii) for all separable Hilbert spaces.

Remark. The general *uniform boundedness principle* states that in every Hilbert space $\mathcal{H}$ and for any $H \subset \mathcal{H}$ one has

$$\sup_{h\in H} |\langle h, g\rangle| < \infty \quad \forall g \in \mathcal{H} \implies \sup_{h\in H} \|h\| < \infty.$$

Interpreting $\Lambda_h : g \mapsto \langle g, h\rangle$ as a linear map, this says that the boundedness of the orbits $\Lambda_h(\mathcal{H})$ for all $h \in H$ implies that the set H is bounded. This formulation persists even in Banach spaces. The proof is normally based on Baire's category theorem, see Rudin [43].

26.17. Let $F, G \subset \mathcal{H}$ be linear subspaces. An operator P defined on G is called ($\mathbb{K}$-)*linear* if $P(\alpha f + \beta g) = \alpha Pf + \beta Pg$ holds for all $\alpha, \beta \in \mathbb{K}$ and $f, g \in G$.

(i) Assume that F is closed and $P : \mathcal{H} \to F$ is the orthogonal projection. Then show that

$$P^2 = P \quad \text{and} \quad \langle Pg, h\rangle = \langle g, Ph\rangle \quad \forall g, h \in \mathcal{H}. \tag{26.17}$$

(ii) Show that if $P : \mathcal{H} \to \mathcal{H}$ is a map satisfying (26.17), then P is linear and P is the orthogonal projection onto the closed subspace $P(\mathcal{H})$.

(iii) Show that if $P : \mathcal{H} \to \mathcal{H}$ is a linear map satisfying

$$P^2 = P \quad \text{and} \quad \|Ph\| \leqslant \|h\| \quad \forall h \in \mathcal{H},$$

then P is the orthogonal projection onto the closed subspace $P(\mathcal{H})$.

26.18. Let $(X, \mathscr{A}, \mu)$ be a measure space and $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ be mutually disjoint sets such that $X = \dot{\bigcup}_{n\in\mathbb{N}} A_n$. Set

$$Y_n := \left\{ u \in L^2(\mu) : \int_{A_n^c} |u|^2 \, d\mu = 0 \right\}, \quad n \in \mathbb{N}.$$

(i) Show that $Y_n \perp Y_k$ if $n \neq k$.

(ii) Show that span $\left(\dot{\bigcup}_{n\in\mathbb{N}} Y_n\right)$ (i.e. the set of all linear combinations of *finitely* many elements from $\dot{\bigcup}_{n\in\mathbb{N}} Y_n$) is dense in $L^2(\mu)$.

(iii) Find the projection $P_n : L^2(\mu) \to Y_n$.

26.19. Let $(X, \mathscr{A}, \mu)$ be a measure space and assume that $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ is a sequence of pairwise disjoint sets such that $\dot{\bigcup}_{n\in\mathbb{N}} A_n = X$ and $0 < \mu(A_n) < \infty$. Then denote $\mathscr{A}_n := \sigma(A_1, A_2, \dots, A_n)$ and $\mathscr{A}_\infty := \sigma(A_n : n \in \mathbb{N})$.

(i) Show that $L^2(\mathscr{A}_n) \subset L^2(\mathscr{A})$ and that $L^2(\mathscr{A}_n)$ is a closed subspace.

(ii) Find an explicit formula for $E^{\mathscr{A}_n} u$, where $E^{\mathscr{A}_n}$ is the orthogonal projection $E^{\mathscr{A}_n} : L^2(\mathscr{A}) \to L^2(\mathscr{A}_n)$.

(iii) Determine the orthogonal complement of $L^2(\mathscr{A}_n)$.

(iv) Show that $\left(E^{\mathscr{A}_n} u\right)_{n\in\mathbb{N}\cup\{\infty\}}$, $u \in L^1(\mathscr{A}) \cap L^2(\mathscr{A})$, is a martingale.

(v) Show that $\lim_{n\to\infty} E^{\mathscr{A}_n} u = E^{\mathscr{A}_\infty} u$ a.e. and in L^2 for all $u \in L^1(\mathscr{A}) \cap L^2(\mathscr{A})$.

(vi) Conclude that $L^2(\mathscr{A}_\infty)$ is separable.

27

Conditional Expectations

Throughout this chapter $(X, \mathscr{A}, \mu)$ will be some measure space.

We have seen in Chapter 26 that $L^2_{\mathbb{C}}(\mathscr{A}) = L^2_{\mathbb{R}}(\mathscr{A}) \oplus iL^2_{\mathbb{R}}(\mathscr{A})$. By considering real and imaginary parts separately, we can reduce many assertions concerning $L^2_{\mathbb{C}}(\mathscr{A})$ to $L^2_{\mathbb{R}}(\mathscr{A})$. From Chapter 26 we know that $L^2_{\mathbb{R}}(\mathscr{A})$ is a Hilbert space with inner product, resp. norm,

$$\langle u, v\rangle = (u, v)_2 = \int uv\,d\mu \quad \text{resp.} \quad \|u\| = \|u\|_2 = \left(\int u^2\,d\mu\right)^{1/2}.$$

Since a function[1] $u \in L^2_{\mathbb{R}}(\mathscr{A})$ is only almost everywhere defined and since (square-) integrable functions with values in $\overline{\mathbb{R}}$ are almost everywhere finite (i.e. $\mathbb{R}$-valued), see Remark 13.5, we can identify $L^2_{\mathbb{R}}(\mathscr{A})$ and $L^2_{\overline{\mathbb{R}}}(\mathscr{A})$. We will do so and simply write $L^2(\mathscr{A})$.

Caution For $u, v \in L^2(\mathscr{A})$ expressions of the type $u = v$, $u \leqslant v \dots$ *always mean* $u(x) = v(x)$, $u(x) \leqslant v(x) \dots$ *for all x outside some μ-null set*.

In this chapter we are mainly interested in linear subspaces of $L^2(\mathscr{A})$ and projections onto them. One particularly important class arises in the following way: if $\mathscr{G} \subset \mathscr{A}$ is a sub-σ-algebra of $\mathscr{A}$, then any $\mathscr{G}$-measurable function is certainly $\mathscr{A}$-measurable. Since $(X, \mathscr{G}, \mu|_{\mathscr{G}})$ is also a measure space, it seems natural to interpret $L^2(\mathscr{G}, \mu|_{\mathscr{G}})$ (with norm $\|\cdot\|_{L^2(\mathscr{G})}$) as a subspace of $L^2(\mathscr{A}, \mu)$ (with norm $\|\cdot\|_{L^2(\mathscr{A})}$). This can indeed be done.

Lemma 27.1 *Let $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra of $\mathscr{A}$. The embedding*

$$\jmath : L^2(\mathscr{G}, \mu|_{\mathscr{G}}) \to L^2(\mathscr{A}, \mu), \quad u \mapsto u,$$

[1] Strictly speaking, we should call it an equivalence class of functions, see Remark 13.5.

is an isometry, i.e. a linear map satisfying $\|\jmath(u)\|_{L^2(\mathscr{A})} = \|u\|_{L^2(\mathscr{G})}$ *for all* $u \in L^2(\mathscr{G}, \mu|_{\mathscr{G}})$. *In particular* $L^2(\mathscr{G})$ *is a closed linear subspace of* $L^2(\mathscr{A})$.

Proof Since $\mathscr{G} \subset \mathscr{A}$, each $u \in L^2(\mathscr{G})$ is also $\mathscr{A}$ measurable. Moreover, by Theorem 14.13

$$\int u^2 \, d\mu = \int_{(0,\infty)} \mu\big(\underbrace{\{u^2 \geqslant t\}}_{\in \mathscr{G}}\big)\lambda^1(dt) = \int_{(0,\infty)} \mu|_{\mathscr{G}}\big(\{u^2 \geqslant t\}\big)\lambda^1(dt) = \int u^2 \, d\mu|_{\mathscr{G}},$$

which shows that $\|\jmath(u)\|_{L^2(\mathscr{A})} = \|u\|_{L^2(\mathscr{G})}$. In particular, $u \in L^2(\mathscr{A})$. □

It is customary to identify $u \in L^2(\mathscr{G})$ with $\jmath(u) \in L^2(\mathscr{A})$ and we will do so in that which follows. Unless we want to stress the σ-algebra, we will write μ instead of $\mu|_{\mathscr{G}}$ and $\|\cdot\|_2$ for the norm in $L^2(\mathscr{G}, \mu|_{\mathscr{G}})$ and $L^2(\mathscr{A}, \mu)$.

A key observation is that the choice of $\mathscr{G} \subset \mathscr{A}$ determines our knowledge about a function u.

Example 27.2 Consider a finite measure space $(X, \mathscr{A}, \mu)$ and the sub-σ-algebra $\mathscr{G} := \{\emptyset, G, G^c, X\}$, where $G \in \mathscr{A}$ and $\mu(G) > 0$, $\mu(G^c) > 0$. Let $f \in \mathcal{E}(\mathscr{A})$ be a simple function in standard representation:

$$f = \sum_{n=0}^{N} y_n \mathbb{1}_{A_n}, \quad y_n \in \mathbb{R},\ A_n \in \mathscr{A}.$$

We have

$$\int_G f \, d\mu = \sum_{n=0}^{N} y_n \int_G \mathbb{1}_{A_n} \, d\mu = \underbrace{\sum_{n=0}^{N} y_n \frac{\mu(A_n \cap G)}{\mu(G)}}_{=:\,\eta_1} \mu(G) = \eta_1 \int \mathbb{1}_G \, d\mu, \tag{27.1}$$

and similarly,

$$\int_{G^c} f \, d\mu = \underbrace{\sum_{n=0}^{N} y_n \frac{\mu(A_n \cap G^c)}{\mu(G^c)}}_{=:\,\eta_2} \mu(G^c) = \eta_2 \int \mathbb{1}_{G^c} \, d\mu. \tag{27.2}$$

This indicates that we could have obtained the same results in the integrations (27.1) and (27.2) if we had used not $f \in \mathcal{E}(\mathscr{A})$ but the $\mathscr{G}$-simple function

$$g := \eta_1 \mathbb{1}_G + \eta_2 \mathbb{1}_{G^c} \in \mathcal{E}(\mathscr{G}) \tag{27.3}$$

with η_1, η_2 from above. In other words, f and g are indistinguishable, if we evaluate (i.e. integrate) both of them on sets of the σ-algebra $\mathscr{G}$. Note that g is much simpler than f, but we have lost nearly all of the information regarding what f looks like on sets from $\mathscr{A}$ save $\mathscr{G}$: if we take a set from the standard representation of f, say, $A_{n_0} \subsetneq G$, $A_{n_0} \in \mathscr{A}$, then

$$\int_{A_{n_0}} f\,d\mu = y_{n_0}\mu(A_{n_0}) \neq \int_{A_{n_0}} g\,d\mu = \eta_1 \mu(A_{n_0} \cap G)$$

$$= \underbrace{\left(\sum_{n=0}^{N} y_n \frac{\mu(A_n \cap G)}{\mu(G)}\right)}_{=\eta_1} \cdot \mu(A_{n_0}),$$

i.e. we would get a weighted *average* of the y_n rather than precisely y_{n_0}.

Let us extend the process sketched in Example 27.2 to σ-finite measures and general square-integrable functions. Our starting point is the observation that, with the notation introduced in Example 27.2,

$$\int fg\,d\mu = \eta_1 \int_G f\,d\mu + \eta_2 \int_{G^c} f\,d\mu = \eta_1^2\mu(G) + \eta_2^2\mu(G^c) = \int g^2\,d\mu,$$

that is, $\langle f-g, g\rangle = 0$ or $(f-g) \perp g$ in the space $L^2(\mathscr{A})$.

Definition 27.3 Let $(X, \mathscr{A}, \mu)$ be a measure space and $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra. The *conditional expectation* of $u \in L^2(\mathscr{A})$ relative to $\mathscr{G}$ is the orthogonal projection $\mathbb{E}^{\mathscr{G}} : L^2(\mathscr{A}) \to L^2(\mathscr{G})$,[2] onto the closed subspace $L^2(\mathscr{G})$.

The terminology 'conditional expectation' comes from probability theory, where this notion is widely used and where $(X, \mathscr{A}, \mu)$ is usually a probability space. In an abuse of language we continue to call $\mathbb{E}^{\mathscr{G}}$ the conditional *expectation* even if μ is not a probability measure. Let us collect some properties of $\mathbb{E}^{\mathscr{G}}$.

Theorem 27.4 *Let $(X, \mathscr{A}, \mu)$ be a measure space and $\mathscr{G} \subset \mathscr{A}$ a sub-σ-algebra. The conditional expectation $\mathbb{E}^{\mathscr{G}}$ has the following properties* ($u, w \in L^2(\mathscr{A})$)*:*

(i) $\mathbb{E}^{\mathscr{G}} u \in L^2(\mathscr{G})$;
(ii) $\|\mathbb{E}^{\mathscr{G}} u\|_{L^2(\mathscr{G})} \leqslant \|u\|_{L^2(\mathscr{A})}$;
(iii) $\langle \mathbb{E}^{\mathscr{G}} u, w\rangle = \langle u, \mathbb{E}^{\mathscr{G}} w\rangle = \langle \mathbb{E}^{\mathscr{G}} u, \mathbb{E}^{\mathscr{G}} w\rangle$;
(iv) $\mathbb{E}^{\mathscr{G}} u$ *is the unique minimizer in* $L^2(\mathscr{G})$ *such that*

$$\|u - \mathbb{E}^{\mathscr{G}} u\|_{L^2(\mathscr{A})} = \inf_{g \in L^2(\mathscr{G})} \|u - g\|_{L^2(\mathscr{A})};$$

(v) $u = w \implies \mathbb{E}^{\mathscr{G}} u = \mathbb{E}^{\mathscr{G}} w$;
(vi) $\mathbb{E}^{\mathscr{G}}(\alpha u + \beta w) = \alpha \mathbb{E}^{\mathscr{G}} u + \beta \mathbb{E}^{\mathscr{G}} w$ *for all* $\alpha, \beta \in \mathbb{R}$;

[2] In probability theory the notation $\mathbb{E}(u \mid \mathscr{G})$ is more common than $\mathbb{E}^{\mathscr{G}} u$.

(vii) *if* $\mathscr{H} \subset \mathscr{G}$ *is a further sub-σ-algebra, then* $\mathbb{E}^{\mathscr{H}}\mathbb{E}^{\mathscr{G}}u = \mathbb{E}^{\mathscr{H}}u$*;*

(viii) $\mathbb{E}^{\mathscr{G}}(gu) = g\mathbb{E}^{\mathscr{G}}u$ *for all* $g \in L^\infty(\mathscr{G})$*;*

(ix) $\mathbb{E}^{\mathscr{G}}g = g$ *for all* $g \in L^2(\mathscr{G})$*;*

(x) $0 \leqslant u \leqslant 1 \implies 0 \leqslant \mathbb{E}^{\mathscr{G}}u \leqslant 1$*;*

(xi) $u \leqslant w \implies \mathbb{E}^{\mathscr{G}}u \leqslant \mathbb{E}^{\mathscr{G}}w$*;*

(xii) $|\mathbb{E}^{\mathscr{G}}u| \leqslant \mathbb{E}^{\mathscr{G}}|u|$*;*

(xiii) $\mathbb{E}^{\{\emptyset, X\}}u = \dfrac{1}{\mu(x)}\displaystyle\int u\,d\mu$ *for all* $u \in L^1(\mathscr{A}) \cap L^2(\mathscr{A})$ $\qquad \left(\dfrac{1}{\infty} := 0\right).$

Before we turn to the proof of the above properties let us stress again that all (in)equalities in (i)–(xiii) are between L^2-functions, i.e. they hold only μ-almost everywhere. In particular, $\mathbb{E}^{\mathscr{G}}u$ is itself determined only up to a μ-null set $N \in \mathscr{G}$.

Proof of Theorem 27.4 Properties (i)–(vi) and (ix) follow directly from Theorem 26.13, Corollary 26.14 and Remark 26.15.

(vii) For all $u, w \in L^2(\mathscr{A})$ we find because of (iii)

$$\langle \mathbb{E}^{\mathscr{H}}\mathbb{E}^{\mathscr{G}}u, w\rangle \overset{\text{(iii)}}{=} \langle \mathbb{E}^{\mathscr{G}}u, \mathbb{E}^{\mathscr{H}}w\rangle = \langle u, \mathbb{E}^{\mathscr{G}}\underbrace{\mathbb{E}^{\mathscr{H}}w}_{\in L^2(\mathscr{G})}\rangle \overset{\text{(ix)}}{=} \langle u, \mathbb{E}^{\mathscr{H}}w\rangle$$
$$= \langle \mathbb{E}^{\mathscr{H}}u, w\rangle$$

as $\mathbb{E}^{\mathscr{H}}w \in L^2(\mathscr{H}) \subset L^2(\mathscr{G})$. Since $w \in L^2(\mathscr{A})$ is arbitrary, we conclude that $\mathbb{E}^{\mathscr{H}}\mathbb{E}^{\mathscr{G}}u = \mathbb{E}^{\mathscr{H}}u$.

(viii) On writing $u = f + f^\perp \in L^2(\mathscr{G}) \oplus L^2(\mathscr{G})^\perp$, we get $gu = gf + gf^\perp$. Moreover, we have for any $\phi \in L^2(\mathscr{G})$ and $g \in L^\infty(\mathscr{G})$ that $g\phi \in L^2(\mathscr{G})$, thus

$$\langle gf^\perp, \phi\rangle = \langle f^\perp, g\phi\rangle = 0,$$

and from the uniqueness of the orthogonal decomposition we infer that

$$(gf)^\perp = gf^\perp \quad \text{or} \quad \mathbb{E}^{\mathscr{G}}(gu) = gf = g\mathbb{E}^{\mathscr{G}}u.$$

(x) Let $0 \leqslant u \leqslant 1$. Since $\mathbb{E}^{\mathscr{G}}u \in L^2(\mathscr{G})$, the Markov inequality (11.4) implies

$$\mu\left\{|\mathbb{E}^{\mathscr{G}}u| > \frac{1}{n}\right\} \leqslant n^2\|\mathbb{E}^{\mathscr{G}}u\|^2_{L^2(\mathscr{G})} \leqslant n^2\|u\|^2_{L^2(\mathscr{A})} < \infty, \tag{27.4}$$

and so $\phi_n := \mathbb{1}_{\{|\mathbb{E}^{\mathscr{G}}u| > 1/n\}} \in L^2(\mathscr{G})$. Therefore,

$$\int \mathbb{E}^{\mathscr{G}}u\mathbb{1}_{\{\mathbb{E}^{\mathscr{G}}u<0\}}\phi_n\,d\mu \overset{\text{(iii)}}{=} \int u\mathbb{E}^{\mathscr{G}}(\mathbb{1}_{\{\mathbb{E}^{\mathscr{G}}u<0\}}\phi_n)d\mu$$
$$\overset{\text{(ix)}}{=} \int u\mathbb{1}_{\{\mathbb{E}^{\mathscr{G}}u<0\}}\phi_n\,d\mu \geqslant 0,$$

which is possible only if $\mu\big(\{\mathbb{E}^{\mathscr{G}} u < 0\} \cap \{|\mathbb{E}^{\mathscr{G}} u| > 1/n\}\big) = 0$, that is, if

$$\begin{aligned}\mu\{\mathbb{E}^{\mathscr{G}} u < 0\} &= \mu\Big(\bigcup_{n\in\mathbb{N}} \{\mathbb{E}^{\mathscr{G}} u < 0\} \cap \{|\mathbb{E}^{\mathscr{G}} u| > 1/n\}\Big) \\ &= \sup_{n\in\mathbb{N}} \mu\big(\{\mathbb{E}^{\mathscr{G}} u < 0\} \cap \{|\mathbb{E}^{\mathscr{G}} u| > 1/n\}\big) \;=\; 0,\end{aligned}$$

hence $\mathbb{E}^{\mathscr{G}} u \geqslant 0$.

With very similar arguments we see that $\mathbb{1}_{\{\mathbb{E}^{\mathscr{G}} u > 1\}} \in L^2(\mathscr{G})$, and since $u \leqslant 1$ we have

$$\int \mathbb{E}^{\mathscr{G}} u \mathbb{1}_{\{\mathbb{E}^{\mathscr{G}} u > 1\}}\, d\mu \overset{\text{(iii),(ix)}}{=} \int u \mathbb{1}_{\{\mathbb{E}^{\mathscr{G}} u > 1\}}\, d\mu \leqslant \mu\{\mathbb{E}^{\mathscr{G}} u > 1\},$$

which entails $\mu\{\mathbb{E}^{\mathscr{G}} u > 1\} = 0$ or $\mathbb{E}^{\mathscr{G}} u \leqslant 1$ [✍], see also Problem 27.2.

(xi) Using that $w - u \geqslant 0$, the first part of the proof of (x) shows $\mathbb{E}^{\mathscr{G}}(w - u) \geqslant 0$, so that by linearity $\mathbb{E}^{\mathscr{G}} w \geqslant \mathbb{E}^{\mathscr{G}} u$.

(xii) Again by the proof of (x) we find for $|u| \pm u \geqslant 0$ that $\mathbb{E}^{\mathscr{G}}(|u| \pm u) \geqslant 0$, and by linearity $\pm\mathbb{E}^{\mathscr{G}} u \leqslant \mathbb{E}^{\mathscr{G}} |u|$. This proves $\big|\mathbb{E}^{\mathscr{G}} u\big| \leqslant \mathbb{E}^{\mathscr{G}} |u|$.

(xiii) If $\mu(X) = \infty$, we have $L^2(\{\emptyset, X\}) = \{0\}$, [✍] thus $\mathbb{E}^{\{\emptyset, X\}} u = 0$ and the formula clearly holds.

If $\mu(X) < \infty$, we have $L^2(\{\emptyset, X\}) \simeq \mathbb{R}$, and $\mathbb{E}^{\{\emptyset, X\}} u = c$ is a constant. By (iv), $c = \mathbb{E}^{\{\emptyset, X\}} u$ minimizes $\|u - c\|_{L^2(\mathscr{A})}$, and

$$\begin{aligned}\int |u - c|^2\, d\mu &= \int u^2\, d\mu - 2c \int u\, d\mu + c^2 \int d\mu \\ &= \int u^2\, d\mu - \frac{1}{\mu(x)}\Big(\int u\, d\mu\Big)^2 + \frac{1}{\mu(x)}\Big(\int u\, d\mu - c\mu(X)\Big)^2\end{aligned}$$

shows that

$$c = \frac{1}{\mu(x)} \int u\, d\mu$$

is the unique minimizer. □

Extension from L^2 to L^p

We want to extend the operator $\mathbb{E}^{\mathscr{G}}$ to $\bigcup_{p\geqslant 1} L^p(\mathscr{A})$. To do so, we use a method which is called *extension by continuity*, see Problem 27.5. Assume that we have a linear operator $T\colon L^2(\mu) \to L^2(\mu)$ such that for some $p \in [1, \infty]$, $p \neq 2$,

$$\|Tu\|_p \leqslant c\|u\|_p \quad \forall u \in L^2(\mu) \cap L^p(\mu)$$

holds. Since $L^2(\mu)\cap L^p(\mu)$ is a dense subset of $L^p(\mu)$ – approximate $u\in L^p(\mu)$ with $u_n := (-n)\vee u\wedge n\cdot \mathbb{1}_{\{|u|>1/n\}}\in L^2(\mu)\cap L^p(\mu)$ [✍] using, say, dominated convergence (Theorem 13.9) – we see from

$$\|Tu_n - Tu_k\|_p \overset{\text{linear}}{=} \|T(u_n-u_k)\|_p \leqslant c\|u_n-u_k\|_p \xrightarrow[k,n\to\infty]{} 0$$

that $(Tu_n)_{n\in\mathbb{N}}$ is an $L^p(\mu)$-Cauchy sequence. Because of the completeness of the space $L^p(\mu)$, the limit

$$\tilde{T}u := \lim_{n\to\infty} Tu_n$$

exists. Since it does not depend on the choice of the approximating sequence u_n, [✍] $\tilde{T}: L^p(\mu)\to L^p(\mu)$ is a linear operator which coincides $L^2(\mu)\cap L^p(\mu)$ with T. Note that, by construction, $Tu_{n_k}\to \tilde{T}u$ μ-a.e. for some subsequence. If for $u_n\uparrow u$ the sequence $(Tu_n)_{n\in\mathbb{N}}$ is also increasing – this is the case for conditional expectations, see Theorem 27.4(xi) – then we even get $Tu_n\uparrow\tilde{T}u$ μ-a.e. for the *full* sequence. [✍]

Theorem 27.5 *Let $(X,\mathscr{A},\mu)$ be a σ-finite measure space. The conditional expectation $\mathbb{E}^{\mathscr{G}}$ has an extension mapping $L^p(\mathscr{A})$ to $L^p(\mathscr{G})$ for any $p\in[1,\infty]$.*

Proof Using the method of extension by continuity, it is enough to check that $\|\mathbb{E}^{\mathscr{G}}u\|_p\leqslant c\|u\|_p$ for all p and $u\in L^2(\mathscr{A})\cap L^p(\mathscr{A})$.

Step 1. $p=\infty$. If $u\in L^2(\mathscr{A})\cap L^\infty(\mathscr{A})$ we get from Theorem 27.4 and the observation that $|u|/\|u\|_\infty\leqslant 1$

$$\left|\mathbb{E}^{\mathscr{G}}(u/\|u\|_\infty)\right|\leqslant \mathbb{E}^{\mathscr{G}}(|u|/\|u\|_\infty)\leqslant 1,$$

thus $\|\mathbb{E}^{\mathscr{G}}u\|_\infty\leqslant\|u\|_\infty$ for all $u\in L^2(\mathscr{A})\cap L^\infty(\mathscr{A})$.

Step 2. $1\leqslant p<\infty$. Fix $u\in L^p(\mathscr{A})\cap L^2(\mathscr{A})$. Since $\mathbb{E}^{\mathscr{G}}u\in L^2(\mathscr{G})$, the Markov inequality (11.4) shows that $\{|\mathbb{E}^{\mathscr{G}}u|=\infty\}\in\mathscr{G}$ is a μ-null set and that the sets $G_n:=\{n>|\mathbb{E}^{\mathscr{G}}u|>1/n\}\in\mathscr{G}$ have finite μ-measure,

$$\mu(G_n)\leqslant\mu\{|\mathbb{E}^{\mathscr{G}}u|>1/n\}\leqslant n^2\int(\mathbb{E}^{\mathscr{G}}u)^2\,d\mu<\infty.$$

Moreover,

$$\begin{aligned}
\left\|\mathbb{E}^{\mathscr{G}}u\mathbb{1}_{G_n}\right\|_p^p &= \left\langle|\mathbb{E}^{\mathscr{G}}u|,|\mathbb{E}^{\mathscr{G}}u|^{p-1}\mathbb{1}_{G_n}\right\rangle\\
&\overset{27.4\text{(xii)}}{\leqslant} \left\langle\mathbb{E}^{\mathscr{G}}|u|,|\mathbb{E}^{\mathscr{G}}u|^{p-1}\mathbb{1}_{G_n}\right\rangle\\
&\overset{27.4\text{(iii),(ix)}}{=} \left\langle|u|,|\mathbb{E}^{\mathscr{G}}u|^{p-1}\mathbb{1}_{G_n}\right\rangle\\
&\leqslant C_q\|u\|_p,
\end{aligned}$$

where we use Hölder's inequality (Theorem 13.2) with $p^{-1}+q^{-1}=1$ and

$$C_q=\left(\int |\mathbb{E}^{\mathscr{G}}u|^{(p-1)q}\mathbb{1}_{G_n}\,d\mu\right)^{1/q}=\left(\int |\mathbb{E}^{\mathscr{G}}u|^{p}\mathbb{1}_{G_n}\,d\mu\right)^{1/q}=\left\|\mathbb{E}^{\mathscr{G}}u\mathbb{1}_{G_n}\right\|_p^{p-1},$$

resp. $C_\infty=1$. Dividing the above inequality by C_q – if $C_q=0$ for all large n there is nothing to show since in this instance $\mathbb{E}^{\mathscr{G}}u=0$ [✍] – gives

$$\left\|\mathbb{E}^{\mathscr{G}}u\mathbb{1}_{G_n}\right\|_p\leqslant\|u\|_p.$$

As we have seen above, $\mu\{\mathbb{E}^{\mathscr{G}}u=\infty\}=0$, so we can use Beppo Levi's theorem (Theorem 9.6) to find for all $u\in L^2(\mathscr{A})\cap L^p(\mathscr{A})$

$$\left\|\mathbb{E}^{\mathscr{G}}u\right\|_p=\left\|\mathbb{E}^{\mathscr{G}}u\mathbb{1}_{\{0<|\mathbb{E}^{\mathscr{G}}u|<\infty\}}\right\|_p=\sup_{n\in\mathbb{N}}\left\|\mathbb{E}^{\mathscr{G}}u\mathbb{1}_{G_n}\right\|_p\leqslant\|u\|_p. \qquad \square$$

Monotone Extensions

It is possible to extend $\mathbb{E}^{\mathscr{G}}$ beyond $\bigcup_{p\geqslant 1}L^p(\mathscr{A})$. For this we use a Daniell-type approach. We need a few preparations.

Definition 27.6 Let $(X,\mathscr{A},\mu)$ be a measure space. Two functions $u,v\in\mathcal{M}(\mathscr{A})$ are called *equivalent*, $u\sim v$, if $\{u\neq v\}\in\mathscr{N}_\mu$ is a μ-null set. We then define $M(\mathscr{A}):=\mathcal{M}(\mathscr{A})/_\sim$ for the set of all equivalence classes of measurable functions $u\in\mathcal{M}(\mathscr{A})$.

As with L^p-functions, all (in)equalities between elements from $M(\mathscr{A})$ hold for *almost every* point $x\in X$.

Lemma 27.7 *Let $(X,\mathscr{A},\mu)$ be a σ-finite measure space. Then $u\in M^+(\mathscr{A})$ if, and only if, there exists an increasing sequence $(u_n)_{n\in\mathbb{N}}\subset L^2_+(\mathscr{A})$ such that $u_n\uparrow u$.*

Proof The 'only if' part '$\Leftarrow$' is trivial since suprema of countably many measurable functions are again measurable (Corollary 8.10). For '$\Rightarrow$' let $(A_n)_{n\in\mathbb{N}}\subset\mathscr{A}$ be an exhausting sequence such that $A_n\uparrow X$ and $\mu(A_n)<\infty$. If $u\in M^+(\mathscr{A})$, then $u_n:=(u\wedge n)\mathbb{1}_{A_n}\in L^2_+(\mathscr{A})$ and $\sup_{n\in\mathbb{N}}u_n=u$. $\square$

Caution Lemma 27.7 may fail if $(X,\mathscr{A},\mu)$ is not σ-finite. In fact, if $u\equiv 1$ can be approximated by an increasing sequence $(u_n)_{n\in\mathbb{N}}\subset L^2_+(\mathscr{A})$, $1=\sup_{n\in\mathbb{N}}u_n$, then the measure space is σ-finite. Indeed, the sets $A_n:=\{u_n>1/n\}$ are an increasing sequence $A_n\uparrow X$ with $\mu(A_n)=\mu\{u_n>1/n\}\leqslant n^2\int u_n^2\,d\mu<\infty$ by the Markov inequality (11.4).

The key technical point is the following result.

Lemma 27.8 *Let $(X,\mathscr{A},\mu)$ be a measure space, $\mathscr{G}\subset\mathscr{A}$ be a sub-σ-algebra, and $(u_n)_{n\in\mathbb{N}},(w_n)_{n\in\mathbb{N}}\subset L^2(\mathscr{A})$ be two increasing sequences. Then*

$$\sup_{n\in\mathbb{N}} u_n=\sup_{n\in\mathbb{N}} w_n \implies \sup_{n\in\mathbb{N}}\mathbb{E}^{\mathscr{G}}u_n=\sup_{n\in\mathbb{N}}\mathbb{E}^{\mathscr{G}}w_n. \tag{27.5}$$

If $u:=\sup_{n\in\mathbb{N}}u_n$ is in $L^2(\mathscr{A})$, the conditional monotone convergence *property*

$$\sup_{n\in\mathbb{N}}\mathbb{E}^{\mathscr{G}}u_n=\mathbb{E}^{\mathscr{G}}\Big(\sup_{n\in\mathbb{N}}u_n\Big)=\mathbb{E}^{\mathscr{G}}u \tag{27.6}$$

holds in $L^2(\mathscr{G})$ and almost everywhere.

Proof Let us first of all assume that $u_n\uparrow u$ and $u\in L^2(\mathscr{A})$. Monotone convergence (Theorem 12.1) and Theorem 13.7 show that $u_n\to u$ also in L^2-sense. By Theorem 27.4(xi), $(\mathbb{E}^{\mathscr{G}}u_n)_{n\in\mathbb{N}}$ is again an increasing sequence and $\mathbb{E}^{\mathscr{G}}u_n\leqslant\mathbb{E}^{\mathscr{G}}u$. From 27.4(ii) and (vi) we get

$$\Big\|\mathbb{E}^{\mathscr{G}}u-\mathbb{E}^{\mathscr{G}}u_n\Big\|_2=\Big\|\mathbb{E}^{\mathscr{G}}(u-u_n)\Big\|_2\leqslant\|u-u_n\|_2,$$

i.e. $L^2\text{-}\lim_{n\to\infty}\mathbb{E}^{\mathscr{G}}u_n=\mathbb{E}^{\mathscr{G}}u$. For a subsequence $(u_{n(k)})_{k\in\mathbb{N}}\subset(u_n)_{n\in\mathbb{N}}$ we even have $\lim_{k\to\infty}\mathbb{E}^{\mathscr{G}}u_{n(k)}=\mathbb{E}^{\mathscr{G}}u$ a.e., see Corollary 13.8. Because of the monotonicity of the sequence $(\mathbb{E}^{\mathscr{G}}u_n)_{n\in\mathbb{N}}$, we get for all $n>n(k)$

$$\Big|\mathbb{E}^{\mathscr{G}}u-\mathbb{E}^{\mathscr{G}}u_n\Big|=\mathbb{E}^{\mathscr{G}}u-\mathbb{E}^{\mathscr{G}}u_n\leqslant\mathbb{E}^{\mathscr{G}}u-\mathbb{E}^{\mathscr{G}}u_{n(k)},$$

and letting first $n\to\infty$ and then $k\to\infty$ gives $\mathbb{E}^{\mathscr{G}}u_n\uparrow\mathbb{E}^{\mathscr{G}}u$ a.e. This finishes the proof of (27.6).

If $(u_n)_{n\in\mathbb{N}},(w_n)_{n\in\mathbb{N}}\subset L^2(\mathscr{A})$ are any two increasing sequences[3] with the same limit $\sup_{n\in\mathbb{N}}u_n=\sup_{n\in\mathbb{N}}w_n$, we can apply (27.6) to the equally increasing sequences $u_n\wedge w_k\uparrow u_n$ (as $k\to\infty$ and for fixed n) and $u_n\wedge w_k\uparrow w_k$ (as $n\to\infty$ and for fixed k). This shows

$$\sup_{n\in\mathbb{N}}\mathbb{E}^{\mathscr{G}}u_n\overset{(27.6)}{=}\sup_{n\in\mathbb{N}}\sup_{k\in\mathbb{N}}\mathbb{E}^{\mathscr{G}}(u_n\wedge w_k)=\sup_{k\in\mathbb{N}}\sup_{n\in\mathbb{N}}\mathbb{E}^{\mathscr{G}}(u_n\wedge w_k)\overset{(27.6)}{=}\sup_{k\in\mathbb{N}}\mathbb{E}^{\mathscr{G}}w_k. \quad\Box$$

A combination of Lemmata 27.7 and 27.8 allows us to define conditional expectations for positive measurable functions in a σ-finite measure space.

Definition 27.9 Let $(X,\mathscr{A},\mu)$ be a σ-finite measure space and $\mathscr{G}\subset\mathscr{A}$ be a sub-σ-algebra. Let $u\in M^+(\mathscr{A})$ and let $(u_n)_{n\in\mathbb{N}}\subset L^2_+(\mathscr{A})$ be an increasing sequence such that $u=\sup_{n\in\mathbb{N}}u_n$. Then

$$E^{\mathscr{G}}u:=\sup_{n\in\mathbb{N}}\mathbb{E}^{\mathscr{G}}u_n \tag{27.7}$$

[3] We do not assume that $\sup_{n\in\mathbb{N}}u_n,\sup_{n\in\mathbb{N}}w_n\in L^2(\mathscr{A})$.

is called the *conditional expectation* of u with respect to $\mathscr{G}$. If $u \in M(\mathscr{A})$ and $E^{\mathscr{G}} u^{\pm} \in \mathbb{R}$ almost everywhere, we define (almost everywhere)

$$E^{\mathscr{G}} u := E^{\mathscr{G}} u^{+} - E^{\mathscr{G}} u^{-} = \lim_{n\to\infty} \left(\mathbb{E}^{\mathscr{G}} u_n^{+} - \mathbb{E}^{\mathscr{G}} u_n^{-}\right), \tag{27.8}$$

where $u_n^{\pm} \uparrow u^{\pm}$ are suitable approximating sequences from $L^2_{+}(\mathscr{A})$.

We write $L^{\mathscr{G}}(\mathscr{A})$ for the set of all functions $u \in M(\mathscr{A})$ such that (almost everywhere) $E^{\mathscr{G}} u$ exists and is finite.

Theorem 27.10 *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space. The conditional expectation $E^{\mathscr{G}}$ extends $\mathbb{E}^{\mathscr{G}}$, i.e. $L^p(\mathscr{A}) \subset L^{\mathscr{G}}(\mathscr{A})$ and $E^{\mathscr{G}} u = \mathbb{E}^{\mathscr{G}} u$ for all $u \in L^p(\mathscr{A})$.*

Proof Applying (27.6) to u^+ and u^- shows that $E^{\mathscr{G}} u^{\pm} = \mathbb{E}^{\mathscr{G}} u^{\pm}$ and, in particular, $E^{\mathscr{G}} u^{\pm} \in L^2(\mathscr{G})$. As such, $E^{\mathscr{G}} u^{\pm}$ is a.e. real-valued, so that (27.8) is always defined in $M(\mathscr{A})$, resp. $M(\mathscr{G})$. The same argument holds for $p \neq 2$ with some obvious modifications in the proof of (27.6). $\square$

Properties of Conditional Expectations

In this section we prove the analogue of Theorem 27.4 for the extension(s) of $\mathbb{E}^{\mathscr{G}}$. From now on we will no longer distinguish between $\mathbb{E}^{\mathscr{G}}$ and its extension $E^{\mathscr{G}}$, and we always write $\mathbb{E}^{\mathscr{G}}$. The following theorem is formulated for functions from $L^{\mathscr{G}}(\mathscr{A})$, i.e. the (maximal) set of $\mathscr{A}$-measurable functions for which the conditional expectation is defined, see e.g. Definition 27.9; if you have not read the subsection on 'monotone extensions', you may safely assume that $L^{\mathscr{G}}(\mathscr{A}) = \bigcup_{1\leqslant p\leqslant\infty} L^p(\mathscr{A})$.

Theorem 27.11 *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and let $\mathscr{H} \subset \mathscr{G} \subset \mathscr{A}$ be sub-σ-algebras. The conditional expectation $\mathbb{E}^{\mathscr{G}}$ has the following properties ($u, w \in L^{\mathscr{G}}(\mathscr{A}), p \in [1,\infty], \alpha, \beta \in \mathbb{R}$):*[4]

(i) *$\mathbb{E}^{\mathscr{G}} u$ is $\mathscr{G}$ measurable;*

(ii) *$u \in L^p(\mathscr{A}) \implies \mathbb{E}^{\mathscr{G}} u \in L^p(\mathscr{G})$ and $\|\mathbb{E}^{\mathscr{G}} u\|_p \leqslant \|u\|_p$;*

(iii) *$\langle \mathbb{E}^{\mathscr{G}} u, w\rangle = \langle u, \mathbb{E}^{\mathscr{G}} w\rangle = \langle \mathbb{E}^{\mathscr{G}} u, \mathbb{E}^{\mathscr{G}} w\rangle$*
(for $u, w \in L^{\mathscr{G}}(\mathscr{A})$, $u\mathbb{E}^{\mathscr{G}} w \in L^1(\mathscr{A})$, e.g. if $u \in L^p(\mathscr{A})$ and $w \in L^q(\mathscr{A})$ with $p^{-1} + q^{-1} = 1$);

(iv) *$u = w \implies \mathbb{E}^{\mathscr{G}} u = \mathbb{E}^{\mathscr{G}} w$;*

(v) *$\mathbb{E}^{\mathscr{G}}(\alpha u + \beta w) = \alpha\mathbb{E}^{\mathscr{G}} u + \beta\mathbb{E}^{\mathscr{G}} w$;*

[4] For parts (vii)–(viii′), if we work solely with the L^p-extension we have to assume that both u and gu are in $\bigcup_{1\leqslant p\leqslant\infty} L^p(\mu)$, e.g. $g \in L^\infty(\mathscr{G})$ as in Theorem 27.4(ix). If we use the monotone extension of $\mathbb{E}^{\mathscr{G}}$, the property $gu \in L^{\mathscr{G}}(\mathscr{A})$ holds for any $\mathscr{G}$ measurable g.

(vi) $\mathscr{G} \supset \mathscr{H} \implies \mathbb{E}^{\mathscr{H}}\mathbb{E}^{\mathscr{G}}u = \mathbb{E}^{\mathscr{H}}u$;
(vii) $\mathbb{E}^{\mathscr{G}}(gu) = g\mathbb{E}^{\mathscr{G}}u$ *for all* $\mathscr{G}$ *measurable* g *and* $u \in L^{\mathscr{G}}(\mathscr{A})$;
(viii) $\mathbb{E}^{\mathscr{G}}g = g\mathbb{E}^{\mathscr{G}}1$ *for all* $\mathscr{G}$*-measurable* g;
(viii′) *if* $\mu|_{\mathscr{G}}$ *is* σ*-finite* $\quad 1 = \mathbb{E}^{\mathscr{G}}1$ *and* $g = \mathbb{E}^{\mathscr{G}}g$ *for all* $\mathscr{G}$*-measurable* g;
(ix) $0 \leqslant u \leqslant 1 \implies 0 \leqslant \mathbb{E}^{\mathscr{G}}u \leqslant 1$;
(x) $u \leqslant w \implies \mathbb{E}^{\mathscr{G}}u \leqslant \mathbb{E}^{\mathscr{G}}w$;
(xi) $|\mathbb{E}^{\mathscr{G}}u| \leqslant \mathbb{E}^{\mathscr{G}}|u|$;
(xii) $\mathbb{E}^{\{\emptyset,X\}}u = \dfrac{1}{\mu(x)}\displaystyle\int u\,d\mu$ *for all* $u \in L^1(\mathscr{A})$ $\qquad \left(\dfrac{1}{\infty} := 0\right)$.

Proof Note that the extension of $\mathbb{E}^{\mathscr{G}}u$ for $u \in L^{\mathscr{G}}(\mathscr{A})$ or $u \in L^p(\mathscr{A})$ can be obtained as an a.e. limit of a suitable sequence $(u_n)_{n\in\mathbb{N}} \subset L^2(\mathscr{A})$, see Definition 27.9 (for $L^{\mathscr{G}}(\mathscr{A})$) and the remark preceding the statement of Theorem 27.5 (for $L^p(\mathscr{A})$). This means that we can use Theorem 27.4 and (pointwise) approximation arguments.

Parts (i) and (v) are clear from the fact that $\mathbb{E}^{\mathscr{G}}u$ is a pointwise limit of $\mathscr{G}$ measurable functions. Part (ii) was proved in Theorem 27.5. Parts (iii), (iv), (ix) and (x) follow by approximation from the corresponding properties of $\mathbb{E}^{\mathscr{G}}$ from Theorem 27.4, and (xi) is derived from (ix) exactly as in the L^2-case.

(vi) Without loss of generality it suffices to consider the case $u \geqslant 0$. Pick a sequence $(u_n)_{n\in\mathbb{N}} \subset L^2(\mathscr{A})$ such that $u_n \uparrow u$. Because of the construction of $\mathbb{E}^{\mathscr{G}}$ and the fact that $\mathbb{E}^{\mathscr{G}}$ is monotone on $L^2(\mathscr{A})$ (Property (xi) of Theorem 27.4), we know that $\mathbb{E}^{\mathscr{G}}u_n \uparrow \mathbb{E}^{\mathscr{G}}u$ as well as $\mathbb{E}^{\mathscr{H}}u_n \uparrow \mathbb{E}^{\mathscr{H}}u$ and $\mathbb{E}^{\mathscr{H}}(\mathbb{E}^{\mathscr{G}}u_n) \uparrow \mathbb{E}^{\mathscr{H}}(\mathbb{E}^{\mathscr{G}}u)$ for a suitable sequence $(u_n)_{n\in\mathbb{N}} \subset L^2_+(\mathscr{A})$. Since $\mathbb{E}^{\mathscr{H}}(\mathbb{E}^{\mathscr{G}}u_n) = \mathbb{E}^{\mathscr{H}}u_n$, by Theorem 27.4(vii), we are done.

(vii) Assume that $g, u \geqslant 0$. Define $g_n := g \wedge n \in L^\infty_+(\mathscr{G})$ and let $(u_n)_{n\in\mathbb{N}} \subset L^2_+(\mathscr{A})$ be an increasing sequence such that $\sup_{n\in\mathbb{N}} u_n = u$. Then $g_n u_n \in L^2(\mathscr{A})$, $g_n u_n \uparrow gu$ and Theorem 27.4 shows that $\mathbb{E}^{\mathscr{G}}(g_n u_n) = g_n\mathbb{E}^{\mathscr{G}}u_n$. Hence,

$$\mathbb{E}^{\mathscr{G}}(gu) = \sup_{n\in\mathbb{N}} \mathbb{E}^{\mathscr{G}}(g_n u_n) = \sup_{n\in\mathbb{N}} \big(g_n \mathbb{E}^{\mathscr{G}}u_n\big) = g\mathbb{E}^{\mathscr{G}}u.$$

If $g \geqslant 0$ and $u \in L^{\mathscr{G}}(\mathscr{A})$, the conditional expectation $\mathbb{E}^{\mathscr{G}}u^+ - \mathbb{E}^{\mathscr{G}}u^-$ is well-defined and we find from the previous calculations

$$g\mathbb{E}^{\mathscr{G}}u = g\big(\mathbb{E}^{\mathscr{G}}u^+ - \mathbb{E}^{\mathscr{G}}u^-\big) = \mathbb{E}^{\mathscr{G}}(gu^+) - \mathbb{E}^{\mathscr{G}}(gu^-) = \mathbb{E}^{\mathscr{G}}(gu).$$

Finally, if g is $\mathscr{G}$ measurable, we see using $g^+g^- = 0$ that

$$(g^+ - g^-)\mathbb{E}^{\mathscr{G}}u = \mathbb{E}^{\mathscr{G}}(g^+u) - \mathbb{E}^{\mathscr{G}}(g^-u) = \mathbb{E}^{\mathscr{G}}(gu).$$

(viii) Since $(X, \mathscr{A}, \mu)$ is σ-finite, $\mathbb{E}^{\mathscr{G}} 1 = \sup_{n\in\mathbb{N}} \mathbb{E}^{\mathscr{G}} \mathbb{1}_{A_n}$ for some exhausting sequence $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ with $A_n \uparrow X$ and $\mu(A_n) < \infty$. We can now argue as in (vii) to get $\mathbb{E}^{\mathscr{G}} g = g\mathbb{E}^{\mathscr{G}} 1$.

(viii′) If $\mu|_{\mathscr{G}}$ is σ-finite, we can find an exhausting sequence $(G_n)_{n\in\mathbb{N}} \subset \mathscr{G}$ with $G_n \uparrow X$ and $\mu(G_n) < \infty$. Since $\mathbb{1}_{G_n} \in L^2(\mathscr{G})$, we find from Theorem 27.4(ix) that

$$\mathbb{E}^{\mathscr{G}} 1 = \sup_{n\in\mathbb{N}} \mathbb{E}^{\mathscr{G}} \mathbb{1}_{G_n} = \sup_{n\in\mathbb{N}} \mathbb{1}_{G_n} = 1.$$

If $g \in M(\mathscr{G})$, we combine $\mathbb{E}^{\mathscr{G}} 1 = 1$ and (viii) to see that $\mathbb{E}^{\mathscr{G}} g = g$.

(ix) If $(u_n)_{n\in\mathbb{N}} \subset L^2(\mathscr{A})$ approximates $0 \leqslant u \leqslant 1$ such that $u_n \uparrow u := \sup_{k\in\mathbb{N}} u_k$, we still have $v_n := u_n^+ \in L^2(\mathscr{A})$ and $v_n \uparrow u$. Thus Theorem 27.4(x) implies $0 \leqslant \mathbb{E}^{\mathscr{G}} u \leqslant 1$.

(xii) Considering the positive and negative parts separately, we may assume that $u \geqslant 0$. Since $u \in L^{\mathscr{G}}(\mathscr{A})$, there is an approximating sequence $(u_n)_{n\in\mathbb{N}} \subset L^2_+(\mathscr{A})$, $u = \sup_{n\in\mathbb{N}} u_n$, and, as $0 \leqslant u_n \leqslant u \in L^1(\mathscr{G})$, we have $u_n \in L^1(\mathscr{A})$. Theorem 27.4(xiii) gives, together with the definition of $\mathbb{E}^{\mathscr{G}}$ and Beppo Levi's theorem (Theorem 9.6),

$$\mathbb{E}^{\{\emptyset,X\}} u = \sup_{n\in\mathbb{N}} \mathbb{E}^{\{\emptyset,X\}} u_n = \sup_{n\in\mathbb{N}} \frac{1}{\mu(x)} \int u_n \, d\mu = \frac{1}{\mu(x)} \int u \, d\mu. \qquad \square$$

If we consider only the extension to $L^1(\mathscr{G})$, we are back in the classical case as it is usually considered in probability theory. This turns out to be a rather elegant way to rewrite the martingale property introduced in Definition 23.1.

Theorem 27.12 (traditional conditional expectation) *Assume that $(X, \mathscr{A}, \mu)$ is a σ-finite measure space and $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra such that $\mu|_{\mathscr{G}}$ is σ-finite. For $u \in L^1(\mathscr{A})$ and $g \in L^1(\mathscr{G})$ the following conditions are equivalent:*

(i) $\mathbb{E}^{\mathscr{G}} u = g$;

(ii) $\displaystyle\int_G u \, d\mu = \int_G g \, d\mu \quad \forall G \in \mathscr{G}$;

(iii) $\displaystyle\int_G u \, d\mu = \int_G g \, d\mu \quad \forall G \in \mathscr{G},\ \mu(G) < \infty$.

If $\mathscr{G}$ is generated by a $\cap$-stable family $\mathscr{F} \subset \mathscr{P}(X)$ containing a sequence $(F_n)_{n\in\mathbb{N}}$, $F_n \uparrow X$, then (i)–(iii) *are also equivalent to*

(iv) $\displaystyle\int_F u \, d\mu = \int_F g \, d\mu \qquad \forall F \in \mathscr{F}$.

Proof We begin with the general remark that by Theorem 27.11(iii) and (viii′) we have for all $G\in\mathscr{G}$ and $u\in L^1(\mathscr{A})$

$$\int_G \mathbb{E}^{\mathscr{G}}u\,d\mu = \langle \mathbb{E}^{\mathscr{G}}u, \mathbb{1}_G\rangle = \langle u, \mathbb{E}^{\mathscr{G}}\mathbb{1}_G\rangle = \langle u, \mathbb{1}_G\rangle = \int_G u\,d\mu. \tag{27.9}$$

(i) $\Rightarrow$ (ii). Because of (27.9) we get for all $G\in\mathscr{G}$

$$\int_G g\,d\mu \overset{\text{(i)}}{=} \int_G u\,d\mu \overset{(27.9)}{=} \int_G \mathbb{E}^{\mathscr{G}}u\,d\mu.$$

(ii) $\Rightarrow$ (iii) is obvious.

(iii) $\Rightarrow$ (i). Take a sequence $(G_n)_{n\in\mathbb{N}}\subset\mathscr{G}$ with $G_n\uparrow X$ and $\mu(G_n)<\infty$. Then we have for all $G\in\mathscr{G}$

$$\int_{G\cap G_n} \mathbb{E}^{\mathscr{G}}u\,d\mu \overset{(27.9)}{=} \int_{G\cap G_n} u\,d\mu \overset{\text{(iii)}}{=} \int_{G\cap G_n} g\,d\mu.$$

Since $|\mathbb{1}_{G\cap G_n}\mathbb{E}^{\mathscr{G}}u|\leqslant|\mathbb{E}^{\mathscr{G}}u|\in L^1$ and $|\mathbb{1}_{G\cap G_n}g|\leqslant|g|\in L^1$, we can use dominated convergence to let $n\to\infty$ and get

$$\int_G \mathbb{E}^{\mathscr{G}}u\,d\mu = \int_G g\,d\mu \quad \forall G\in\mathscr{G},$$

from which we conclude that $\mathbb{E}^{\mathscr{G}}u=g$ a.e. by Corollary 11.7(i).

Assume now, in addition, that $\mathscr{G}=\sigma(\mathscr{F})$. In this case, (ii) $\Rightarrow$ (iv) is obvious, while (iv) $\Rightarrow$ (ii) follows with the technique used in Remark 23.2(i): because of (iv) the finite measures

$$\rho(G) := \int_G (u^+ + g^-)d\mu \quad\text{and}\quad \nu(G) := \int_G (u^- + g^+)d\mu$$

coincide on $\mathscr{F}$, and by the uniqueness theorem for measures (Theorem 5.7), on the σ-algebra $\mathscr{G}$ which is generated by $\mathscr{F}$. □

If we combine Theorem 27.12 with the Beppo Levi theorem (Theorem 9.6) or other convergence theorems we can derive all sorts of 'conditional' versions of these theorems.

Corollary 27.13 (conditional Beppo Levi) *Let $(X,\mathscr{A},\mu)$ be a σ-finite measure space and $\mathscr{G}\subset\mathscr{A}$ a sub-σ-algebra such that $\mu|_{\mathscr{G}}$ is σ-finite. For every increasing sequence $(u_n)_{n\in\mathbb{N}}\subset L^1_+(\mathscr{A})$ of positive functions the limit $u:=\sup_{n\in\mathbb{N}}u_n$ admits a conditional expectation with values in $[0,\infty]$ and*

$$\sup_{n\in\mathbb{N}}\mathbb{E}^{\mathscr{G}}u_n = \mathbb{E}^{\mathscr{G}}\Big(\sup_{n\in\mathbb{N}}u_n\Big) = \mathbb{E}^{\mathscr{G}}u. \tag{27.10}$$

Proof Let $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ be an exhausting sequence of sets, i.e. $A_n \uparrow X$ and $\mu(A_n) < \infty$. Then the functions $w_n := (u_n \wedge n)\mathbb{1}_{A_n} \in L^\infty_+(\mathscr{A}) \cap L^1_+(\mathscr{A})$ and, in particular, $w_n \in L^2_+(\mathscr{A})$. [✍] Moreover, the sequence w_n increases towards u. From Definition 27.9 – or the remark preceding the statement of Theorem 27.5 combined with the monotonicity of $\mathbb{E}^\mathscr{G}$ (Theorem 27.4(xi)) – we get that

$$\mathbb{E}^\mathscr{G} u = \sup_{n\in\mathbb{N}} \mathbb{E}^\mathscr{G} w_n,$$

which is a function with values in $[0,\infty]$. On the other hand, we know from Theorem 27.12 that for all $G \in \mathscr{G}$

$$\int_G \mathbb{E}^\mathscr{G} w_n \, d\mu = \int_G w_n \, d\mu \quad \text{and} \quad \int_G \mathbb{E}^\mathscr{G} u_n \, d\mu = \int_G u_n \, d\mu$$

holds. Since $\sup_{n\in\mathbb{N}} w_n = u = \sup_{n\in\mathbb{N}} u_n$ and since the sequences $\left(\mathbb{E}^\mathscr{G} u_n\right)_{n\in\mathbb{N}}$ and $\left(\mathbb{E}^\mathscr{G} w_n\right)_{n\in\mathbb{N}}$ are positive and increasing, see Theorems 27.4(xi) and 27.11(x), we conclude from Beppo Levi's theorem (Theorem 9.6) that

$$\int_G \sup_{n\in\mathbb{N}} \mathbb{E}^\mathscr{G} u_n \, d\mu = \sup_{n\in\mathbb{N}} \int_G \mathbb{E}^\mathscr{G} u_n \, d\mu = \sup_{n\in\mathbb{N}} \int_G u_n \, d\mu = \int_G \sup_{n\in\mathbb{N}} u_n \, d\mu = \int_G u \, d\mu.$$

With a similar calculation we find

$$\int_G \sup_{n\in\mathbb{N}} \mathbb{E}^\mathscr{G} w_n \, d\mu = \int_G u \, d\mu,$$

and, consequently,

$$\int_G \sup_{n\in\mathbb{N}} \mathbb{E}^\mathscr{G} u_n \, d\mu = \int_G \sup_{n\in\mathbb{N}} \mathbb{E}^\mathscr{G} w_n \, d\mu \qquad \forall G \in \mathscr{G}.$$

With Corollary 11.7 we conclude that $\sup_{n\in\mathbb{N}} \mathbb{E}^\mathscr{G} u_n = \sup_{n\in\mathbb{N}} \mathbb{E}^\mathscr{G} w_n \stackrel{\text{def}}{=} \mathbb{E}^\mathscr{G} u$ almost everywhere. □

In the same way as we deduced Fatou's lemma (Theorem 9.11) and Lebesgue's dominated convergence theorem (Theorem 12.2) from the monotonicity property of the integral and Beppo Levi's theorem (Theorem 9.6), we can get their conditional versions from Theorem 27.4(xi) and (xii) and Corollary 27.13. We leave the simple proofs to the reader.

Corollary 27.14 (conditional Fatou's lemma) *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space, $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra, such that $\mu|_\mathscr{G}$ is σ-finite, and $(u_n)_{n\in\mathbb{N}} \subset L^1_+(\mathscr{A})$. Then*

$$\mathbb{E}^\mathscr{G}\left(\liminf_{n\to\infty} u_n\right) \leqslant \liminf_{n\to\infty} \mathbb{E}^\mathscr{G} u_n. \tag{27.11}$$

Corollary 27.15 (conditional dominated convergence theorem) *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space, $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra such that $\mu|_{\mathscr{G}}$ is σ-finite, and $(u_n)_{n\in\mathbb{N}} \subset L^1(\mathscr{A})$ such that $u_n \to u$ a.e. and $|u_n| \leqslant w$ for some $w \in L^1_+(\mathscr{A})$. Then*

$$\mathbb{E}^{\mathscr{G}}\Big(\lim_{n\to\infty} u_n\Big) = \lim_{n\to\infty} \mathbb{E}^{\mathscr{G}} u_n. \tag{27.12}$$

Corollary 27.16 (conditional Jensen inequality) *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra such that $\mu|_{\mathscr{G}}$ is σ-finite. Assume that $V : \mathbb{R} \to \mathbb{R}$ is a convex function and $\Lambda : \mathbb{R} \to \mathbb{R}$ a concave function. Then*

$$\mathbb{E}^{\mathscr{G}} \Lambda(u) \leqslant \Lambda\big(\mathbb{E}^{\mathscr{G}} u\big) \qquad \forall u \in L^{\mathscr{G}}(\mathscr{A}), \tag{27.13}$$

and, in particular, $\Lambda(u) \in L^{\mathscr{G}}(\mathscr{A})$. Moreover,

$$V\big(\mathbb{E}^{\mathscr{G}} u\big) \leqslant \mathbb{E}^{\mathscr{G}} V(u) \qquad \forall u \in L^{\mathscr{G}}(\mathscr{A}) \text{ s.t. } V(u) \in L^{\mathscr{G}}(\mathscr{A}). \tag{27.14}$$

Proof The argument is similar to the proof of Jensen's inequality Theorem 13.13. Note, however, that we do not have to require the finiteness of the reference measure – which is $w\mu$ in Theorem 13.13. Let us, for example, prove (27.14). Using Lemma 13.12 and denoting by $\sup_\ell$ the supremum over all affine linear functions ℓ such that $\ell(x) = ax + b \leqslant V(x)$ for all $x \in \mathbb{R}$, we get using Theorem 27.11(v), (viii′) and (x) that

$$V\big(\mathbb{E}^{\mathscr{G}} u\big) = \sup_\ell \big(a\mathbb{E}^{\mathscr{G}} u + b\big) = \sup_\ell \mathbb{E}^{\mathscr{G}}(au + b) \leqslant \mathbb{E}^{\mathscr{G}} V(u).$$

The inequality (27.13) is proved in the same way. □

We close this section with a very useful conditional version of Fubini's theorem.

Theorem 27.17 (conditional Fubini theorem) *Let $(X, \mathscr{A}, \mu), (Y, \mathscr{B}, \nu)$ be σ-finite measure spaces and $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra containing a sequence $(G_n)_{n\in\mathbb{N}}$ such that $G_n \uparrow X$ and $\mu(G_n) < \infty$. If $u = u(x, y)$ satisfies $u \in \mathcal{L}^1(\mu \times \nu)$, then*

$$\mathbb{E}^{\mathscr{G}} \int_Y u(\cdot\,, y)\nu(dy) = \int_Y \mathbb{E}^{\mathscr{G}} u(\cdot\,, y)\nu(dy). \tag{27.15}$$

Proof Since $\int_Y \int_X |u(x, y)|\mu(dx)\nu(dy) < \infty$, we know from Fubini's theorem (Corollary 14.9) that $\int_X |u(x, y)|\mu(dx) < \infty$ for ν-a. e. $y \in Y$. Therefore, $\mathbb{E}^{\mathscr{G}} u(\cdot\,, y)$ is well-defined for all $y \in Y \setminus N$ outside of the ν-null set $N \in \mathscr{B}$. If $y \in N$, we set $\mathbb{E}^{\mathscr{G}} u(\cdot\,, y) := 0$.

Assume for a moment that $(x,y)\mapsto \mathbb{E}^{\mathscr{G}}(u(\cdot\,,y))(x)$ is $\mathscr{G}\otimes\mathscr{B}$-measurable. We have

$$\begin{aligned}\int_Y\int_X |\mathbb{E}^{\mathscr{G}}u(\cdot\,,y)(x)|\mu(dx)\nu(dy) &= \int_Y \|\mathbb{E}^{\mathscr{G}}u(\cdot\,,y)\|_{L^1(\mu)}\,\nu(dy)\\ &\overset{27.11(\mathrm{ii})}{\leqslant} \int_Y \|u(\cdot\,,y)\|_{L^1(\mu)}\,\nu(dy)<\infty,\end{aligned}$$

i.e. the function $(x,y)\mapsto \mathbb{E}^{\mathscr{G}}u(\cdot\,,y)(x)$ is in $L^1(\mu\times\nu)$.

Moreover, for all $G\in\mathscr{G}$, Fubini's theorem and Theorem 27.12 show

$$\begin{aligned}\int_G\left(\int_Y u(x,y)\nu(dy)\right)\mu(dx) &= \int_Y\int_G u(x,y)\mu(dx)nu(dy)\\ &= \int_Y\int_G \mathbb{E}^{\mathscr{G}}u(\cdot\,,y)(x)\mu(dx)\nu(dy)\\ &= \int_G\int_Y \mathbb{E}^{\mathscr{G}}u(\cdot\,,y)(x)\nu(dy)\mu(dx).\end{aligned}$$

Using Theorem 27.12, we conclude that (27.15) holds.

The $\mathscr{G}\otimes\mathscr{B}$-measurability of $(x,y)\mapsto \mathbb{E}^{\mathscr{G}}(u(\cdot\,,y))(x)$ follows with a monotone class argument, see Theorem 8.15. For $A\times B\in\mathscr{A}\times\mathscr{B}$ we deduce from Theorem 27.11(vii)–(viii′) that

$$\mathbb{E}^{\mathscr{G}}\mathbb{1}_{A\times B}(\cdot\,,y)=\mathbb{1}_B(y)\mathbb{E}^{\mathscr{G}}\mathbb{1}_A(\cdot),$$

i.e. $\mathbb{E}^{\mathscr{G}}\mathbb{1}_{A\times B}(\cdot\,,y)(x)$ is clearly $\mathscr{G}\otimes\mathscr{B}$-measurable. Denote by $\mathcal{V}$ the set of all $u\in\mathcal{M}(\mathscr{A}\otimes\mathscr{B})$ such that $\mathbb{E}^{\mathscr{G}}u(\cdot\,,y)$ is well-defined and $\mathscr{G}\otimes\mathscr{B}$-measurable. We have seen that $\mathbb{1}_{A\times B}\in\mathcal{V}$ for all $A\in\mathscr{A}, B\in\mathscr{B}$. Because of the linearity of the conditional expectation, $\mathcal{V}$ is a vector space, and the conditional Beppo Levi theorem ensures that the condition (ii) of Theorem 8.15 is satisfied. This shows that $\mathcal{V}$ contains all functions from $\mathcal{M}(\mathscr{A}\otimes\mathscr{B})$ for which the conditional expectation can be defined, in particular the family $L^1(\mu\times\nu)$. □

Conditional Expectations and Martingales

Because of Theorem 27.12 it is now very easy and convenient to express the martingale property Definition 23.1 in terms of conditional expectations. In fact, we have the following.

Corollary 27.18 *Let $(X,\mathscr{A},\mathscr{A}_n,\mu)$ be a σ-finite filtered measure space. A sequence $(u_n)_{n\in\mathbb{N}}\subset L^1(\mathscr{A})$ such that $u_n\in L^1(\mathscr{A}_n)$ is a martingale* (*resp. sub- or supermartingale*) *if, and only if, for all $n\in\mathbb{N}$*

$$\mathbb{E}^{\mathscr{A}_n}u_{n+1}=u_n \qquad \Big(\textit{resp.}\quad \mathbb{E}^{\mathscr{A}_n}u_{n+1}\geqslant u_n \quad \textit{or}\quad \mathbb{E}^{\mathscr{A}_n}u_{n+1}\leqslant u_n\Big).$$

A great advantage of this way of putting things is that we can now formulate the convergence theorem for uniformly integrable martingales Theorem 24.6 in a very striking way, as follows.

Theorem 27.19 (closability of martingales) *Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space and $\mathscr{A}_\infty = \sigma\big(\bigcup_{n\in\mathbb{N}} \mathscr{A}_n\big)$.*

(i) *For every $u \in L^1(\mathscr{A})$ the sequence $\big(\mathbb{E}^{\mathscr{A}_n}u\big)_{n\in\mathbb{N}}$ is a uniformly integrable martingale. In particular, $\lim_{n\to\infty} \mathbb{E}^{\mathscr{A}_n}u = \mathbb{E}^{\mathscr{A}_\infty}u$ in L^1 and a.e.*

(ii) *Conversely, if $(u_n)_{n\in\mathbb{N}}$ is a uniformly integrable martingale, there exists a function $u_\infty \in L^1(\mathscr{A}_\infty)$ such that $\lim_{n\to\infty} u_n = u_\infty$ in L^1 and a.e. In particular, $(u_n)_{n\in\mathbb{N}\cup\{\infty\}}$ is a martingale and $u_n = \mathbb{E}^{\mathscr{A}_n}u_\infty$. In this sense, u_∞ closes the martingale $(u_n)_{n\in\mathbb{N}}$.*

Proof (i) That $\big(\mathbb{E}^{\mathscr{A}_n}u\big)_{n\in\mathbb{N}}$ is a martingale follows at once from Theorem 27.11(vi). By assumption there exists an exhausting sequence $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}_0$ with $A_n \uparrow X$ and $\mu(A_n) < \infty$. Therefore, the function

$$w := \sum_{n=1}^{\infty} \frac{2^{-n}}{1+\mu(A_n)} \mathbb{1}_{A_n}$$

is strictly positive and integrable; in particular, $\{|u| \geqslant Nw\} \downarrow \emptyset$ as $N \to \infty$.

Since $\{|\mathbb{E}^{\mathscr{A}_n}u| > Nw\} \in \mathscr{A}_n$, we get

$$\begin{aligned}
\int_{\{|\mathbb{E}^{\mathscr{A}_n}u|>2Nw\}} \mathbb{E}^{\mathscr{A}_n}|u|\,d\mu &\overset{27.12}{=} \int_{\{|\mathbb{E}^{\mathscr{A}_n}u|>2Nw\}} |u|\,d\mu \\
&= \int_{\{|\mathbb{E}^{\mathscr{A}_n}u|>2Nw\}\cap\{|u|>Nw\}} |u|\,d\mu + \int_{\{|\mathbb{E}^{\mathscr{A}_n}u|>2Nw\}\cap\{|u|\leqslant Nw\}} |u|\,d\mu \\
&\leqslant \int_{\{|u|>Nw\}} |u|\,d\mu + \frac{1}{2}\int_{\{|\mathbb{E}^{\mathscr{A}_n}u|>2Nw\}} \mathbb{E}^{\mathscr{A}_n}|u|\,d\mu
\end{aligned}$$

where we use the fact that $|u| \leqslant Nw < \frac{1}{2}|\mathbb{E}^{\mathscr{A}_n}u| \leqslant \frac{1}{2}\mathbb{E}^{\mathscr{A}_n}|u|$ for the integrand of the second integral. Thus,

$$\int_{\{|\mathbb{E}^{\mathscr{A}_n}u|>2Nw\}} |\mathbb{E}^{\mathscr{A}_n}u|\,d\mu \overset{27.11(\mathrm{xi})}{\leqslant} \int_{\{|\mathbb{E}^{\mathscr{A}_n}u|>2Nw\}} \mathbb{E}^{\mathscr{A}_n}|u|\,d\mu \leqslant \frac{1}{2}\int_{\{|u|>Nw\}} |u|\,d\mu$$

and the right-hand side converges, by dominated convergence, to zero as $N \to \infty$. Since this convergence is uniform in $n \in \mathbb{N}$, this proves uniform integrability.

The convergence assertions follow now from the convergence theorem for UI submartingales, Theorem 24.6.

(ii) follows directly from Theorem 24.6. □

Since the conditional Jensen inequality needs fewer assumptions than the classical Jensen inequality we can improve Example 23.3(v) and (vi).

Corollary 27.20 *Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space and $(u_n)_{n\in\mathbb{N}}$ be a family of measurable functions $u_n \in L^{\mathscr{A}_n}(\mathscr{A})$ which satisfies the [sub]martingale property*[5]

$$u_n = \mathbb{E}^{\mathscr{A}_n} u_{n+1} \qquad \Big[\textit{resp.} \quad u_n \leqslant \mathbb{E}^{\mathscr{A}_n} u_{n+1}\Big].$$

If $V : \mathbb{R} \to \mathbb{R}$ is a [monotone increasing] convex function and $V(u_n) \in L^1(\mathscr{A}_n)$, then $(V(u_n))_{n\in\mathbb{N}}$ is a submartingale.

Proof Since $(u_n)_{n\in\mathbb{N}}$ satisfies the [sub]martingale property, we find from Jensen's inequality Corollary 27.16 [and the monotonicity of V] that

$$V(u_n) \leqslant V\big(\mathbb{E}^{\mathscr{A}_n} u_{n+1}\big) \leqslant \mathbb{E}^{\mathscr{A}_n} V(u_{n+1}). \qquad \square$$

Example 27.21 In Example 23.3(ix) we introduced a *dyadic filtration* on the measure space $\big([0,\infty)^n, \mathscr{B}([0,\infty)^n), \lambda = \lambda^n|_{[0,\infty)^n}\big)$ given by

$$\mathscr{A}_k^{\Delta} = \sigma\big(z + [0, 2^{-k})^n : z \in 2^{-k}\mathbb{N}_0^n\big), \quad k \in \mathbb{N}_0.$$

For $u \in L^1([0,\infty)^n, \lambda)$ and all $k \in \mathbb{N}_0$ we can now rewrite (23.4) as

$$\mathbb{E}^{\mathscr{A}_k^{\Delta}} u(x) = \sum_{z \in 2^{-k}\mathbb{N}_0^n} \left\{ \int u \frac{\mathbb{1}_{z+[0,2^{-k})^n}}{\lambda\big(z + [0, 2^{-k})^n\big)}\, d\lambda \right\} \mathbb{1}_{z+[0,2^{-k})^n}(x).$$

On the Structure of Subspaces of L^2

As we have seen, $\mathbb{E}^{\mathscr{G}} : L^2(\mathscr{A}) \to L^2(\mathscr{G})$ is the symmetric orthogonal projection onto the closed subspace $L^2(\mathscr{G})$ of the Hilbert space $L^2(\mathscr{A})$. It is natural to ask whether *every* orthogonal projection $\pi : L^2(\mathscr{A}) \to \mathcal{H}$ onto a closed subspace $\mathcal{H} \subset L^2(\mathscr{A})$ is a conditional expectation. Equivalently we could ask under which conditions a closed subspace $\mathcal{H}$ of $L^2(\mathscr{A})$ is of the form $L^2(\mathscr{G}) = \mathcal{H}$ for a suitable sub-σ-algebra $\mathscr{G} \subset \mathscr{A}$.

Theorem 27.22 *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space. For a closed linear subspace $\mathcal{H} \subset L^2(\mathscr{A})$ and its orthogonal projection $\pi = P_{\mathcal{H}} : L^2(\mathscr{A}) \to \mathcal{H}$, the following assertions are equivalent.*

(i) *$\mathcal{H} = L^2(\mathscr{G})$ and $\pi = \mathbb{E}^{\mathscr{G}}$ for some sub-σ-algebra $\mathscr{G} \subset \mathscr{A}$ containing an exhausting sequence $(G_n)_{n\in\mathbb{N}} \subset \mathscr{G}$ with $G_n \uparrow X$ and $\mu(G_n) < \infty$.*

[5] This is slightly more general than assuming that $(u_n)_{n\in\mathbb{N}}$ is a [sub]martingale, since [sub]martingales are, by definition, integrable.

(ii) *π is a sub-Markovian operator, i.e.* $0 \leqslant u \leqslant 1 \implies 0 \leqslant \pi(u) \leqslant 1$, $u \in L^2(\mathscr{A})$, *and for some* $u_0 \in L^2(\mathscr{A})$ *with* $u_0 > 0$ *we have* $\pi(u_0) > 0$.

(iii) $\mathcal{H} \cap L^\infty(\mathscr{A})$ *is an algebra – i.e. it is closed under pointwise products, namely* $f, h \in \mathcal{H} \cap L^\infty(\mathscr{A}) \implies fh \in \mathcal{H} \cap L^\infty(\mathscr{A})$ *– which is* L^2*-dense in* $\mathcal{H}$ *and contains an (everywhere) strictly positive function* $h_0 > 0$.

(iv) $\mathcal{H}$ *is a lattice – i.e.* $f, h \in \mathcal{H} \implies f \wedge h \in \mathcal{H}$ *– containing an (everywhere) strictly positive function* $h_0 > 0$*, and for all* $h \in \mathcal{H}$ *also* $h \wedge 1 \in \mathcal{H}$.

Proof We show that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iv) $\Rightarrow$ (i) $\Rightarrow$ (iii) $\Rightarrow$ (iv).

(i) $\Rightarrow$ (ii). The sub-Markov property of $\pi = \mathbb{E}^{\mathscr{G}}$ follows from Theorem 27.4(x), while

$$u_0 := \sum_{n=1}^{\infty} \frac{2^{-n}}{\sqrt{\mu(G_n) + 1}} \mathbb{1}_{G_n} \in \mathcal{M}^+(\mathscr{G})$$

clearly satisfies $0 < u_0 \leqslant 1$,

$$\begin{aligned} \|u_0\|_2 = \left\| \sum_{n=1}^{\infty} \frac{2^{-n}}{\sqrt{\mu(G_n) + 1}} \mathbb{1}_{G_n} \right\|_2 &\leqslant \sum_{n=1}^{\infty} \frac{2^{-n}}{\sqrt{\mu(G_n) + 1}} \|\mathbb{1}_{G_n}\|_2 \\ &= \sum_{n=1}^{\infty} 2^{-n} \sqrt{\frac{\mu(G_n)}{\mu(G_n) + 1}} \leqslant 1, \end{aligned}$$

so that $u_0 \in L^2(\mathscr{G})$ and, therefore, $0 < \pi(u_0) = u_0 \leqslant 1$.

(ii) $\Rightarrow$ (iv). Since π preserves positivity, we find for all $u \in L^2(\mathscr{A})$ that $\pi(u^+) \geqslant 0$ and $\pi(u) = \pi(u^+) - \pi(u^-) \leqslant \pi(u^+)$, thus $\pi(u) \vee 0 \leqslant \pi(u^+)$.

On the other hand, $\mathcal{H} = \{h \in L^2(\mathscr{A}) : \pi(h) = h\}$ and the above calculation shows, for $h \in \mathcal{H}$, that

$$h^+ = (\pi(h))^+ = \pi(h) \vee 0 \leqslant \pi(h^+). \tag{27.16}$$

Since π is a contraction, see (26.10), we find also

$$\|h^+\|_2 \overset{(27.16)}{\leqslant} \|\pi(h^+)\|_2 \leqslant \|h^+\|_2,$$

which implies $\langle \pi(h^+), \pi(h^+) \rangle = \langle \pi(h^+), h^+ \rangle = \langle h^+, h^+ \rangle$. Because of (27.16) we get $\langle \underbrace{\pi(h^+) - h^+}_{\geqslant 0}, h^+ \rangle = 0$ or $\pi(h^+) = h^+$ on the set $\{h^+ > 0\}$. But then

$$\|\pi(h^+)\|_2 = \|h^+\|_2 \implies \int_{\{h^+ = 0\}} \pi(h^+)^2 \, d\mu = \int_{\{h^+ = 0\}} (h^+)^2 \, d\mu = 0,$$

which shows that $\pi(h^+) = 0$ on $\{h^+ = 0\}$ or $\mu\{h^+ = 0\} = 0$. In either case, $\pi(h^+) = h^+$ (almost everywhere) and $h^+ \in \mathcal{H}$.

Consequently, $f \wedge h = f - (f - h)^+ \in \mathcal{H}$. Similarly, $h \wedge 1 = h - (h-1)^+$ and, if $h \in \mathcal{H}$, we see $\pi(h \wedge 1) \leqslant \pi(h) \wedge 1 = h \wedge 1$. Further,

$$\pi((h-1)^+) = \pi(h) - \pi(h \wedge 1) \geqslant h - h \wedge 1 = (h-1)^+,$$

and, since π is a contraction, the same argument as what we use to get $\pi(h^+) = h^+$ yields $\pi((h-1)^+) = (h-1)^+$, hence $(h-1)^+, h \wedge 1 \in \mathcal{H}$. Finally, $h_0 := \pi(u_0)$, u_0 as in (ii), satisfies $h_0 \in \mathcal{H}$ and $h_0 > 0$.

(iv) $\Rightarrow$ (i). We set

$$\mathscr{G} := \{G \in \mathscr{A} : h \wedge \mathbb{1}_G \in \mathcal{H} \quad \forall h \in \mathcal{H}\}.$$

Let us first show that $\mathscr{G}$ is a σ-algebra. Clearly, $\emptyset, X \in \mathscr{G}$. If $G \in \mathscr{G}$, then

$$h \wedge \mathbb{1}_{G^c} + \underbrace{h \wedge \mathbb{1}_G}_{\in \mathcal{H}} = h \wedge 1 + h \wedge 0 \in \mathcal{H} \quad \forall h \in \mathcal{H},$$

which means that $h \wedge \mathbb{1}_{G^c} \in \mathcal{H}$ and $G^c \in \mathscr{G}$. For any two sets $G, H \in \mathscr{G}$ we see[6]

$$h \wedge \mathbb{1}_{G \cup H} = h \wedge (\mathbb{1}_G \vee \mathbb{1}_H) = (h \wedge \mathbb{1}_G) \vee (h \wedge \mathbb{1}_H) \in \mathcal{H} \quad \forall h \in \mathcal{H},$$

so that $G \cup H \in \mathscr{G}$. Finally, let $(G_n)_{n \in \mathbb{N}} \subset \mathscr{G}$; since $\mathscr{G}$ is $\cup$-stable, we may assume that $G_n \uparrow G := \bigcup_{n \in \mathbb{N}} G_n$. Then

$$(h \wedge \mathbb{1}_{G_n})_{n \in \mathbb{N}} \subset \mathcal{H} \quad \text{and} \quad \lim_{n \to \infty} (h \wedge \mathbb{1}_{G_n}) = h \wedge \mathbb{1}_G \in L^2(\mathscr{A}) \quad \forall h \in \mathcal{H}.$$

Since $|h \wedge \mathbb{1}_{G_n}| \leqslant h \wedge \mathbb{1}_G \in L^2$, an application of the dominated convergence theorem (Theorem 13.9) shows that, in $L^2(\mathscr{A})$, $\lim_{n \to \infty} h \wedge \mathbb{1}_{G_n} = h \wedge \mathbb{1}_G$ for all $h \in \mathcal{H}$. Since $\mathcal{H}$ is a closed subspace, we conclude that $h \wedge \mathbb{1}_G \in \mathcal{H}$ for all $h \in \mathcal{H}$, thus $G \in \mathscr{G}$.

We will now show that $L^2(\mathscr{G}) = \mathcal{H}$. If $f \in \mathcal{H}$ we know from our assumptions that $(\pm f) \wedge 0 \in \mathcal{H}$, so that $f^+ = -((-f) \wedge 0)$, $f^- = -(f \wedge 0) \in \mathcal{H}$. Thus $\mathcal{H} = \mathcal{H}^+ - \mathcal{H}^+$, and since also $L^2(\mathscr{G}) = L^2_+(\mathscr{G}) - L^2_+(\mathscr{G})$ it clearly suffices to show that $\mathcal{H}^+ = L^2_+(\mathscr{G})$.

Assume that $f \in \mathcal{H}^+$. Then, for $a > 0$, $f \wedge a = a(f/a \wedge 1) \in \mathcal{H}$, and, by monotone convergence Theorem 12.1 and the closedness of $\mathcal{H}$,

$$h \wedge \mathbb{1}_{\{f > a\}} = h \wedge \sup_{n \in \mathbb{N}} \big\{[nf - n(f \wedge a)] \wedge 1\big\} \in \mathcal{H} \quad \forall h \in \mathcal{H},$$

proving that $\{f > a\} \in \mathscr{G}$ for all $a > 0$. Moreover, $\{f > a\} = X$ if $a < 0$ and $\{f > 0\} = \bigcup_{k \in \mathbb{N}} \{f > 1/k\}$, which shows that $\{f > a\} \in \mathscr{G}$ for all $a \in \mathbb{R}$ and, consequently, $\mathcal{H}^+ \subset L^2_+(\mathscr{G})$.

[6] Use that $\mathcal{H}$ is a vector space and $a \vee b = -((-a) \wedge (-b))$.

Conversely, if $g\in L^2_+(\mathscr{G})$, we can write g as a limit of simple functions, namely $g_k=\sum_{n=1}^{N(k)} y_n^{(k)}\mathbb{1}_{G_n^{(k)}}$ with disjoint sets $G_1^{(k)},\dots,G_{N(k)}^{(k)}\in\mathscr{G}$ and $y_n^{(k)}>0$. For all $h\in\mathcal{H}$ we find

$$g_k\wedge h=\sum_{n=1}^{N(k)}\big(y_n^{(k)}\mathbb{1}_{G_n^{(k)}}\big)\wedge h=\sum_{n=1}^{N(k)} y_n^{(k)}\left(\mathbb{1}_{G_n^{(k)}}\wedge\frac{h}{y_n^{(k)}}\right)\in\mathcal{H},$$

and dominated convergence and the closedness of $\mathcal{H}$ imply that $g\wedge h\in\mathcal{H}$. Choosing, in particular, $h=nh_0$ for some a.e. strictly positive function h_0 and letting $n\to\infty$ gives

$$g=L^2\text{-}\lim_{n\to\infty}(nh_0)\wedge g\in\mathcal{H}^+,$$

where we again use monotone convergence and the closedness of $\mathcal{H}$. This proves $L^2_+(\mathscr{G})\subset\mathcal{H}^+$.

Finally, the sets $G_n:=\{h_0>1/n\}\in\mathscr{G}$ satisfy $G_n\uparrow X$ and, because of the Markov inequality Proposition 11.5, $\mu(G_n)=\mu\{h_0>1/n\}\leqslant n^2\int h_0^2\,d\mu<\infty$.

(i)$\Rightarrow$(iii). Note that $L^2(\mathscr{G})\cap L^\infty(\mathscr{G})=L^2(\mathscr{G})\cap L^\infty(\mathscr{A})$. An application of the dominated convergence theorem (Theorem 13.9) then shows that the sequence $f_n:=(-n)\vee f\wedge n$, $n\in\mathbb{N}$, $f\in L^2(\mathscr{G})$, converges in $L^2(\mathscr{G})$ to f, i.e. we have that $L^2(\mathscr{G})\cap L^\infty(\mathscr{G})$ is a dense subset of $L^2(\mathscr{G})$. The element $h_0>0$ is now constructed as in the proof of (i)$\Rightarrow$(ii). That $L^2(\mathscr{G})\cap L^\infty(\mathscr{G})$ is an algebra is trivial.

(iii)$\Rightarrow$(iv). Let us show, first of all, that $\mathcal{H}\cap L^\infty(\mathscr{A})$ is stable under minima. To this end we define recursively a sequence of polynomials in $\mathbb{R}$,

$$p_0(x):=0,\qquad p_{n+1}(x):=p_n(x)+\tfrac12(x^2-p_n^2(x)),\quad n\in\mathbb{N}_0.$$

By induction it is easy to see that $p_n(0)=0$ for all $n\in\mathbb{N}_0$ and that

$$0\leqslant p_n(x)\leqslant p_{n+1}(x)\leqslant|x|\qquad\forall x\in[-1,1].$$

For $n=0$ there is nothing to show. Otherwise we can use the induction assumption $p_n(x)\leqslant p_{n+1}(x)\leqslant|x|$ to get

$$\begin{aligned}0\leqslant p_{n+1}(x)&\leqslant\overbrace{p_{n+1}(x)+\underbrace{\frac12\big(x^2-p_{n+1}^2(x)\big)}_{\geqslant0}}^{\overset{\text{def}}{=}p_{n+2}(x)}\\&=|x|-\underbrace{(|x|-p_{n+1}(x))}_{\geqslant0}\cdot\underbrace{\left(1-\frac12(|x|+p_{n+1}(x))\right)}_{\geqslant0\text{ for }x\in[-1,1]}\\&\leqslant|x|.\end{aligned}$$

Therefore, $\lim_{n\to\infty} p_n(x) = \sup_{n\in\mathbb{N}} p_n(x) = \ell(x)$ exists for all $|x| \leqslant 1$ and, according to the recursion relation, $\ell(x) = |x|$.

Since $\mathcal{H}$ is a linear subspace which is stable under products, we get for every $h \in \mathcal{H} \cap L^\infty(\mathscr{A})$ that $p_n(h/\|h\|_\infty) \in \mathcal{H}$, and monotone convergence, Theorem 12.1, and the closedness of $\mathcal{H}$ show that

$$\sup_{n\in\mathbb{N}} p_n\Big(\frac{h}{\|h\|_\infty}\Big) = \frac{|h|}{\|h\|_\infty} \in \mathcal{H} \implies |h| \in \mathcal{H}.$$

As $\mathcal{H} \cap L^\infty(\mathscr{A})$ is dense in $\mathcal{H}$, we find for $h \in \mathcal{H}$ a sequence $(h_k)_{k\in\mathbb{N}} \subset \mathcal{H} \cap L^\infty(\mathscr{A})$ such that $L^2\text{-}\lim_{k\to\infty} h_k = h$. From the above we know, however, that $|h_k| \in \mathcal{H}$ and $|h_k| \to |h|$ in $L^2(\mathscr{A})$, thus $|h| \in \mathcal{H}$. This shows, in particular, that

$$f \wedge h = \frac{1}{2}(f + h - |f - h|) \in \mathcal{H} \quad \forall f, h \in \mathcal{H}.$$

Since $0 < h_0 \leqslant \|h_0\|_\infty$, we get for all $n > \|h_0\|_\infty$ that

$$\frac{n}{n - h_0} = \sum_{i=0}^{\infty} \left(\frac{h_0}{n}\right)^i = \sup_{N\in\mathbb{N}} \sum_{i=0}^{N} \left(\frac{h_0}{n}\right)^i$$

and, for $h \in \mathcal{H}$,

$$h \wedge \frac{n}{n - h_0} = \lim_{N\to\infty} \underbrace{\sum_{i=0}^{N} \left(\frac{h_0}{n}\right)^i \wedge h}_{\in\mathcal{H}}.$$

By monotone convergence (Theorem 12.1) we conclude that $h \wedge n/(n - h_0) \in \mathcal{H}$. Finally, as $n/(n - h_0) \downarrow 1$ and $h^2 \wedge ((n/n - h_0))^2 \leqslant h^2$, we can use the dominated convergence theorem (Theorem 13.9) and the closedness of $\mathcal{H}$ to see that $h \wedge 1 \in \mathcal{H}$. $\square$

Separability Criteria for the Spaces $L^p(X, \mathscr{A}, \mu)$

Let $(X, \mathscr{A}, \mu)$ be a measure space. Recall that $L^p(\mathscr{A})$ is *separable* if it contains a countable dense subset $(d_n)_{n\in\mathbb{N}} \subset L^p(\mathscr{A})$. We have seen in Chapter 26 that the Hilbert space $L^2(\mathscr{A})$ is separable if we can find a countable *complete* orthonormal system $(e_n)_{n\in\mathbb{N}} \subset L^2(\mathscr{A})$ since the system $\{q_1 e_1 + \cdots + q_N e_N : N \in \mathbb{N}, q_n \in \mathbb{Q}\}$ is both countable and dense. Conversely, using any countable dense subset $(d_n)_{n\in\mathbb{N}}$ as input for the Gram–Schmidt orthonormalization procedure (26.16) produces a complete countable orthonormal system.

Here is a simple sufficient criterion for the separability of L^p.

Lemma 27.23 *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and assume that the σ-algebra $\mathscr{A}$ is countably generated, i.e. $\mathscr{A} = \sigma(A_n : n \in \mathbb{N})$, $A_n \subset X$. Then $L^p(X, \mathscr{A}, \mu)$, $1 \leqslant p < \infty$, is separable.*

Proof Step 1. Let us first assume that μ is a *finite measure*. Consider the σ-algebras $\mathscr{A}_n := \sigma(A_1, \dots, A_n)$; clearly,

$$\mathscr{A}_1 \subset \mathscr{A}_2 \subset \cdots \subset \mathscr{A}_\infty = \sigma(\mathscr{A}_n : n \in \mathbb{N})$$

is a filtration, $\mu|_{\mathscr{A}_n}$ is trivially σ-finite for every $n \in \mathbb{N}$, and $\mathscr{A} = \mathscr{A}_\infty$.

Set $u_n := \mathbb{E}^{\mathscr{A}_n} u$ for $u \in L^1(\mathscr{A})$. By Theorem 27.19, $(u_n, \mathscr{A}_n)_{n \in \mathbb{N}}$ is a uniformly integrable martingale, hence $u_n \to u$ in L^1 and a.e.

If $v \in L^p(\mathscr{A})$, we set $v_n := |\mathbb{E}^{\mathscr{A}_n} v|^p \leqslant \mathbb{E}^{\mathscr{A}_n}(|v|^p)$ (by Corollary 27.16) and observe that $(v_n, \mathscr{A}_n)_{n \in \mathbb{N}}$ is a submartingale, see Theorem 27.19, which is uniformly integrable. The latter follows easily from

$$\int_{\{v_n > w\}} v_n \, d\mu \leqslant \int_{\{v_n > w\}} \mathbb{E}^{\mathscr{A}_n}(|v|^p) d\mu \leqslant \int_{\{\mathbb{E}^{\mathscr{A}_n}(|v|^p) > w\}} \mathbb{E}^{\mathscr{A}_n}(|v|^p) d\mu$$

and the uniform integrability of the family $\left(\mathbb{E}^{\mathscr{A}_n}(|v|^p)\right)_{n \in \mathbb{N}}$, see Theorem 27.19. From the (sub)martingale convergence theorem (Theorem 24.6) we conclude that $v_n \to |\mathbb{E}^{\mathscr{A}_\infty} v|^p = |v|^p$ in L^1 and a.e., and Riesz's theorem (Theorem 13.10) shows that $|v_n|^{1/p} = \left|\mathbb{E}^{\mathscr{A}_n} v\right| \to |v|$ in L^p. Consequently, $\mathbb{E}^{\mathscr{A}_n} v \to v$ in L^p.

Since the σ-algebra $\mathscr{A}_n$ is generated by finitely many sets, $\mathbb{E}^{\mathscr{A}_n} u$, resp., $\mathbb{E}^{\mathscr{A}_n} v$ are simple functions with canonical representations of the form

$$s = \sum_{i=1}^{N} y_i \mathbb{1}_{B_i}, \quad y_i \neq 0, \ B_1, \dots, B_N \in \mathscr{A}_n \text{ disjoint;}$$

as $\mathscr{A}_n$ is kept fixed, we suppress the dependence of y_i, B_i, N on n. If $y_n \notin \mathbb{Q}$, we find for every $\epsilon > 0$ numbers $y_i^\epsilon \in \mathbb{Q}$ such that

$$|y_i - y_i^\epsilon| \leqslant \frac{\epsilon}{N \mu(X)^{1/p}} \, .$$

The triangle inequality now shows that

$$\left\| s - \sum_{i=1}^{N} y_i^\epsilon \mathbb{1}_{B_i} \right\|_p \leqslant \sum_{i=1}^{N} |y_i - y_i^\epsilon| \, \mu(B_i)^{1/p} \leqslant \epsilon,$$

which proves that the system

$$D := \left\{ \sum_{n=1}^{N} q_n \mathbb{1}_{B_n} : N \in \mathbb{N}, \ q_n \in \mathbb{Q}, \ B_n \in \bigcup_{n \in \mathbb{N}} \mathscr{A}_n \right\}$$

is a countable dense subset of the space $L^p(X, \mathscr{A}, \mu)$, $1 \leqslant p < \infty$.

Step 2. If μ is σ-finite but not finite, we choose a sequence $(C_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ such that $C_n \uparrow X$ and $\mu(C_n) < \infty$, and consider the finite measures $\mu_n := \mu(\cdot \cap C_n)$, $n \in \mathbb{N}$, on $C_n \cap \mathscr{A}$. Since every $u \in L^p(\mu_n) = L^p(C_n, C_n \cap \mathscr{A}, \mu_n)$ can be extended by 0 on the set $X \setminus C_n$ and becomes an element of $L^p(\mu_{n+1})$, we can interpret the sets $L^p(\mu_n)$ as a chain of increasing subspaces of each other and of $L^p(\mu)$:

$$L^p(C_n, C_n \cap \mathscr{A}, \mu_n) \subset L^p(C_{n+1}, C_{n+1} \cap \mathscr{A}, \mu_{n+1}) \subset \cdots \subset L^p(X, \mathscr{A}, \mu).$$

Applying the construction from Step 1 to each of the sets $L^p(\mu_n)$ furnishes countable dense subsets D_n. Obviously, $D := \bigcup_{n\in\mathbb{N}} D_n$ is a countable set but it is also dense in $L^p(\mu)$. To see this, fix $\epsilon > 0$ and $u \in L^p(\mu)$. Since $X \setminus C_n \downarrow \emptyset$, we find by Lebesgue's dominated convergence theorem (Theorem 13.9) some $N \in \mathbb{N}$ such that $\int_{X\setminus C_n} |u|^p \, d\mu < \epsilon^p$ for all $n \geqslant N$. Since D_n is dense in $L^p(\mu_n)$ and since $u\mathbb{1}_{C_n} \in L^p(\mu_n)$, there exists some $d_n \in D_n \subset D$ with $\|u\mathbb{1}_{C_n} - d_n\|_{L^p(\mu_n)} \leqslant \epsilon$, and altogether we get for large $n \geqslant N$

$$\|u - d_n\|_p \leqslant \|u\mathbb{1}_{C_n} - d_n\|_{L^p(\mu_n)} + \|u\mathbb{1}_{X\setminus C_n}\|_{L^p(\mu)} \leqslant 2\epsilon. \qquad \square$$

If the underlying set X is a *separable metric space*, the criterion of Lemma 27.23 becomes particularly simple.

Corollary 27.24 *Let (X, ρ) be a separable metric space equipped with its Borel σ-algebra $\mathscr{B} = \mathscr{B}(X)$. Then $L^p(X, \mathscr{B}, \mu)$, $1 \leqslant p < \infty$, is separable for every σ-finite measure μ on $(X, \mathscr{B})$. If μ is not σ-finite, $L^p(X, \mathscr{B}, \mu)$ need not be separable.*

Proof Denote by $D \subset X$ a countable dense subset and consider the countable system of open balls $B_r(d) := \{x \in X : \rho(x, d) < r\}$

$$\mathscr{F} := \{B_r(d) : d \in D,\ r \in \mathbb{Q}^+\} \subset \mathscr{O}(X).$$

Since every open set $U \in \mathscr{O}(X)$ can be written as[7]

$$U = \bigcup_{\mathscr{F} \ni B_r(d) \subset U} B_r(d)$$

which shows that $\mathscr{O}(X) \subset \sigma(\mathscr{F}) \subset \sigma(\mathscr{O}(X)) = \mathscr{B}(X)$. Thus the Borel sets $\mathscr{B}(X) = \sigma(\mathscr{F})$ are countably generated, and the assertion follows from Lemma 27.23.

If μ is not σ-finite, we have the following counterexample: take $X = [0, 1]$ with its natural Euclidean metric $\rho(x, y) = |x - y|$ and let μ be the counting measure

[7] The inclusion '$\subset$' is obvious, for '$\supset$' fix $x \in U$. Then there exists some $r \in \mathbb{Q}^+$ with $B_r(x) \subset U$. Since D is dense, $x \in B_{r/2}(d)$ for some $d \in D$ with $\rho(d, x) < r/4$, so that $x \in B_{r/2}(d) \subset U$.

on $([0,1],\mathscr{B}[0,1])$, i.e. $\mu(B):=\#B$. Obviously, μ is not σ-finite. The pth-power μ-integrable *simple* functions are of the form

$$\mathcal{E}(\mathscr{B})\cap L^p(\mu)=\left\{\sum_{n=1}^{N} y_n \mathbb{1}_{A_n} : N\in\mathbb{N},\ y_n\in\mathbb{R},\ A_n\in\mathscr{B},\ \#A_n<\infty\right\},$$

so that

$$L^p(\mu)=\Big\{u:[0,1]\to\mathbb{R}:\exists(x_n)_{n\in\mathbb{N}}\subset[0,1],\forall x\neq x_n\quad u(x)=0 \\ \text{and}\quad \sum_{n=1}^{\infty}|u(x_n)|^p<\infty\Big\}.$$

Obviously, $(\mathbb{1}_{\{x\}})_{x\in[0,1]}\subset L^p(\mu)$, but no single countable system can approximate this family since

$$\|\mathbb{1}_{\{x\}}-\mathbb{1}_{\{y\}}\|_p^p=0\quad\text{or}\quad 2,$$

according to whether $x=y$ or $x\neq y$. □

With somewhat more effort we can show that the conditions of Lemma 27.23 are even necessary.

Theorem 27.25 *Let $(X,\mathscr{A},\mu)$ be a σ-finite measure space. Then the following assertions are equivalent.*[8]

(i) *$\mathscr{A}$ is (almost) separable, i.e. there exists a countable family $\mathscr{F}\subset\mathscr{A}$ such that $\mu(F)<\infty$ for all $F\in\mathscr{F}$ and $\sigma(\mathscr{F})\approx\mathscr{A}$ in the sense that every set in $\mathscr{A}$ has, up to a null set, a version in $\sigma(\mathscr{F})$.*
(ii) *μ is separable,*[9] *i.e. there exists a countable family $\mathscr{F}\subset\mathscr{A}$ with $\mu(F)<\infty$ and for every $A\in\mathscr{A}$ with $\mu(A)<\infty$ we have*

$$\forall\epsilon>0\quad\exists F_\epsilon\in\mathscr{F}\ :\ \mu(A\setminus F_\epsilon)+\mu(F_\epsilon\setminus A)\leqslant\epsilon.$$

(iii) *$L^p(X,\mathscr{A},\mu)$ is separable, $1\leqslant p<\infty$.*

Proof (i)$\Rightarrow$(iii). The proof of Lemma 27.23 shows that $L^p(X,\sigma(\mathscr{F}),\mu)$ is separable. Since for each $A\in\mathscr{A}$ there is an $A^*\in\sigma(\mathscr{F})$ with

$$\mu\big((A\setminus A^*)\cup(A^*\setminus A)\big)=0\iff\int|\mathbb{1}_A-\mathbb{1}_{A^*}|\,d\mu=0,$$

[8] Compare this result with Problem 5.12.
[9] This notion derives from the fact that $(\mathscr{A},\rho_\mu)$, $\rho_\mu(A,B):=\mu(A\setminus B)+\mu(B\setminus A)$, $A,B\in\mathscr{A}$ becomes a separable pseudo-metric space in the usual sense, see Problem 22.8.

every function $\phi \in \mathcal{E}(\mathscr{A})$ has a version $\phi^* \in \mathcal{E}(\sigma(\mathscr{F}))$ such that $\int |\phi - \phi^*| d\mu = 0$. This proves that $L^p(X, \sigma(\mathscr{F}), \mu) \supset L^p(X, \mathscr{A}, \mu)$ (we have, in fact, equality since $\sigma(\mathscr{F}) \subset \mathscr{A}$), and we see that $L^p(X, \mathscr{A}, \mu)$ is separable.

(iii) $\Rightarrow$ (ii). Let $(d_n)_{n\in\mathbb{N}} \subset L^p(\mathscr{A})$ be a countable dense set. Since $\mathcal{E}(\mathscr{A}) \cap L^p(\mathscr{A})$ is dense in $L^p(\mathscr{A})$, see Lemma 17.2 or combine Corollary 8.9 and Theorem 13.10, we find for each d_n a sequence $(f_{nk})_{k\in\mathbb{N}} \subset \mathcal{E}(\mathscr{A}) \cap L^p(\mathscr{A})$ such that $L^p\text{-}\lim_{k\to\infty} f_{nk} = d_n$. Thus $(f_{nk})_{k,n\in\mathbb{N}}$ is also dense in $L^p(\mathscr{A})$, and the system of subsets

$$\mathscr{F} := \left\{ \bigcup_{\ell=1}^{N} \{f_{nk} = r_\ell\} : N \in \mathbb{N},\ k, n \in \mathbb{N},\ r_\ell \in \mathbb{R} \right\}$$

is countable since each f_{nk} attains only finitely many values.

For every $A \in \mathscr{A}$, $\mu(A) < \infty$, we have $\mathbb{1}_A \in L^p(\mathscr{A})$, and we find a subsequence $(f_\ell^A)_{\ell\in\mathbb{N}} \subset (f_{nk})_{k,n\in\mathbb{N}}$ with $\lim_{\ell\to\infty} \|f_\ell^A - \mathbb{1}_A\|_p^p = 0$.

Set $F_\ell := \{|f_\ell^A - \mathbb{1}_A| \leqslant 1/2\} \cap \{|f_\ell^A| > 1/2\}$. Obviously $F_\ell \in \mathscr{F}$, and $F_\ell \subset A$ since

$$\begin{aligned} A^c \cap F_\ell &= A^c \cap \{|f_\ell^A - \mathbb{1}_A| \leqslant 1/2\} \cap \{|f_\ell^A| > 1/2\} \\ &= A^c \cap \{|f_\ell^A| \leqslant 1/2\} \cap \{|f_\ell^A| > 1/2\} \\ &= \emptyset. \end{aligned}$$

Thus $\mu(F_\ell \setminus A) = 0$, while

$$\mu(A \setminus F_\ell) \leqslant \mu(A \cap \{|f_\ell^A - \mathbb{1}_A| > 1/2\}) + \mu(A \cap \{|f_\ell^A| \leqslant 1/2\}).$$

Using the triangle inequality, we infer

$$A \cap \{|f_\ell^A| \leqslant 1/2\} \subset A \cap \{|f_\ell^A - 1| \geqslant 1/2\} = A \cap \{|f_\ell^A - \mathbb{1}_A| \geqslant 1/2\},$$

and with the above calculation and an application of Markov's inequality we conclude that

$$\begin{aligned} \mu(A \setminus F_\ell) + \mu(F_\ell \setminus A) &\leqslant 2\mu(A \cap \{|f_\ell^A - \mathbb{1}_A| \geqslant 1/2\}) \\ &\overset{(11.4)}{\leqslant} 2^{p+1} \|f_\ell^A - \mathbb{1}_A\|_p^p. \end{aligned}$$

The right-hand side of the above inequality tends to 0 as $\ell \to \infty$, and (ii) follows.

(ii) $\Rightarrow$ (i). Fix $A \in \mathscr{A}$ with $\mu(A) < \infty$. Then we find, by assumption, sets $F_n \in \mathscr{F}$ with $\mu(A \setminus F_n) + \mu(F_n \setminus A) \leqslant 2^{-n}$. Consider the sets

$$F^* := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} F_n \quad \text{and} \quad F_* := \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} F_n.$$

Then using the continuity and σ-subadditivity of measures (Theorem 4.3(vi)–(viii)),

$$\begin{aligned}\mu(F^*\setminus A)+\mu(A\setminus F_*) &= \mu\bigg(\bigg[\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}F_n\bigg]\bigg\backslash A\bigg)+\mu\bigg(A\cap\bigg[\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}F_n\bigg]^c\bigg)\\ &= \mu\bigg(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}(F_n\setminus A)\bigg)+\mu\bigg(A\cap\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}F_n^c\bigg)\\ &= \lim_{k\to\infty}\bigg[\mu\bigg(\bigcup_{n=k}^{\infty}(F_n\setminus A)\bigg)+\mu\bigg(\bigcup_{n=k}^{\infty}(A\setminus F_n)\bigg)\bigg]\\ &\leqslant \lim_{k\to\infty}\sum_{n=k}^{\infty}\big[\mu\big(F_n\setminus A\big)+\mu\big(A\setminus F_n\big)\big]\\ &\leqslant \lim_{k\to\infty}\sum_{n=k}^{\infty}2^{-n}=0.\end{aligned}$$

This shows that for all $A\in\mathscr{A}$ with $\mu(A)<\infty$

$$\exists F_*,F^*\in\sigma(\mathscr{F}):F_*\subset F^*\quad\text{and}\quad\mu(F^*\setminus A)+\mu(A\setminus F_*)=0,$$

implying that $\mu(F^*\setminus A)+\mu(A\setminus F^*)=0$, too.

If $\mu(A)=\infty$ we pick some exhausting sequence $(A_k)_{k\in\mathbb{N}}\subset\mathscr{A}$ with $A_k\uparrow X$ and $\mu(A_k)<\infty$. Then the sets $A\cap A_k$ have finite μ-measure and we can construct, as before, sets F_k^* and $F_{*,k}$. Setting $F^*:=\bigcup_{k\in\mathbb{N}}F_k^*$, we find

$$\bigg(\bigcup_{k\in\mathbb{N}}F_k^*\bigg)\bigg\backslash\bigcup_{n\in\mathbb{N}}(A\cap A_n)=\bigcup_{k\in\mathbb{N}}\bigg(F_k^*\bigg\backslash\bigcup_{n\in\mathbb{N}}(A\cap A_n)\bigg)\subset\bigcup_{k\in\mathbb{N}}(F_k^*\setminus(A\cap A_k)),$$

and so

$$\mu(F^*\setminus A)\leqslant\mu\bigg(\bigcup_{k\in\mathbb{N}}\big(F_k^*\setminus(A\cap A_k)\big)\bigg)\leqslant\sum_{k=1}^{\infty}\underbrace{\mu\big(F_k^*\setminus(A\cap A_k)\big)}_{=0}=0.$$

The expression $\mu(A\setminus F^*)$ is handled analogously.

This shows that sets from $\mathscr{A}$ and $\sigma(\mathscr{F})$ differ by at most a null set. $\square$

Problems

27.1. Let $(X,\mathscr{A},\mu)$ be a measure space and $\mathscr{H}\subset\mathscr{G}$ be two sub-σ-algebras. Show that

$$\mathbb{E}^{\mathscr{H}}\mathbb{E}^{\mathscr{G}}u=\mathbb{E}^{\mathscr{G}}\mathbb{E}^{\mathscr{H}}u=\mathbb{E}^{\mathscr{H}}u\qquad\forall u\in L^2(\mathscr{A}).$$

27.2. Let $u \in \mathcal{L}^2(\mu)$ be a positive, integrable function. Show that $\int_{\{u>1\}} u\,d\mu \leqslant \mu\{u>1\}$ entails that $u(x) \leqslant 1$ for μ-a.e. x.

Remark. This argument is needed for the second part of the proof of Lemma 27.4(x).

27.3. Let $(X, \mathscr{A}, \mu)$ be a measure space, $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra and let $\nu := f\mu$ where $f \in \mathcal{M}^+(\mathscr{A})$ is a density $f > 0$. Denote by $\mathbb{E}_\nu^{\mathscr{G}}$, resp. $\mathbb{E}_\mu^{\mathscr{G}}$, the projections in the spaces $L^2(\mathscr{A}, \mu)$, resp. $L^2(\mathscr{A}, \nu)$.

(i) Let $G^* := \{\mathbb{E}_\mu^{\mathscr{G}} f > 0\}$ and show that $\nu|_{G^* \cap \mathscr{G}} = (\mathbb{E}_\mu^{\mathscr{G}} f)\mu|_{G^* \cap \mathscr{G}}$ ($G^* \cap \mathscr{G}$ denotes the trace σ-algebra).

(ii) Show that $Pu := \mathbb{1}_{G^*} \mathbb{E}_\mu^{\mathscr{G}}(fu)/\mathbb{E}_\mu^{\mathscr{G}} f$ maps bounded $L^2(\mathscr{A}, \nu)$-functions into $L^2(\mathscr{G}, \nu)$ and satisfies $\|Pu\|_{L^2(\mathscr{G}, \nu)} \leqslant \|u\|_{L^2(\mathscr{A}, \mu)}$.

(iii) Show that $P = \mathbb{E}_\nu^{\mathscr{G}}$.

(iv) When do we have $\mathbb{E}_\mu^{\mathscr{G}} u = \mathbb{E}_\nu^{\mathscr{G}} u$ for all $u \in L^2(\mathscr{A}, \mu) \cap L^2(\mathscr{A}, \nu)$?

Remark. The above result allows us to study conditional expectations for finite measures μ only and to *define* for more measures a conditional expectation by

$$\mathbb{E}_\nu^{\mathscr{G}} u := \frac{\mathbb{E}_\mu^{\mathscr{G}}(fu)}{\mathbb{E}_\mu^{\mathscr{G}} f} \mathbb{1}_{\{\mathbb{E}_\mu^{\mathscr{G}} f > 0\}}.$$

27.4. Let $(X, \mathscr{A}, \mu)$ be a finite measure space, $G_1, \ldots, G_n \in \mathscr{A}$ such that $\dot{\bigcup}_{i=1}^n G_i = X$ and $\mu(G_i) > 0$ for all $i = 1, 2, \ldots, n$. Then

$$\mathbb{E}^{\mathscr{G}} u = \sum_{i=1}^n \left[\int_{G_i} u(x) \frac{\mu(dx)}{\mu(G_i)} \right] \mathbb{1}_{G_i}.$$

Remark. The measure $\mathbb{1}_{G_i}\mu/\mu(G_i) = \mu(\cdot \cap G_i)/\mu(G_i)$ is often called the *conditional probability* given G_i.

27.5. **Extension by continuity.** Let $T: L^2(\mu) \to L^2(\mu)$ be a linear operator such that $\|Tu\|_p \leqslant c\|u\|_p$ for $u \in L^2(\mu) \cap L^p(\mu)$ for some $p \neq 2$. Show that there is a unique extension $\tilde{T}: L^p(\mu) \to L^p(\mu)$ defined by L^p-$\lim_{n\to\infty} Tu_n$ for any sequence $u_n \in L^2(\mu) \cap L^p(\mu)$ with $u_n \to u$ in $L^p(\mu)$. If T is monotone, i.e. $u \leqslant w \implies Tu \leqslant Tw$, then $u_n \uparrow u$ a.e. implies that $Tu_n \uparrow Tu$ a.e.

[Hint: have a look at the remark preceding Theorem 27.5.]

27.6. Let $(X, \mathscr{A}, \mu)$ be a measure space and let $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra such that $\mu|_{\mathscr{G}}$ is σ-finite. Let $p, q \in (1, \infty)$ be conjugate numbers, i.e. $1/p + 1/q = 1$, and let $u, w \in \bigcap_{p \in [1,\infty]} L^p(\mathscr{A}, \mu)$. Show that

$$\left|\mathbb{E}^{\mathscr{G}}(uw)\right| \leqslant \left[\mathbb{E}^{\mathscr{G}}(|u|^p)\right]^{1/p} \left[\mathbb{E}^{\mathscr{G}}(|w|^q)\right]^{1/q}.$$

[Hint: use Lemma 13.1 with $A = |u|/\left[\mathbb{E}^{\mathscr{G}}(|u|^p)\right]^{1/p}$ and $B = |w|/\left[\mathbb{E}^{\mathscr{G}}(|w|^q)\right]^{1/q}$ whenever the numerator is not 0.]

27.7. Complete the proof of Theorem 27.11.

27.8. Show that $\mathbb{E}^{\mathscr{G}} 1 = 1$ if, and only if, $\mu|_{\mathscr{G}}$ is σ-finite. Find a counterexample showing that $\mathbb{E}^{\mathscr{G}} 1 \leqslant 1$ is, in general, the best possible.

[Hint: use $p = 2$ and $\mathbb{E}^{\mathscr{G}} = \mathbb{E}^{\mathscr{G}}$.]

27.9. Let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{A}$. Show that $\mathbb{E}^{\mathscr{G}} g = g$ for all $g \in L^p(\mathscr{G})$.

[Hint: observe that, a.e., $g = g\mathbb{1}_{\bigcup_n \{|g|>1/n\}}$ and $\mu\{|g| > 1/n\} < \infty$. This emulates σ-finiteness.]

27.10. Let $\mathscr{H} \subset \mathscr{G}$ be two sub-σ-algebras of $\mathscr{A}$. Show that

$$\mathbb{E}^{\mathscr{G}}\mathbb{E}^{\mathscr{H}}u = \mathbb{E}^{\mathscr{H}}\mathbb{E}^{\mathscr{G}}u = \mathbb{E}^{\mathscr{H}}u$$

for all $u \in L^p(\mathscr{A})$, resp. for all $u \in M^+(\mathscr{A})$, provided that $\mu|_{\mathscr{H}}$ is σ-finite.
[Hint: if $\mu|_{\mathscr{H}}$ is not σ-finite, the set $L^p(\mathscr{H})$ can be *very* small …]

27.11. Consider on the measure space $([0,\infty), \mathscr{B}[0,\infty), \lambda^1|_{[0,\infty)})$ the filtration defined by $\mathscr{A}_n := \sigma\big([0,1), [0,2), \dots, [n-1,n), [n,\infty)\big)$. Find $\mathbb{E}^{\mathscr{A}_n}u$ for $u \in L^p$.

27.12. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and $\mathscr{G} \subset \mathscr{A}$ be a sub-σ-algebra. Show that, in general,

$$\int \mathbb{E}^{\mathscr{G}}u\,d\mu \leqslant \int u\,d\mu, \qquad u \in L^1(\mathscr{A}),\ u \geqslant 0,$$

with equality holding only if $\mu|_{\mathscr{G}}$ is σ-finite.

27.13. Prove Corollaries 27.14 and 27.15.

27.14. Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space and denote by $\langle u, \phi\rangle$ the canonical dual pairing between $u \in L^p$ and $\phi \in L^q$, where $p^{-1} + q^{-1} = 1$, namely $\langle u, \phi\rangle := \int u\phi\,d\mu$. A sequence $(u_n)_{n\in\mathbb{N}} \subset L^p$ is *weakly relatively compact* if there exists a subsequence $(u_{n_k})_{k\in\mathbb{N}}$ such that

$$\langle u_{n_k} - u, \phi\rangle \xrightarrow[k\to\infty]{} 0$$

holds for all $\phi \in L^q$ and some $u \in L^p$. Show that for a martingale $(u_n)_{n\in\mathbb{N}}$ and every $p \in (1,\infty)$ the following assertions are equivalent:

(i) there exists some $u \in L^p(\mathscr{A})$ such that $\lim_{n\to\infty} \|u_n - u\|_p = 0$;
(ii) there exists some $u_\infty \in L^p(\mathscr{A}_\infty)$ such that $u_n = \mathbb{E}^{\mathscr{A}_n}u_\infty$;
(iii) the sequence $(u_n)_{n\in\mathbb{N}}$ is weakly relatively compact.

27.15. Let $(X, \mathscr{A}, \mu)$ be a measure space and $(u_n)_{n\in\mathbb{N}} \subset L^1(\mathscr{A})$. Show that

$$m_1 := u_1, \qquad m_{n+1} - m_n := u_{n+1} - \mathbb{E}^{\mathscr{A}_n}u_{n+1}$$

is a martingale under the filtration $\mathscr{A}_n := \sigma(u_1, \dots, u_n)$.

27.16. (Continuation of Problem 27.15). If $\int u_1\,d\mu = 0$ and $\mathbb{E}^{\mathscr{A}_n}u_{n+1} = 0$, then $(u_n)_{n\in\mathbb{N}}$ is called a *martingale difference sequence*. Assume that $u_n \in L^2(\mathscr{A})$ and denote by $s_k := u_1 + \cdots + u_k$ the partial sums. Show that $(s_n^2, \mathscr{A}_n)_{n\in\mathbb{N}}$ is a submartingale satisfying

$$\int s_k^2\,d\mu = \sum_{n=1}^{k} \int u_n^2\,d\mu.$$

27.17. **Doob decomposition.** Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a σ-finite filtered measure space and let $(u_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ be a submartingale. Define $u_0 := 0$ and $\mathscr{A}_0 := \{\emptyset, X\}$. Show that there exists an a.e. unique martingale $(m_n, \mathscr{A}_n)_{n\in\mathbb{N}}$ and an increasing sequence of functions $(a_n)_{n\in\mathbb{N}}$ such that $a_n \in L^1(\mathscr{A}_{n-1})$ for all $n \geqslant 2$ and

$$u_n = m_n + a_n, \qquad n \in \mathbb{N}.$$

[Hint: set $m_0 := u_0$, $m_{n+1} - m_n := u_{n+1} - \mathbb{E}^{\mathscr{A}_n}u_{n+1}$ and $a_0 := 0$, $a_{n+1} - a_n := \mathbb{E}^{\mathscr{A}_n}u_{n+1} - u_n$. For uniqueness assume $\tilde{m}_n + \tilde{a}_n$ is a further Doob decomposition and study the measurability properties of the martingale $M_n := m_n - \tilde{m}_n = \tilde{a}_n - a_n$.]

27.18. Let $(\Omega, \mathscr{A}, P)$ be a probability space and let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent identically distributed random variables such that $P(X_n = 0) = P(X_n = 2) = \frac{1}{2}$. Set $M_k := \prod_{n=1}^{k} X_n$. Show that there exists no any filtration $(\mathscr{A}_n)_{n\in\mathbb{N}}$ and no random variable M such that $M_k = \mathbb{E}^{\mathscr{A}_k} M$.

[Hint: compare this with Example 23.3(xi).]

Remark. This example shows that not all martingales can be obtained as conditional expectations of a single function.

27.19. Let $(X, \mathscr{A}, \mathscr{A}_n, \mu)$ be a finite filtered measure space. Let $u_n \in L^1(\mathscr{A}_n)$ be a sequence of measurable functions such that $|u_n| \leqslant \mathbb{E}^{\mathscr{A}_n} f$ for some $f \in L^1(\mathscr{A})$. Show that the family $(u_n)_{n\in\mathbb{N}}$ is uniformly integrable.

28

Orthonormal Systems and Their Convergence Behaviour

In Chapter 26 we discussed the importance of orthonormal systems (ONSs) in Hilbert spaces. In particular, countable complete ONSs turned out to be bases of separable Hilbert spaces. We have also seen that a countable ONS gives rise to a family of finite-dimensional subspaces and a sequence of orthogonal projections onto these spaces. In this chapter we want to

- give concrete examples of (complete) ONSs;
- see when the associated canonical projections are conditional expectations;
- understand the L^p $(p \neq 2)$ and a.e. convergence behaviour of series expansions with respect to certain ONSs.

The third item in the list above is, in general, not a trivial matter. Here we will see how we can use the powerful martingale machinery of Chapters 23 and 24 to get L^p $(1 \leqslant p < \infty)$ and a.e. convergence.

Throughout this chapter we consider the Hilbert space $L^2(I, \mathscr{B}(I), \rho\lambda)$, where $I \subset \mathbb{R}$ is a finite or infinite interval of the real line, $\mathscr{B}(I) = I \cap \mathscr{B}(\mathbb{R})$ are the Borel sets in I, $\lambda = \lambda^1|_I$ is Lebesgue measure on I and $\rho(x)$ is a density function. We will usually write $\rho(x)dx$ and $\int \ldots dx$ instead of $\rho\lambda$ and $\int \ldots d\lambda$.

One of the most important techniques to construct ONSs is the Gram–Schmidt orthonormalization procedure (26.16), which we can use to turn any countable family $(f_n)_{n\in\mathbb{N}}$ into an orthonormal sequence $(e_n)_{n\in\mathbb{N}}$. It is something of a problem, however, to find a reasonable sequence $(f_n)_{n\in\mathbb{N}}$ which can be used as input to the orthonormalization procedure.

Orthogonal Polynomials

For many practical applications, such as interpolation, approximation or numerical integration, a natural set of f_n to begin with is given by the polynomials on I. Usually one applies (26.16) to the sequence of *monomials*

$$(1, t, t^2, t^3, \dots) = (t^n)_{n\in\mathbb{N}_0}$$

to construct an ONS consisting of polynomials. Of course, this depends heavily on the underlying measure space, where polynomials should be square integrable. With some (partly pretty tedious) calculations[1] one can get the following important classes of orthogonal polynomials in $L^2(I, \mathscr{B}(I), \rho(x)dx)$.

Jacobi polynomials $\big(J_n^{(\alpha,\beta)}\big)_{n\in\mathbb{N}_0}$, $\alpha, \beta > -1$ 28.1 We choose

$$I = [-1, 1], \quad \rho(x)dx = (1-x)^\alpha(1+x)^\beta\, dx, \quad \alpha, \beta > -1,$$

and we get

$$\begin{aligned} J_n^{(\alpha,\beta)}(x) &= \frac{(-1)^n}{n!2^n} \frac{\dfrac{d^n}{dx^n}\Big((1-x)^{\alpha+n}(1+x)^{\beta+n}\Big)}{(1-x)^\alpha(1+x)^\beta} \\ &= \frac{1}{2^n}\sum_{i=0}^n \binom{n+\alpha}{i}\binom{n+\beta}{n-i}(x-1)^{n-i}(x+1)^i \\ \|J_n^{(\alpha,\beta)}\|_2^2 &= \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+1)\Gamma(n+\alpha+\beta+1)}. \end{aligned}$$

Choosing particular values for α and β yields other important families.

Chebyshev polynomials (of the first kind) $(T_n)_{n\in\mathbb{N}_0}$ 28.2 We choose

$$I = [-1, 1], \quad \rho(x)dx = (1-x^2)^{-1/2}\, dx,$$

and we get

$$\begin{aligned} T_n(x) &= J_n^{(-1/2,-1/2)}(x) = \begin{cases} \sqrt{\dfrac{2}{\pi}}\cos(n \arccos x) & \text{if } n \in \mathbb{N}, \\ \dfrac{1}{\sqrt{\pi}} & \text{if } n = 0, \end{cases} \\ \|T_n\|_2^2 &= \frac{1}{2}\left(\frac{\Gamma\big(n+\frac{1}{2}\big)}{\Gamma(n+1)}\right)^2. \end{aligned}$$

The first few Chebyshev polynomials are

$$1,\ x,\ 2x^2-1,\ 4x^3-3x,\ 8x^4-8x^2+1,\ 16x^5-20x^3+5x, \dots$$

[1] The material in items 28.1–28.5 below is taken from Alexits [1, pp. 30–37], the NIST handbook [33, Section 18] and Kaczmarz and Steinhaus [24, Sections IV.1–2, 8–9].

and the following recursion formula holds:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n \in \mathbb{N}.$$

Legendre polynomials $(P_n)_{n\in\mathbb{N}_0}$ **28.3** We choose

$$I = [-1, 1], \quad \rho(x)dx = dx,$$

and we get

$$P_n(x) = J_n^{(0,0)}(x) = \frac{(-1)^n}{n!2^n} \frac{d^n}{dx^n}(1 - x^2)^n, \quad \|P_n\|_2^2 = \frac{2}{2n+1}.$$

The first few Legendre polynomials are

$1, \ x, \ \frac{1}{2}(3x^2 - 1), \ \frac{1}{2}(5x^3 - 3x), \ \frac{1}{8}(35x^4 - 30x^2 + 3), \ \frac{1}{8}(63x^5 - 70x^3 + 15x)$

and the following recursion formula holds:

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x), \quad n \in \mathbb{N}.$$

Laguerre polynomials $(L_n)_{n\in\mathbb{N}_0}$ **28.4** We choose

$$I = [0, \infty), \quad \rho(x)dx = e^{-x}dx,$$

and we get

$$L_n(x) = e^x \frac{d^n}{dx^n}\left(e^{-x}x^n\right) = n! \sum_{i=0}^{n} (-1)^i \binom{n}{i} \frac{x^i}{i!}, \quad \|L_n\|_2^2 = (n!)^2.$$

The first few Laguerre polynomials are

$1, \ 1 - x, \ x^2 - 4x + 2, \ -x^3 + 9x^2 - 18x + 6, \ x^4 - 16x^3 + 72x^2 - 96x + 24, \ldots$

and the following recursion formula holds:

$$L_{n+1}(x) = (2n + 1 - x)L_n(x) - n^2 L_{n-1}(x), \quad n \in \mathbb{N}.$$

Hermite polynomials $(H_n)_{n\in\mathbb{N}_0}$ **28.5** We choose

$$I = (-\infty, \infty), \quad \rho(x)dx = e^{-x^2}\,dx,$$

and we get

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n}\left(e^{-x^2}\right), \quad \|H_n\|_2^2 = 2^n n! \sqrt{\pi}.$$

The first few Hermite polynomials are

$$1, \quad 2x, \quad 4x^2 - 2, \quad 8x^3 - 12x, \quad 16x^4 - 48x^2 + 12, \ldots$$

and the following recursion formula holds:

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x), \quad n \in \mathbb{N}.$$

In order to decide if a family of polynomials $(p_k)_{k\in\mathbb{N}} \subset L^2(I, \rho(x)dx)$ is a complete ONS we have to show that

$$\int u(x)p_k(x)\rho(x)dx = 0 \quad \forall k \in \mathbb{N}_0 \Longrightarrow u = 0 \text{ a.e.}$$

The key technical result is the *Weierstraß approximation theorem*.

Theorem 28.6 (Weierstraß) *Polynomials are dense in* $C[0,1]$ *w.r.t. uniform convergence.*

Proof (S. N. Bernstein) Take a sequence $(\xi_i)_{i\in\mathbb{N}}$ of independent[2] measurable functions on $([0,1], \mathscr{B}[0,1], dx)$ which are Bernoulli $(p, 1-p)$-distributed, $0 < p < 1$:

$$\lambda\{\xi_i = 1\} = p \quad \text{and} \quad \lambda\{\xi_i = 0\} = 1 - p \qquad \forall i \in \mathbb{N},$$

see Scholium 23.4 for details of the construction of such a sequence. We then write $S_n := \xi_1 + \cdots + \xi_n$ for the partial sum and observe that, due to independence,

$$\begin{aligned}\lambda\{S_n = k\} &= \lambda\Big(\dot{\bigcup}_{1\leqslant i_1\leqslant\cdots\leqslant i_k\leqslant n}\Big(\{\xi_{i_1} = 1\} \cap \cdots \cap \{\xi_{i_k} = 1\} \\ &\qquad\qquad \cap \{\xi_{i_{k+1}} = 0\} \cap \cdots \cap \{\xi_{i_n} = 0\}\Big)\Big) \\ &= \binom{n}{k} p^k (1-p)^{n-k},\end{aligned}$$

which shows that

$$\int u\left(\frac{1}{n}S_n(x)\right)dx = \sum_{k=0}^{n} u\left(\frac{k}{n}\right)\binom{n}{k}p^k(1-p)^{n-k} =: B_n(u;p),$$

where $B_n(u;p)$ stands for the nth *Bernstein polynomial*.[3] From Scholium 23.4 we know that

$$\int \left|\frac{1}{n}S_n(x) - p\right|^2 dx = \frac{p(1-p)}{n} \leqslant \frac{1}{4n}, \tag{28.1}$$

[2] In the sense of Example 23.3(x).

[3] In view of the strong law of large numbers, Example 24.8, we observe that $n^{-1}S_n \to p$ a.e., so that by dominated convergence $B_n(u,p) \to u(p)$ for each $p \in (0,1)$. Since our argument includes this result as a particular case, we leave it as a side-remark.

since the function $p \mapsto p(1-p)$ attains its maximum at $p = 1/2$. As $u \in C[0,1]$ is uniformly continuous, $|u(x) - u(y)| < \epsilon$ whenever $|x-y| < \delta$ is small enough. So

$$
\begin{aligned}
&|B_n(u;p) - u(p)| \\
&\quad \leqslant \int \left|u\left(\frac{1}{n}S_n\right) - u(p)\right| d\lambda \\
&\quad = \int_{\{|\frac{1}{n}S_n - p| < \delta\}} \left|u\left(\frac{1}{n}S_n\right) - u(p)\right| d\lambda + \int_{\{|\frac{1}{n}S_n - p| \geqslant \delta\}} \left|u\left(\frac{1}{n}S_n\right) - u(p)\right| d\lambda \\
&\quad \leqslant \epsilon\lambda\left\{\left|\frac{1}{n}S_n - p\right| < \delta\right\} + 2\|u\|_\infty \lambda\left\{\left|\frac{1}{n}S_n - p\right| \geqslant \delta\right\} \\
&\quad \overset{(11.4)}{\leqslant} \epsilon + 2\|u\|_\infty \frac{1}{\delta^2}\int \left|\frac{1}{n}S_n - p\right|^2 d\lambda \\
&\quad \overset{(28.1)}{\leqslant} \epsilon + \frac{\|u\|_\infty}{2n\delta^2},
\end{aligned}
$$

by Markov's inequality and (28.1). This inequality is independent of $p \in [0,1]$, and the assertion follows by letting first $n \to \infty$ and then $\epsilon \to 0$. $\square$

Remark 28.7 The key ingredient in the above proof is (28.1), which shows that the variance of the random variable S_n vanishes uniformly (in p) as $n \to \infty$. A short calculation confirms that this is equivalent to saying that

$$B_n\big((\cdot - p)^2; p\big) \xrightarrow[n\to\infty]{} 0 \quad \text{uniformly for } p \in [0,1].$$

With this information, the proof then yields that $B_n(u;p) \to u(p)$ uniformly in p for all continuous u. This is, in fact, a special case of **Korovkin's theorem.** *A sequence of linear operators $T_n : C[0,1] \to C[0,1]$ which preserve positivity, i.e. $T_n u \geqslant 0$ if $u \geqslant 0$, converges uniformly for every $u \in C[0,1]$ if, and only if, it converges uniformly for each of the following three test functions:* $1,\ x,\ x^2$. *In the present case, $T_n u = B_n(u,\cdot)$. More on this topic can be found in Korovkin's monograph [25, pp. 1–30] or the expository paper [2] by Bauer.*

Corollary 28.8 *The monomials $(t^n)_{n\in\mathbb{N}_0}$ are complete in $L^1 = L^1([0,1], dt)$, that is, $\int_{[0,1]} u(t)t^n\,dt = 0$ for all $n \in \mathbb{N}_0$ implies that $u = 0$ a.e.*

Proof Assume first that $u \in C[0,1]$ satisfies $\int_{[0,1]} u(t)t^n\,dt = 0$ for all $n \in \mathbb{N}_0$. This implies, in particular, that

$$\int_{[0,1]} u(t)p(t)dt = 0 \ \text{ for all polynomials } \ p(t).$$

By Weierstraß' approximation theorem there is a sequence of polynomials $(p_n)_{n\in\mathbb{N}}$ which approximate u uniformly on $[0,1]$. Since $C[0,1]\subset L^2[0,1]\subset L^1[0,1]$, we see

$$\int_{[0,1]} u^2\,dt = \int_{[0,1]} u\cdot(u-p_n)dt \leqslant \|u\|_2\cdot\|u-p_n\|_2$$
$$\leqslant \|u\|_2\cdot\|u-p_n\|_\infty \xrightarrow[n\to\infty]{} 0,$$

and conclude that $u=0$ a.e. (even everywhere, since u is continuous).

Assume now that $u\in L^1([0,1],dt)\setminus C[0,1]$ such that $\int_{[0,1]} u(t)t^n\,dt=0$ for all $n\in\mathbb{N}_0$. The primitive

$$U(x):=\int_{(x,1]} u(t)dt$$

is a continuous function, see Problem 12.16, and by Fubini's theorem we see for all $n\in\mathbb{N}$

$$\begin{aligned}\int_{[0,1]} U(x)x^{n-1}\,dx &= \int_{[0,1]}\int_{[0,1]} \mathbb{1}_{(x,1]}(t)u(t)x^{n-1}\,dt\,dx\\ &= \int_{[0,1]}\left(\int_{[0,1]} \mathbb{1}_{[0,t)}(x)x^{n-1}\,dx\right)u(t)dt\\ &= \int_{[0,1]} \frac{t^n}{n!}u(t)dt \;=\; 0.\end{aligned}$$

This means that $\int_{[0,1]} U(x)x^n\,dx=0$ for all $n\in\mathbb{N}_0$ and, by the first part of the proof, that $U\equiv 0$. Lebesgue's differentiation theorem (Theorem 25.20) finally shows that $u(x)=U'(x)=0$ a.e. □

It is not hard to see that Theorem 28.6 and Corollary 28.8 also hold for the interval $[-1,1]$ and even for general compact intervals $[a,b]$ (see Problem 28.3). This we can use to show that the Jacobi (and hence Legendre and Chebyshev) polynomials are dense in $L^2(I,\rho(x)dx)$ and form a complete ONS. Note that

$$\begin{aligned}\int_{[-1,1]} |u\rho|dx &\leqslant \left(\int_{[-1,1]} \left(u\sqrt{\rho}\right)^2 dx\right)^{1/2}\cdot\left(\int_{[-1,1]} \left(\sqrt{\rho}\right)^2 dx\right)^{1/2}\\ &= \left(\int_{[-1,1]} u^2\rho\,dx\right)^{1/2}\cdot\left(\int_{[-1,1]} \rho\,dx\right)^{1/2} < \infty\end{aligned}$$

implies that $u\rho\in L^1([-1,1],dx)$, and from Corollary 28.8 and the fact that $\rho>0$ we get

$$\int_{[-1,1]} u(x)\rho(x)x^j\,dx=0 \implies u\rho=0 \text{ a.e.} \implies u=0 \text{ a.e.}$$

This does not quite work for the Hermite and Laguerre polynomials, which are defined on infinite intervals. For the latter we take $u \in L^2([0,\infty), e^{-x}\,dx)$, and find for all $s \geqslant 1$

$$\begin{aligned}\int_{[0,\infty)} u(x)e^{-sx}\,dx &= \int_{[0,\infty)} u(x)e^{(1-s)x}e^{-x}\,dx\\ &= \sum_{n=0}^{\infty}\frac{(1-s)^n}{n!}\underbrace{\int_{[0,\infty)} u(x)x^n e^{-x}\,dx}_{=0} = 0\end{aligned}$$

(note that the integral and the sum can be interchanged by dominated convergence). Using Jacobi's formula (Theorem 16.4) to change coordinates according to $t = e^{-x}$, $dt/dx = -e^{-x}$, we get

$$0 = \int_{[0,\infty)} u(x)e^{-sx}\,dx = \int_{[0,1)} u(-\ln t)t^{s-1}\,dt, \quad s \geqslant 1,$$

and for $s \in \mathbb{N}$ the above equality reduces to the case covered by Corollary 28.8. A very similar calculation can be used for the Hermite polynomials since

$$\begin{aligned}\int_{\mathbb{R}} u(x)e^{-sx^2}\,dx &= \int_{[0,\infty)} (u(x)+u(-x))e^{-sx^2}\,dx\\ &= \int_{[0,\infty)} \big(u(\sqrt{t}) + u(-\sqrt{t})\big)e^{-st}\,\frac{dt}{2\sqrt{t}},\end{aligned}$$

where we use the obvious substitution $x = \sqrt{t}$.

The Trigonometric System and Fourier Series

We consider now $L^2 = L^2((-\pi,\pi), \mathscr{B}(-\pi,\pi), \lambda = \lambda^1|_{(-\pi,\pi)})$. As before we use dx as a shorthand for $\lambda(dx)$. The *trigonometric system* consists of the functions

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{\cos x}{\sqrt{\pi}}, \quad \frac{\sin x}{\sqrt{\pi}}, \quad \frac{\cos 2x}{\sqrt{\pi}}, \quad \frac{\sin 2x}{\sqrt{\pi}}, \ldots, \frac{\cos nx}{\sqrt{\pi}}, \quad \frac{\sin nx}{\sqrt{\pi}}, \ldots \tag{28.2}$$

or, equivalently,

$$\frac{1}{\sqrt{2\pi}}e^{inx}, \quad n \in \mathbb{Z},\ i = \sqrt{-1}. \tag{28.3}$$

Since $e^{ix} = \cos x + i\sin x$, we can see that (28.2) and (28.3) are equivalent, and from now on we will consider only (28.2). Orthogonality of the functions in

(28.2) follows easily from the classical result that for $j,k\in\mathbb{N}_0$ and $\ell,m\in\mathbb{N}$

$$\begin{aligned}
&\int_{(-\pi,\pi)} \cos(jx)\cos(kx)dx = \begin{cases} 0, & \text{if } k\neq j,\\ \pi, & \text{if } k=j\geqslant 1,\\ 2\pi, & \text{if } k=j=0,\end{cases}\\
&\int_{(-\pi,\pi)} \sin(\ell x)\sin(mx)dx = \begin{cases} 0, & \text{if } \ell\neq m,\\ \pi, & \text{if } \ell=m,\end{cases}\\
&\int_{(-\pi,\pi)} \cos(kx)\sin(\ell x)\,dx = 0 \quad \text{for all } k,\ell,
\end{aligned} \tag{28.4}$$

which we leave as an exercise for the reader, see Problem 28.4.

Definition 28.9 A *trigonometric polynomial* (*of order n*) is an expression of the form

$$T(x) = \alpha_0 + \sum_{k=1}^{n} \big(\alpha_k \cos(kx) + \beta_k \sin(kx)\big), \tag{28.5}$$

where $n\in\mathbb{N}_0$, $\alpha_k,\beta_k\in\mathbb{R}$ and $\alpha_n^2+\beta_n^2>0$.

It is not hard to see that the representation (28.5) of $T(x)$ is equivalent to

$$T(x) = \sum_{j,k=0}^{n} \gamma_{j,k}\cos^j x\sin^k x$$

with coefficients $\gamma_{j,k}\in\mathbb{R}$, see Problem 28.5. It is this way of writing $T(x)$ that justifies the name trigonometric *polynomial*.

Theorem 28.10 *The trigonometric system* (28.2) *is a complete orthonormal system in* $L^2 = L^2((-\pi,\pi),dx)$.

Proof We have to show that

$$\left.\begin{aligned} &\int_{(-\pi,\pi)} u(x)\cos(kx)dx = 0 \quad \forall k\in\mathbb{N}_0\\ &\int_{(-\pi,\pi)} u(x)\sin(\ell x)dx = 0 \quad \forall \ell\in\mathbb{N}\end{aligned}\right\} \Longrightarrow u=0 \quad \text{a.e.} \tag{28.6}$$

Assume first that u is continuous and that, contrary to (28.6), $u(x_0)=c\neq 0$ for some $x_0\in(-\pi,\pi)$. Without loss of generality we may assume that $c>0$. Since the trigonometric functions are 2π-periodic, we can extend u periodically onto the whole real line. Then $w(x) := c^{-1}u(x+x_0)$ is continuous around $x=0$, orthogonal on $(-\pi,\pi)$ to any of the functions in (28.2), [✍] and satisfies $w(0)=1$. As w is continuous, there is some $0<\delta<\pi$ such that

$$w(x)>\frac{1}{2}\quad\forall x\in(-\delta,\delta).$$

Consider the trigonometric polynomial (Fig. 28.1)

$$t(x)=1-\cos\delta+\cos x.$$

Obviously, $t(x)$ and all powers $t^N(x)$ are polynomials in $\cos x$. From de Moivre's formula

$$e^{inx}=(\cos x+i\sin x)^n$$

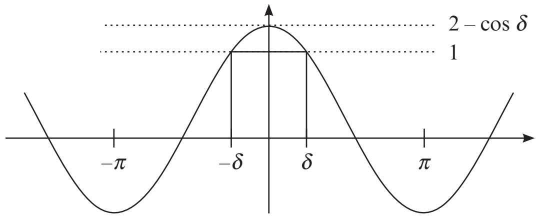


Fig. 28.1. A plot of the trigonometric polynomial $t(x)=1-\cos\delta+\cos x$.

it is easy to see that $\cos^n x=\sum_{k=0}^n c_k\cos(kx)$ ([✍], see also Gradshteyn and Ryzhik [19, 1.320]), i.e. we can write the $t^N(x)$ as linear combinations of $\cos(kx)$, $k=0,1,\dots,N$. By assumption, w is orthogonal to all of them, and so

$$0=\int_{(-\pi,\pi)}w(x)t^N(x)dx=\left(\int_{(-\pi,-\delta]}+\int_{(-\delta,\delta)}+\int_{[\delta,\pi)}\right)w(x)t^N(x)dx.\qquad(28.7)$$

On $(-\delta,\delta)$ we have $w(x)>\frac{1}{2}$ as well as $t(x)>1$, hence

$$\int_{(-\delta,\delta)}w(x)t^N(x)dx\geqslant\frac{1}{2}\int_{(-\delta,\delta)}t^N(x)dx\xrightarrow[N\to\infty]{}\infty$$

by monotone convergence (Theorem 9.6). On the other hand, we have $|t(x)|\leqslant 1$ for $x\in(-\pi,-\delta]\cup[\delta,\pi)$ and

$$\left|\int_{[\delta,\pi)}w(x)t^N(x)dx\right|\leqslant(\pi-\delta)\|w\|_\infty<\infty\quad\forall N\in\mathbb{N},$$

which means that (28.7) is impossible, i.e. $w\equiv 0$ and $u\equiv 0$.

An arbitrary function $u\in L^2((-\pi,\pi),dx)$ is, due to the finiteness of the measure, integrable, [✍] and we may consider the primitive

$$U(x):=\int_{(-\pi,x)}u(t)dt,$$

which is a continuous function, see Problem 12.16. Moreover,

$$U(-\pi)=0=\int_{(-\pi,\pi)}u(t)dt=U(\pi),$$

because of the assumption that u is orthogonal to *every* function from (28.2) and, in particular, to $t \mapsto 1/\sqrt{2\pi}$. By Fubini's theorem (Corollary 14.9) we get

$$\begin{aligned}
\int_{(-\pi,\pi)} U(x)\cos(kx)dx &= \int_{(-\pi,\pi)}\int_{(-\pi,\pi)} \mathbb{1}_{(-\pi,x)}(t)u(t)\cos(kx)dt\,dx \\
&= \int_{(-\pi,\pi)} \left(\int_{(-\pi,\pi)} \mathbb{1}_{(t,\pi)}(x)\cos(kx)dx\right)u(t)dt \\
&= -\int_{(-\pi,\pi)} u(t)\frac{\sin(kt)}{k}\,dt = 0,
\end{aligned}$$

and we conclude from the first part of the proof that $U \equiv 0$. Lebesgue's differentiation theorem (Theorem 25.20) finally shows that $u(x) = U'(x) = 0$ a.e. $\square$

Since the trigonometric system is one of the most important ONSs, we provide a further proof of the completeness theorem which gives some more insight into Fourier series and even yields a new proof of Weierstraß' approximation theorem (Theorem 28.6) for trigonometric polynomials, see Corollary 28.12 below.

We begin with an elementary but fundamental consideration which goes back to Féjer. If $u \in L^2((-\pi,\pi),dt)$, we write

$$a_k := \frac{1}{\pi}\int_{(-\pi,\pi)} u(t)\cos(kt)dt, \quad b_n := \frac{1}{\pi}\int_{(-\pi,\pi)} u(t)\sin(nt)dt \tag{28.8}$$

($k \in \mathbb{N}_0$, $n \in \mathbb{N}$) for the *Fourier cosine and sine coefficients* of u and set

$$\begin{aligned}
s_N(u;x) &:= \sum_{n=1}^{N}\big(a_n\cos(nx) + b_n\sin(nx)\big) + \frac{a_0}{2} \\
&= \frac{1}{\pi}\int_{(-\pi,\pi)}\left(\sum_{n=1}^{N}\big(\cos(nt)\cos(nx) + \sin(nt)\sin(nx)\big) + \frac{1}{2}\right)u(t)dt \\
&= \frac{1}{\pi}\int_{(-\pi,\pi)} u(t)\underbrace{\left(\frac{1}{2} + \sum_{n=1}^{N}\cos(n(t-x))\right)}_{=:\,D_N(t-x)}dt,
\end{aligned} \tag{28.9}$$

where we use the elementary trigonometric formula

$$\cos a\cos b + \sin a\sin b = \cos(a-b). \tag{28.10}$$

The function $D_N(\cdot)$ is called the *Dirichlet kernel*. In Problem 28.6 we will see that $D_N(\cdot)$ has the following closed-form expression:

$$D_N(x) = \frac{\sin\big(N+\frac{1}{2}\big)x}{2\sin(x/2)}, \tag{28.11}$$

but we do not need this formula in that which follows.

Now we introduce the *Cesàro C-1 mean*

$$\sigma_N(u;x) := \frac{1}{N+1}\big(s_0(u;x) + s_1(u;x) + \cdots + s_N(u;x)\big), \tag{28.12}$$

and in view of (28.9) we want to compute what is known as the *Féjer kernel*

$$K_N(x) := \frac{1}{N+1}\big(D_0(x) + D_1(x) + \cdots + D_N(x)\big).$$

Using again (28.10) and observing that the cosine is even we find that for every $n=0,1,\dots,N$,

$$\begin{aligned}(1-\cos x)D_n(x) &= \frac{1}{2}(1-\cos x)\sum_{k=-n}^{n}\cos(kx)\\ &\overset{(28.10)}{=} \frac{1}{2}\sum_{k=-n}^{n}\big(\cos(kx)-\cos((k-1)x)+\sin x\sin(kx)\big)\\ &= \frac{1}{2}\big(\cos(nx)-\cos((n+1)x)\big),\end{aligned}$$

since $\sin(kx) = -\sin(-kx)$ is an odd function which cancels if we sum over $-n \leqslant k \leqslant n$. Summing over all values of $n=0,1,\dots,N$ shows that

$$\begin{aligned}K_N(x) &= \frac{1}{N+1}\big(D_0(x)+D_1(x)+\cdots+D_N(x)\big)\\ &= \frac{1}{2(N+1)}\frac{1-\cos((N+1)x)}{1-\cos x}.\end{aligned} \tag{28.13}$$

Denote by $C_{\text{per}}[-\pi,\pi] = \{u \in C[-\pi,\pi] : u(\pi) = u(-\pi)\}$ the set of 2π-periodic continuous functions. If needed, we extend $u \in C_{\text{per}}[-\pi,\pi]$ periodically onto $\mathbb{R}$.

Lemma 28.11 (Féjer) *If* $u \in C_{\text{per}}[-\pi,\pi]$, *then* $\lim_{N\to\infty}\|\sigma_N(u)-u\|_p = 0$ *for all* $1 \leqslant p \leqslant \infty$.

Proof From (28.9), (28.12) and (28.13) we get after a change of variables in the integrals

$$\begin{aligned}\sigma_N(u;x) &= \frac{1}{\pi}\int_{[-\pi,\pi]} u(t)K_N(x-t)\,dt\\ &= \frac{1}{2(N+1)\pi}\int_{[-\pi,\pi]} u(x-t)\frac{1-\cos((N+1)t)}{1-\cos t}\,dt.\end{aligned}$$

Since $\frac{1}{\pi}\int_{[-\pi,\pi]} K_N(t)dt = 1$, [✍] we see for all $\epsilon > 0$ and sufficiently small $\delta > 0$

$$\begin{aligned}
\|\sigma_N(u) - u\|_p &= \left\| \frac{1}{2(N+1)\pi}\int_{[-\pi,\pi]} \frac{1-\cos((N+1)t)}{1-\cos t}(u(\cdot - t) - u)dt \right\|_p \\
&\overset{13.13}{\leqslant} \frac{1}{2(N+1)\pi}\int_{[-\pi,\pi]} \frac{1-\cos((N+1)t)}{1-\cos t}\|u(\cdot - t) - u\|_p\, dt \\
&\leqslant \frac{1}{2(N+1)\pi}\int_{(-\delta,\delta)} \frac{1-\cos((N+1)t)}{1-\cos t}\|u(\cdot - t) - u\|_p\, dt \\
&\quad + \frac{\|u\|_p}{(N+1)\pi}\int_{[-\pi,-\delta]\cup[\delta,\pi]} \frac{1-(\cos(N+1)t)}{1-\cos t}\, dt \\
&\leqslant \epsilon + \frac{\|u\|_p}{(N+1)\pi}\frac{4\pi}{1-\cos\delta},
\end{aligned}$$

where we use Jensen's inequality and the fact that $\lim_{t\to 0}\|u(\cdot - t) - u\|_p = 0$ by dominated convergence ($p < \infty$), resp. uniform continuity ($p = \infty$). Letting first $N \to \infty$ and then $\epsilon \to 0$ finishes the proof. □

Corollary 28.12 (Weierstraß) *The trigonometric polynomials are uniformly dense in* $C_{\mathrm{per}}[-\pi,\pi]$ *and dense in* $L^p([-\pi,\pi], dt)$ *w.r.t.* $\|\cdot\|_p$, $1 \leqslant p < \infty$.

Proof From (28.9), (28.12) it is obvious that $\sigma_N(u,\cdot)$ is a trigonometric polynomial. The density of the trigonometric polynomials in $C_{\mathrm{per}}[-\pi,\pi]$ is just Lemma 28.11. Since $C_{\mathrm{per}}[-\pi,\pi]$ is dense in $L^p([-\pi,\pi], dt)$, see Theorem 17.8, we can find for every $\epsilon > 0$ and $u \in L^p[-\pi,\pi]$ some $g_\epsilon \in C_{\mathrm{per}}[-\pi,\pi]$ with $\|u - g_\epsilon\|_p \leqslant \epsilon$ and a trigonometric polynomial t_ϵ such that $\|g_\epsilon - t_\epsilon\|_\infty \leqslant (2\pi)^{-1/p}\epsilon$. This shows that

$$\|u - t_\epsilon\|_p \leqslant \|u - g_\epsilon\|_p + \|g_\epsilon - t_\epsilon\|_p \leqslant \epsilon + (2\pi)^{1/p}\|g_\epsilon - t_\epsilon\|_\infty \leqslant 2\epsilon.$$

For the last estimate we also use that $\|w\|_p \leqslant (2\pi)^{1/p}\|w\|_\infty$. □

Corollary 28.13 *The trigonometric system* (28.2) *is a complete orthonormal system in* $L^2 = L^2([-\pi,\pi], dt)$.

Proof of Corollary 28.13 and, again, of Theorem 28.10 Let $u \in L^2[-\pi,\pi]$ and pick a trigonometric polynomial t_ϵ such that $\|u - t_\epsilon\|_2 \leqslant \epsilon$, see Corollary 28.12. Let $n = \mathrm{degree}(t_\epsilon)$. As in the proof of Theorem 28.10 we use de Moivre's formula to see that $\cos^n x$ and $\sin^n x$ can be represented as linear combinations of $1, \cos x, \dots, \cos(nx)$ and $\sin x, \dots, \sin(nx)$. [✍]

Recall that the partial sum $s_n(u;x) = a_0/2 + \sum_{k=1}^n (a_k\cos(kx) + b_k\sin(kx))$ is the projection of u onto $\mathrm{span}\{1, \cos x, \sin x, \dots, \cos(nx), \sin(nx)\}$. Therefore, Theorem 26.19(i) applies and we get completeness since

$$\|u - s_n(u)\|_2 \leqslant \|u - t_\epsilon\|_2 \leqslant \epsilon. \qquad \square$$

The above proof of the completeness of the trigonometric system has a further advantage as it allows a glimpse into other modes of convergence of Fourier series.

Corollary 28.14 (M. Riesz's theorem) *Let* $u \in L^p([-\pi, \pi], dt)$ *and* $1 \leqslant p < \infty$. *Then*

$$\lim_{n\to\infty} \|u - s_n(u)\|_p = 0 \iff \|s_n(u)\|_p \leqslant C_p \|u\|_p \quad \forall n \in \mathbb{N} \tag{28.14}$$

with an absolute constant C_p *not depending on* u *or* $n \in \mathbb{N}$.

Proof The 'only if' part is a consequence of the uniform boundedness principle (Banach–Steinhaus theorem) from functional analysis, see e.g. Rudin [43, Section 5.8] or Problem 26.16.

The 'if' part follows from the observation that every trigonometric polynomial T of degree $\leqslant n$ satisfies $s_n(T) = T$. [✍] Choosing for $u \in L^p$ the polynomial $T = t_\epsilon$ with $\|u - t_\epsilon\|_p \leqslant \epsilon$, see Corollary 28.12, we infer that for sufficiently large $n > \text{degree}(t_\epsilon)$

$$\begin{aligned} \|u - s_n(u)\|_p &\leqslant \|u - t_\epsilon\|_p + \underbrace{\|t_\epsilon - s_n(t_\epsilon)\|_p}_{=0} + \|s_n(t_\epsilon) - s_n(u)\|_p \\ &\leqslant (1 + C_p)\|u - t_\epsilon\|_p. \end{aligned} \qquad \square$$

Establishing the estimate $\|s_n(u)\|_p \leqslant C_p\|u\|_p$ is an altogether different matter and so is the whole L^p- and pointwise convergence theory for Fourier series. Here we want to mention only a few facts:

- L^p-convergence ($1 \leqslant p < \infty$) of the Cesàro means $\sigma_n(u)$ follows immediately from Lemma 28.11.

This is in stark contrast to the facts that

- L^p-convergence ($1 \leqslant p < \infty$, $p \neq 2$) of the partial sums $s_n(u)$ requires the estimate (28.14); see Corollary 28.14 and, for more details, Wheeden and Zygmund [58, Section 2.88];
- pointwise a.e. convergence of the partial sums $s_n(u) \to u$ when $u \in L^2$ or $u \in L^p$, $1 < p < \infty$, which had been an open problem until 1966. A. N. Kolmogorov constructed in 1922/23 a function $u \in L^1$ whose Fourier series diverges a.e. In his famous 1966 paper L. Carleson proved that a.e. convergence holds for $u \in L^2$, and R. A. Hunt extended this result in 1968 to $u \in L^p$, $1 < p < \infty$.

These deep results depend on estimates of the type (28.14) and, more importantly, on estimates for $\max_{0\leqslant k\leqslant n} s_k(u)$, which resemble the maximal martingale estimates which we encountered in Chapter 25, e.g. Theorem 25.12. But there is a catch.

Lemma 28.15 $\Sigma_n = \operatorname{span}\{1, \cos x, \sin x, \dots, \cos(nx), \sin(nx)\} \subset L^2([-\pi,\pi], dx)$ *is not of the form* $L^2(\mathscr{G}_n)$ *where* $\mathscr{G}_n$ *is a sub-*σ*-algebra of the Borel sets* $\mathscr{B}[-\pi,\pi]$.

Proof The space $L^2(\mathscr{G}_n)$ is a lattice, i.e. if $f \in L^2(\mathscr{G}_n)$, then $|f| \in L^2(\mathscr{G}_n)$. Take $f(x) = \sin x$. Unfortunately,

$$|\sin x| = \frac{2}{\pi} - \frac{4}{\pi}\left(\frac{\cos(2x)}{1\cdot 3} + \frac{\cos(4x)}{3\cdot 5} + \frac{\cos(6x)}{5\cdot 7} + \cdots\right),$$

so that $\sin(\cdot) \in \Sigma_n$ but $|\sin(\cdot)| \notin \Sigma_n$. (You might also want to have a look at Theorem 27.22 for a more systematic treatment.) $\square$

This means that martingale methods are not (immediately) applicable to Fourier series.

The Haar system

In contrast to Fourier series, the Haar system allows a complete martingale treatment. Throughout this section we consider $L^2 = L^2([0,1), \mathscr{B}[0,1), \lambda)$, $\lambda = \lambda^1|_{[0,1)}$.

Definition 28.16 The *Haar system* consists of the functions

$$\left.\begin{aligned} \chi_{0,0}(x) &:= \mathbb{1}_{[0,1)}(x), \\ \chi_{k,j}(x) &:= 2^{k/2}\left(\mathbb{1}_{\left[\frac{2j-2}{2^{k+1}}, \frac{2j-1}{2^{k+1}}\right)}(x) - \mathbb{1}_{\left[\frac{2j-1}{2^{k+1}}, \frac{2j}{2^{k+1}}\right)}(x)\right), \\ & \qquad 1 \leqslant j \leqslant 2^k,\ k \in \mathbb{N}_0. \end{aligned}\right\} \tag{28.15}$$

Obviously, each Haar function is normalized to give $\|\chi_{k,j}\|_2 = 1$. The first few Haar functions are shown in Fig. 28.2.

It is often more convenient to arrange the double sequence (28.15) in lexicographical order: $\chi_{0,0}$; $\chi_{0,1}$; $\chi_{1,1}, \chi_{1,2}$; $\chi_{2,1}, \chi_{2,2}, \chi_{2,3}, \chi_{2,4}$; ... and to relabel the functions in the following way:

$$H_0 := \chi_{0,0}; \quad H_n = H_{2^k+\ell} := \chi_{k,\ell+1}, \quad 0 \leqslant \ell \leqslant 2^k - 1 \tag{28.16}$$

(note that the representation $n = 2^k + \ell$, $0 \leqslant \ell \leqslant 2^k - 1$ is unique). We can now associate with the sequence $(H_n)_{n\in\mathbb{N}}$ a canonical filtration

$$\mathscr{A}_n^H := \sigma(H_0, H_1, \dots, H_n), \qquad n \in \mathbb{N}_0,$$

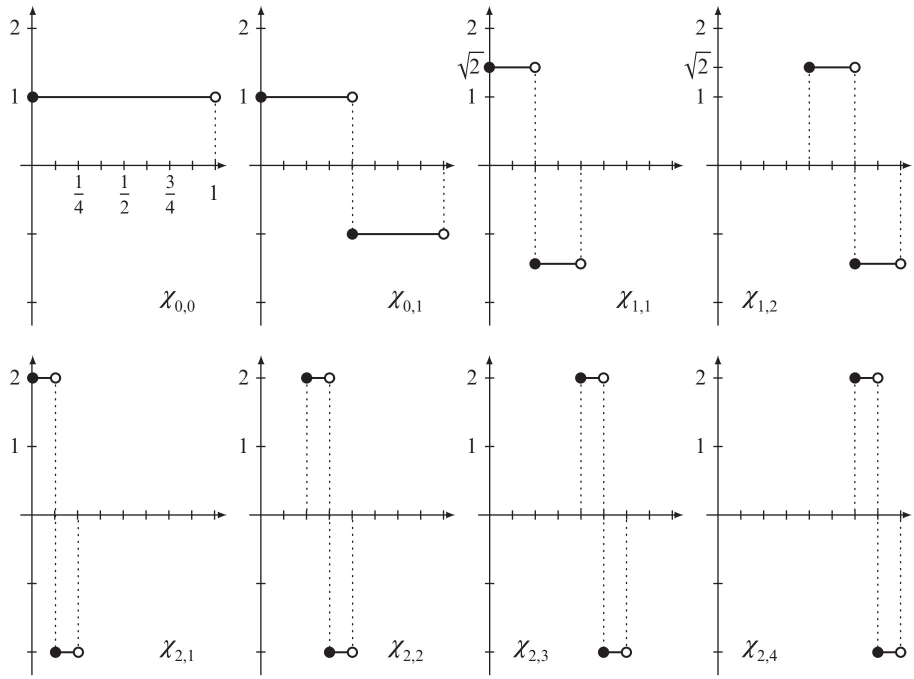


Fig. 28.2. The first few Haar functions.

which is the smallest σ-algebra that makes all functions $H_0, \ldots, H_n$ measurable, see Definition 7.5.

Theorem 28.17 *The Haar functions are a complete ONS in* $L^2([0,1), dx)$. *Moreover,*

$$M_N := \sum_{n=0}^{N} a_n H_n, \qquad N \in \mathbb{N}_0,\ a_n \in \mathbb{R},$$

is a martingale w.r.t. the filtration $(\mathscr{A}_N^H)_{N\in\mathbb{N}_0}$, *and for every* $u \in L^p([0,1), dx)$, $1 \leqslant p < \infty$, *the Haar–Fourier series*

$$s_N(u; x) := \sum_{n=0}^{N} \langle u, H_n \rangle H_n$$

converges to u in L^p *and almost everywhere, and the maximal inequality*

$$\left\| \sup_{n\in\mathbb{N}} s_n(u) \right\|_p \leqslant \frac{p}{p-1} \|u\|_p$$

holds for all $u \in L^p$ *and* $1 < p < \infty$.

Proof Step 1. Orthonormality. That $\|\chi_{k,j}\|_2 = 1$ is obvious. If the functions $\chi_{k,j} \neq \chi_{m,\ell}$ satisfy $\{\chi_{k,j} \neq 0\} \cap \{\chi_{m,\ell} \neq 0\} = \emptyset$, it is clear that $\int \chi_{k,j}\chi_{m,\ell}\, d\lambda = 0$. Otherwise, we can assume that $k < m$, so that

$$\text{either} \quad \{\chi_{m,\ell} \neq 0\} \subset \{\chi_{k,j} = +1\} \quad \text{or} \quad \{\chi_{m,\ell} \neq 0\} \subset \{\chi_{k,j} = -1\}$$

obtains. In either case,

$$\int \chi_{m,\ell}\chi_{k,j}\, d\lambda = \pm \int \chi_{m,\ell}\, d\lambda = 0.$$

Step 2. Martingale property. Let $n = 2^k + \ell$. Then, for all $n \in \mathbb{N}$,

$$\begin{aligned}\mathscr{A}_n^H &= \sigma\big(\chi_{0,0}, \chi_{0,1}, \chi_{1,1}, \ldots, \chi_{k-1,2^{k-1}}, \chi_{k,1}, \chi_{k,2}, \ldots, \chi_{k,\ell+1}\big)\\ &= \sigma\Bigg(\underbrace{\Big[0, \frac{1}{2^{k+1}}\Big), \ldots, \Big[\frac{2\ell+1}{2^{k+1}}, \frac{2\ell+2}{2^{k+1}}\Big)}_{=:\,\mathscr{E}_n}, \underbrace{\Big[\frac{\ell+1}{2^k}, \frac{\ell+2}{2^k}\Big), \ldots, \Big[\frac{2^k-1}{2^k}, 1\Big)}_{=:\,\mathscr{F}_n}\Bigg),\end{aligned}$$

where we use that the dyadic intervals are nested and refine. Assume, for simplicity, that $\ell < 2^k - 1$. Then $\{H_{n+1} \neq 0\} \in \mathscr{F}_n$, and so

$$\int_J H_{n+1}(x)dx = 0 \quad \forall J \in \mathscr{E}_n \;\text{ or }\; J \in \mathscr{F}_n.$$

(If $\ell = 2^k - 1$ we get an analogous conclusion with a rollover as $\mathscr{A}_n^H$ is just the dyadic σ-algebra generated by all disjoint half-open intervals of length 2^{-k-1} in $[0,1)$.) By Theorem 27.12 we have $\mathbb{E}^{\mathscr{A}_n^H} H_{n+1} = 0$, and by Theorem 27.11

$$\mathbb{E}^{\mathscr{A}_N^H} M_{N+1} = \mathbb{E}^{\mathscr{A}_N^H}(M_N + a_{N+1}H_{N+1}) = M_N + a_{N+1}\mathbb{E}^{\mathscr{A}_N^H} H_{N+1} = M_N.$$

This shows that $\big(M_N, \mathscr{A}_N^H\big)_{N\in\mathbb{N}}$ is indeed a martingale, see Corollary 27.18.

Step 3. Convergence in L^1 and a.e. if $u \in L^1 \cap L^\infty$. Set $a_k := \langle u, H_k\rangle$, so that $M_N = s_N(u)$ becomes the Haar–Fourier partial sum. Using Bessel's inequality (Theorem 26.19) we see

$$\|s_N(u)\|_2^2 = \sum_{k=0}^{N} |\langle u, H_k\rangle|^2 \leqslant \|u\|_2^2, \tag{28.17}$$

where the right-hand side is finite since $L^1 \cap L^\infty \subset L^2$, [✍] and from the Cauchy–Schwarz and Markov inequalities we get for all $R > 0$

$$\begin{aligned}\int_{\{|s_N(u)|>R\}} |s_N(u)|\, d\lambda &\overset{(13.6)}{\leqslant} \|s_N(u)\|_2 \lambda\big(\{|s_N(u)| > R\}\big)^{1/2}\\ &\overset{(11.4)}{\leqslant} \frac{1}{R}\|s_N(u)\|_2^2 \leqslant \frac{1}{R}\|u\|_2^2.\end{aligned}$$

Since the constant function R is in $L^2([0,1),dx)$, the martingale $(s_N(u))_{N\in\mathbb{N}}$ is uniformly integrable in the sense of Definition 22.1, and we conclude from Theorems 24.6 and 27.19 that

$$s_N(u) \xrightarrow[N\to\infty]{} u_\infty \quad \text{in } L^1 \text{ and almost everywhere.}$$

Since $(\mathscr{A}_n^H)_{n\in\mathbb{N}}$ contains the sequence $(\mathscr{A}_k^\Delta)_{k\in\mathbb{N}}$ of dyadic σ-algebras – we do indeed have $\mathscr{A}_n^\Delta = \mathscr{A}_{2^n-1}^H$ – we know that $\mathscr{A}_\infty^H := \sigma(\mathscr{A}_n^H : n\in\mathbb{N}) = \mathscr{B}[0,1)$. Just as in Example 27.21 we see that

$$\mathbb{E}^{\mathscr{A}_{2^n-1}^H} u = \mathbb{E}^{\mathscr{A}_n^\Delta} u = s_{2^n-1}(u),$$

and in view of Theorem 27.19 we conclude that $u = u_\infty$ a.e.

Step 4. Convergence in L^p if $u\in L^1\cap L^\infty$. Observe that $L^1\cap L^\infty \subset L^p$ for all $1<p<\infty$. [✍] Applying the inequality

$$\begin{aligned} |a|^p - |b|^p \leqslant \big||a|^p - |b|^p\big| &= \left| p\int_{|a|}^{|b|} t^{p-1}\,dt \right| \\ &\leqslant p\big||a|-|b|\big| \max\left\{|a|^{p-1}, |b|^{p-1}\right\} \\ &\leqslant p|a-b|\max\left\{|a|^{p-1}, |b|^{p-1}\right\}, \end{aligned}$$

$a,b\in\mathbb{R}$, $1<p<\infty$, to the martingale $\mathbb{E}^{\mathscr{A}_N^H} u = s_N(u)$, we get after integrating over $[0,1)$

$$\pm \int \left(|s_N(u)|^p - |u|^p\right) d\lambda \leqslant p\|s_N(u) - u\|_1 \|u\|_\infty^{p-1}, \quad p>1,$$

where we also use that $|s_N(u)| = \left|\mathbb{E}^{\mathscr{A}_N^H} u\right| \leqslant \|u\|_\infty$ as $|u| \leqslant \|u\|_\infty < \infty$, see Theorem 27.11(ix). From Riesz's convergence theorem (Theorem 13.10) we conclude that $s_N(u)\to u$ in L^p for all $1<p<\infty$ and $u\in L^1\cap L^\infty$.

Step 5. Convergence in L^p if $u\in L^p$. If $u\in L^p$, $1\leqslant p<\infty$, is not bounded, we set $u_k := (-k)\vee u\wedge k$. Since we have a finite measure space, it holds that $u_k\in L^p\cap L^\infty \subset L^1\cap L^\infty$, and we see from the triangle inequality and Theorem 27.11(v) and (ii) that

$$\begin{aligned} \|s_N(u) - u\|_p &\leqslant \|s_N(u) - s_N(u_k)\|_p + \|s_N(u_k) - u_k\|_p + \|u_k - u\|_p \\ &\leqslant \|s_N(u_k) - u_k\|_p + 2\|u_k - u\|_p. \end{aligned}$$

The claim follows on letting first $N\to\infty$ and then $k\to\infty$.

Step 6. Convergence a.e. if $u\in L^p$. Since $s_N(u^\pm) = \mathbb{E}^{\mathscr{A}_N^H}(u^\pm) \geqslant 0$, we know from Corollary 27.20 that $\left((s_N(u^\pm))^p\right)_{N\in\mathbb{N}}$ are submartingales which satisfy, by Theorem 27.11(ii),

$$\int (s_N(u^\pm))^p\,d\mu \leqslant \int \left(\mathbb{E}^{\mathscr{A}_N^H}(u^\pm)\right)^p d\mu = \left\|\mathbb{E}^{\mathscr{A}_N^H}(u^\pm)\right\|_p^p \leqslant \|u^\pm\|_p^p.$$

Therefore, the submartingale convergence theorem (Theorem 24.2) applies and shows that $\lim_{N\to\infty} |s_N(u^{\pm}; x)|^p$ exists a.e., hence, $\lim_{N\to\infty} s_N(u; x)$ exists a.e. As Step 5 and Corollary 13.8 already imply $\lim_{j\to\infty} s_{N_j}(u; x) = u(x)$ a.e. for some subsequence, we can identify the limit and get $\lim_{N\to\infty} s_N(u; x) = u(x)$ a.e.

Step 7. Completeness follows from $\lim\limits_{N\to\infty} \|s_N(u) - u\|_2 = 0$ and Theorem 26.21.

Step 8. The maximal inequality is just Doob's maximal L^p-inequality for martingales, Theorem 25.12, since $(s_n(u))_{n\in\mathbb{N}}$ is a uniformly integrable martingale which is, by Step 5 and Theorem 27.19, closed by $s_\infty(u) = u$. □

Remark 28.18 As a matter of fact, ordering the Haar functions in a sequence like $(H_n)_{n\in\mathbb{N}_0}$ does play a rôle. If $p = 1$, we can find (after some elementary but very tedious calculations) that

$$\left\| \chi_{0,0} + \chi_{0,1} + \sum_{k=1}^{2n} 2^{k/2} \chi_{k,1} \right\|_1 \leqslant \sqrt{2},$$

while the lacunary series satisfies

$$\left\| \chi_{0,1} + \sum_{k=1}^{n} 2^{k} \chi_{2k,1} \right\|_1 \geqslant cn$$

for some absolute constant $c > 0$. Therefore, we can rearrange $\sum_{n=0}^{\infty} a_n H_n$ in such a way that it becomes a divergent series $\sum_{n=0}^{\infty} a_{\sigma(n)} H_{\sigma(n)}$ for some necessarily infinite permutation $\sigma : \mathbb{N}_0 \to \mathbb{N}_0$.

This phenomenon does not happen if $1 < p < \infty$. In fact, $(H_n)_{n\in\mathbb{N}_0}$ is what one calls an *unconditional basis* of L^p, $1 < p < \infty$, which means that every rearrangement of the series $\sum_{n=0}^{\infty} a_n H_n$ converges in L^p and leads to the same limit. The Haar system is even the litmus test for the existence of unconditional bases: *every Banach space B where $(H_n)_{n\in\mathbb{N}_0}$ is a basis has an unconditional basis if, and only if, the basis $(H_n)_{n\in\mathbb{N}_0}$ is unconditional*, see Olevskiĭ [34, p. 72, Corollary] or Lindenstrauss and Tzafriri [28, vol. II, p. 161, Corollary 2.c.11].

Since the unconditionality of $(H_n)_{n\in\mathbb{N}_0}$ rests on a martingale argument, we include a sketch of its proof. First we need the following *Burkholder–Davis–Gundy inequalities* for a martingale $(u_n)_{n\in\mathbb{N}_0}$ on a *probability space* $(X, \mathscr{A}, \mu)$:

$$\kappa_p \left\| \sup_{0\leqslant n\leqslant N} |u_n| \right\|_p \leqslant \left\| \sqrt{[u_\bullet, u_\bullet]_N} \right\|_p \leqslant K_p \left\| \sup_{0\leqslant n\leqslant N} |u_n| \right\|_p \tag{BDG}$$

for all $N\in\mathbb{N}_0$, all $0<p<\infty$ and some absolute constants $K_p\geqslant\kappa_p>0$. The expression $[u_\bullet,u_\bullet]_N$ stands for the *quadratic variation* of the martingale

$$[u_\bullet,u_\bullet]_N:=|u_0|^2+\sum_{n=0}^{N-1}|u_{n+1}-u_n|^2.$$

A proof of (BDG) can be found in Rogers and Williams [41, vol. 2, pp. 93–5]. If we combine (BDG) with Doob's maximal L^p-inequality, Theorem 25.12, we get

$$\kappa_p\|u_N\|_p\leqslant\left\|\sqrt{[u_\bullet,u_\bullet]_N}\right\|_p\leqslant\frac{pK_p}{p-1}\|u_N\|_p \tag{BDG$'$}$$

for all $N\in\mathbb{N}_0$ and $1<p<\infty$ – mind the different range for p in (BDG$'$) compared with (BDG). Obviously,

$$u_N:=\sum_{n=0}^{N}\langle u,H_n\rangle H_n\quad\text{and}\quad w_N:=\sum_{n=0}^{N}\epsilon_n\langle u,H_n\rangle H_n,$$

$\epsilon_n\in\{-1,+1\}$, are uniformly integrable martingales (use the argument of the proof of Theorem 28.17) and their quadratic variations $[u_\bullet,u_\bullet]_N=[w_\bullet,w_\bullet]_N$ coincide. Therefore, (BDG$'$) shows that the martingales $(u_N-u_n)_{N\geqslant n}$ and $(w_N-w_n)_{N\geqslant n}$ satisfy

$$\|u_N-u_n\|_p\sim\left\|[u_\bullet-u_n,u_\bullet-u_n]_N^{\frac12}\right\|_p=\left\|[w_\bullet-w_n,w_\bullet-w_n]_N^{\frac12}\right\|_p\sim\|w_N-w_n\|_p,$$

where $a\sim b$ means that $\kappa a\leqslant b\leqslant Ka$ for some absolute constants $\kappa,K>0$, so that either both sequences converge or both sequences diverge. Let us assume that $(u_N)_{N\in\mathbb{N}_0}$ converges. Then every lacunary series

$$\sum_{k=1}^{\infty}\langle u,H_{n_k}\rangle H_{n_k} \tag{28.18}$$

converges, since we can produce its partial sums by adding and subtracting u_N and w_N with suitable ±1-sequences $(\epsilon_n)_{n\in\mathbb{N}}$. This entails that for every fixed permutation $\sigma:\mathbb{N}_0\to\mathbb{N}_0$

$$\left\|\sum_{k=n}^{N}\langle u,H_{\sigma(k)}\rangle H_{\sigma(k)}\right\|_p\leqslant\epsilon,\quad N>n\quad\text{sufficiently large.}$$

Otherwise, we could find finite sets $\Sigma_0,\Sigma_1,\Sigma_2,\ldots\subset\mathbb{N}_0$ with $(k_j)_{j\in\mathbb{N}}=\bigcup_{n\in\mathbb{N}}\Sigma_n$ and

$$\left\|\sum_{k\in\Sigma_n}\langle u,H_k\rangle H_k\right\|_p>\epsilon\quad\forall n\in\mathbb{N},$$

contradicting (28.18). For more on this topic we refer the reader to Lindenstrauss and Tzafriri [28].

The Haar Wavelet

Let us now consider a Haar system in $L^2 = L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), dx)$. We begin with the remark that the functions $\chi_{0,0} = \mathbb{1}_{[0,1)}$ and $\chi_{0,1} = \mathbb{1}_{[0,1/2)} - \mathbb{1}_{[1/2,1)}$ are the two basic Haar functions, since we can reconstruct all Haar functions $\chi_{k,j}$ from them by scaling and shifting:

$$\chi_{k,j}(x) = 2^{k/2}\chi_{0,1}(2^k x - j + 1), \quad k \in \mathbb{N}_0, j = 1, 2, \ldots, 2^k. \tag{28.19}$$

The advantage of (28.19) over the definition (28.15) is that (28.19) easily extends to all pairs $(j,k) \in \mathbb{Z}^2$ and thus to a system of functions on $\mathbb{R}$.

Definition 28.19 The *Haar wavelets* are the system $\big(\psi_{k,j}\big)_{j,k\in\mathbb{Z}}$, where the *mother wavelet* is $\psi(x) := \mathbb{1}_{[0,1/2)}(x) - \mathbb{1}_{[1/2,1)}(x)$ and

$$\psi_{k,j}(x) := 2^{k/2}\psi(2^k x - j) = 2^{k/2}\Big(\mathbb{1}_{\left[\frac{2j}{2^{k+1}}, \frac{2j+1}{2^{k+1}}\right)}(x) - \mathbb{1}_{\left[\frac{2j+1}{2^{k+1}}, \frac{2j+2}{2^{k+1}}\right)}(x)\Big)$$

for all $j, k \in \mathbb{Z}$.

Note that $\psi = \psi_{0,1} = \chi_{0,1}$, $\psi_{k,j-1} = \chi_{k,j}$ for all $j = 1, 2, \ldots, 2^k$ and $k \in \mathbb{N}_0$ while $\psi_{0,-1}(x) = 2^{-1/2}\chi_{0,0}(x)$ for $0 \leqslant x < 1$.

The Haar wavelets can be treated by martingale methods. To do so, we introduce the two-sided dyadic filtration

$$\begin{aligned}
\mathscr{A}^{\Delta}_{n+1} &= \sigma\Big(\Big[\frac{j}{2^{n+1}}, \frac{j+1}{2^{n+1}}\Big) : j \in \mathbb{Z}\Big) = \sigma\big(\psi_{n,j} : j \in \mathbb{Z}\big), \quad n \in \mathbb{Z}, \\
\mathscr{A}^{\Delta}_{-\infty} &= \bigcap_{n\in\mathbb{Z}} \mathscr{A}^{\Delta}_n = \{\emptyset, \mathbb{R}\}, \quad \mathscr{A}^{\Delta}_{\infty} = \sigma\Big(\bigcup_{n\in\mathbb{Z}} \mathscr{A}^{\Delta}_n\Big) = \mathscr{B}(\mathbb{R}).
\end{aligned} \tag{28.20}$$

The last assertion follows from the fact that $D = \{j2^{-n-1} : j \in \mathbb{Z},\, n \in \mathbb{Z}\}$ is a dense subset of $\mathbb{R}$ and that $\mathscr{B}(\mathbb{R})$ is generated by all intervals of the form $[a, b)$, where $a, b \in D$ (or, indeed, any other dense subset). [✍]

In what follows we have to consider double summations. To keep the notation simple, we write

$$\sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_{j,k}$$

as a shorthand for

$$\sum_{k=-\infty}^{\infty}\left[\sum_{j=-\infty}^{\infty} a_{j,k}\right]$$

and call $\sum_{k\geqslant \text{const.}}\sum_{j=-\infty}^{\infty}$ the *right tail* and $\sum_{-\infty<k\leqslant\text{const.}}\sum_{j=-\infty}^{\infty}$ the *left tail* of the double sum.

Theorem 28.20 *The Haar wavelets $(\psi_{k,j})_{j,k\in\mathbb{Z}}$ are a complete orthonormal system in $L^2 = L^2(\mathbb{R},\mathscr{B}(\mathbb{R}),dx)$. Moreover, for all $1<p<\infty$,*

$$u = \sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\langle u,\psi_{k,j}\rangle\psi_{k,j},\quad u\in L^p, \tag{28.21}$$

in L^p and almost everywhere, and for all $1<p<\infty$ and $u\in L^p$

$$\left\|\sup_{M,N\in\mathbb{N}}\sum_{k=-M}^{N}\sum_{j=-\infty}^{\infty}\langle u,\psi_{k,j}\rangle\psi_{k,j}\right\|_p \leqslant \frac{2p-1}{p-1}\|u\|_p. \tag{28.22}$$

Proof Step 1. Orthonormality. of the family $(\psi_{k,j})_{j,k\in\mathbb{Z}}$ can be seen with arguments similar to those in Step 1 of the proof of Theorem 28.17.

Step 2. L^p $(1\leqslant p<\infty)$ and a.e. convergence of the right tail of (28.21) *if $u\in L^1\cap L^\infty$.* Note that the inner sum is pointwise convergent since $\psi_{k,j}\psi_{k,\ell}=0$ whenever $j\neq\ell$. Consider now $u\in L^1\cap L^\infty\subset L^p$. [✍] Set

$$u_{N,-M} := \sum_{k=-M}^{N}\sum_{j=-\infty}^{\infty}\langle u,\psi_{k,j}\rangle\psi_{k,j} = \mathbb{E}^{\mathscr{A}_{N+1}^{\Delta}}u - \mathbb{E}^{\mathscr{A}_{-M}^{\Delta}}u. \tag{28.23}$$

The second equality follows from the fact that $\mathbb{E}^{\mathscr{A}_n^{\Delta}}$ is the orthogonal projection onto $L^2(\mathscr{A}_n^{\Delta})$ – whose basis is $\{\psi_{k,j} : j\in\mathbb{Z},\ k\in\mathbb{Z},\ k\leqslant n-1\}$ – and by Theorem 27.4(vii),

$$\mathbb{E}^{\mathscr{A}_{n+1}^{\Delta}}u - \mathbb{E}^{\mathscr{A}_n^{\Delta}}u = \mathbb{E}^{\mathscr{A}_{n+1}^{\Delta}}u - \mathbb{E}^{\mathscr{A}_{n+1}^{\Delta}}\left(\mathbb{E}^{\mathscr{A}_n^{\Delta}}u\right) = \mathbb{E}^{\mathscr{A}_{n+1}^{\Delta}}\left(u - \mathbb{E}^{\mathscr{A}_n^{\Delta}}u\right),$$

which is the orthogonal projection of $L^2(\mathscr{A}_n^{\Delta})^{\perp}$ onto $L^2(\mathscr{A}_{n+1}^{\Delta})$. This means that

$$\mathbb{E}^{\mathscr{A}_{n+1}^{\Delta}}u - \mathbb{E}^{\mathscr{A}_n^{\Delta}}u = \sum_{j\in\mathbb{Z}}\langle u,\psi_{n,j}\rangle\psi_{n,j}, \tag{28.24}$$

since the resulting function must be $\mathscr{A}_{n+1}^{\Delta}$-measurable as well as orthogonal to $L^2(\mathscr{A}_n^{\Delta})$: i.e. we must include $(\psi_{n,j})_{j\in\mathbb{Z}}$ and exclude $(\psi_{k,j})_{j\in\mathbb{Z},\,k<n}$. Summing (28.24) over $n=-M,\dots,N$ yields (28.23).

Since $(\mathbb{E}^{\mathscr{A}_N^\Delta} u)_{N\in\mathbb{N}}$ is by Theorem 27.19 a uniformly integrable martingale, and since $\mathscr{A}_\infty^\Delta = \mathscr{B}(\mathbb{R})$, we find that

$$\mathbb{E}^{\mathscr{A}_N^\Delta} u \xrightarrow[N\to\infty]{} u \quad \text{in } L^1 \text{ and a.e. for all } u \in L^1 \cap L^\infty.$$

As in Step 4 of the proof of Theorem 28.17 we see that this also holds in L^p.

Step 3. L^p $(1 \leqslant p < \infty)$ *convergence of the right tail of* (28.21) *if* $u \in L^p$. For a general $u \in L^p$ we can use dominated convergence to see that the functions $u_k := ((-k) \vee u \wedge k)\mathbb{1}_{[-k,k]} \in L^1 \cap L^\infty$ approximate u in L^p-sense. By Theorem 27.11(ii) and (v),

$$\begin{aligned}\left\|\mathbb{E}^{\mathscr{A}_N^\Delta} u - u\right\|_p &\leqslant \left\|\mathbb{E}^{\mathscr{A}_N^\Delta} u - \mathbb{E}^{\mathscr{A}_N^\Delta} u_k\right\|_p + \left\|\mathbb{E}^{\mathscr{A}_N^\Delta} u_k - u_k\right\|_p + \|u_k - u\|_p \\ &\leqslant \left\|\mathbb{E}^{\mathscr{A}_N^\Delta} u_k - u_k\right\|_p + 2\|u_k - u\|_p.\end{aligned}$$

In view of the result of the previous step, we can let first $N \to \infty$, then $k \to \infty$, and find that $\mathbb{E}^{\mathscr{A}_N^\Delta} u \to u$ in L^p for every $u \in L^p$.

Step 4. Convergence a.e. of the right tail of (28.21)*) if* $u \in L^p$*,* $1 < p < \infty$, follows from exactly the same arguments as were used in Step 6 of the proof of Theorem 28.17.

Step 5. L^p*-convergence* $(1 < p < \infty)$ *of the left tail of* (28.21) *if* $u \in L^p$. It remains to consider $\left(\mathbb{E}^{\mathscr{A}_{-M}^\Delta} u\right)_{M\in\mathbb{N}}$. Although this is a backwards martingale, we cannot use Theorem 24.7 as $\lambda|_{\mathscr{A}_{-\infty}^\Delta}$ is not σ-finite. Instead, we take $u \in L^p$, $1 < p < \infty$, and set $u_R := u\mathbb{1}_{[-R,R]}$, $R > 0$. For all $M \in \mathbb{N}$ with $2^M > R$ we find that

$$\mathbb{E}^{\mathscr{A}_{-M}^\Delta} u_R = 2^{-M} \int_{[-R,0]} u(x)dx\, \mathbb{1}_{[-2^M,0)} + 2^{-M} \int_{[0,R]} u(x)dx\, \mathbb{1}_{[0,2^M)},$$

where we use that $\mathbb{E}^{\mathscr{A}_{-M}^\Delta}$ projects onto the intervals $[j2^M, (j+1)2^M)$, and we find from the Hölder inequality (Theorem 13.2) with $p^{-1} + q^{-1} = 1$ that

$$\left|\mathbb{E}^{\mathscr{A}_{-M}^\Delta} u_R\right| \leqslant 2^{-M} R^{1/q} \|u\|_p \mathbb{1}_{[-2^M,2^M)},$$

which implies

$$\left\|\mathbb{E}^{\mathscr{A}_{-M}^\Delta} u_R\right\|_p \leqslant 2^{-M} R^{1/q} \|u\|_p (2 \cdot 2^M)^{1/p} = c_R 2^{-M(1-1/p)} \|u\|_p.$$

Finally, by Theorem 27.11(ii) and (v),

$$\begin{aligned}\left\|\mathbb{E}^{\mathscr{A}_{-M}^\Delta} u\right\|_p &\leqslant \left\|\mathbb{E}^{\mathscr{A}_{-M}^\Delta} (u - u_R)\right\|_p + \left\|\mathbb{E}^{\mathscr{A}_{-M}^\Delta} u_R\right\|_p \\ &\leqslant \|u - u_R\|_p + c_R 2^{-M(1-1/p)} \|u\|_p,\end{aligned}$$

and we get $\lim_{M\to\infty} \|\mathbb{E}^{\mathscr{A}_{-M}^\Delta} u\|_p = 0$ for all $u \in L^p$, $1 < p < \infty$, letting first $M \to \infty$ and then $R \to \infty$.

This shows for $M, N \to \infty$ that $u_{-M,N} \to u$ in L^p, $1 < p < \infty$, and the proof of the convergence of (28.21) in L^p, $1 < p < \infty$, is complete.

Step 6. Completeness of the Haar wavelets in L^2 follows if we apply (28.21) in the case $p = 2$, see Theorem 26.21.

Step 7. Convergence a.e. of the left tail of (28.21). Observe that

$$\begin{aligned} A &:= \left\{ \left|\mathbb{E}^{\mathscr{A}^{\Delta}_{-M}} u\right| > \epsilon : \text{ for infinitely many } M \in \mathbb{N} \right\} \\ &= \bigcap_{M=1}^{\infty} \underbrace{\bigcup_{j=M}^{\infty} \left\{ \left|\mathbb{E}^{\mathscr{A}^{\Delta}_{-j}} u\right| > \epsilon \right\}}_{\in \mathscr{A}^{\Delta}_{-M}} \in \mathscr{A}^{\Delta}_{-\infty}. \end{aligned}$$

By the martingale maximal inequality, Lemma 25.11, for the reversed martingale $\left(\mathbb{E}^{\mathscr{A}^{\Delta}_{-j}} u\right)_{j \in \mathbb{N}}$ and Theorem 27.11(ii) we see that

$$\begin{aligned} \lambda(A) &\leqslant \lambda\left(\bigcup_{j=M}^{\infty} \left\{ \left|\mathbb{E}^{\mathscr{A}^{\Delta}_{-j}} u\right| > \epsilon \right\} \right) \\ &\leqslant \lambda\left\{ \sup_{j \in \mathbb{N}} \left|\mathbb{E}^{\mathscr{A}^{\Delta}_{-j}} u\right| > \epsilon \right\} \\ &\leqslant \frac{1}{\epsilon^p} \left\|\mathbb{E}^{\mathscr{A}^{\Delta}_{-1}} u\right\|_p \leqslant \frac{1}{\epsilon^p} \|u\|_p. \end{aligned}$$

This shows that $\lambda(A) < \infty$. Since $\mathscr{A}^{\Delta}_{-\infty} = \{\emptyset, \mathbb{R}\}$ is the trivial σ-algebra, we conclude that $\lambda(A) = 0$ or $A = \emptyset$. Therefore, $\mathbb{E}^{\mathscr{A}^{\Delta}_{-M}} u \to 0$ almost everywhere and so $u_{N,-M} \to u$ $(M, N \to \infty)$ almost everywhere.

Step 8. Proof of the maximal inequality (28.22). From Step 2 we know that

$$\begin{aligned} \Big\| \sup_{N,M \in \mathbb{N}} u_{N,-M} \Big\|_p &= \left\| \sup_{N,M \in \mathbb{N}} \sum_{k=-M}^{N} \sum_{j=-\infty}^{\infty} \langle u, \psi_{k,j} \rangle \psi_{k,j} \right\|_p \\ &= \left\| \sup_{N \in \mathbb{N}} \mathbb{E}^{\mathscr{A}^{\Delta}_{N+1}} u - \inf_{M \in \mathbb{N}} \mathbb{E}^{\mathscr{A}^{\Delta}_{-M}} u \right\|_p \\ &\leqslant \frac{p}{p-1} \|u\|_p + \left\|\mathbb{E}^{\mathscr{A}^{\Delta}_{-1}} u\right\|_p. \end{aligned}$$

The last estimate follows from a combination of Minkowski's inequality, Doob's maximal L^p-inequality for martingales (Theorem 25.12) applied to the closed (by u) martingale $\left(\mathbb{E}^{\mathscr{A}^{\Delta}_{N}} u\right)_{N \in \mathbb{N} \cup \{\infty\}}$, see Step 3 and Theorem 27.19, and the fact that $\left(\left|\mathbb{E}^{\mathscr{A}^{\Delta}_{-M}} u\right|^p\right)_{M \in \mathbb{N}}$ is a reversed submartingale, which entails $\|\mathbb{E}^{\mathscr{A}^{\Delta}_{-M}} u\|_p \leqslant \|\mathbb{E}^{\mathscr{A}^{\Delta}_{-1}} u\|_p$, see Example 23.3(vi) or Corollary 27.20.

By Theorem 27.11(ii), conditional expectations are contractions on L^p, hence $\|\mathbb{E}^{\mathscr{A}_{-1}^{\Delta}}u\|_p \leqslant \|u\|_p$, and the proof is completed. $\square$

A nice introduction to the Haar and other wavelets is Pinsky 37.

The Rademacher Functions

Let $L^2 = L^2\big([0,1), \mathscr{B}[0,1), \lambda\big)$, $\lambda = \lambda^1|_{[0,1)}$. The *Rademacher functions* $(R_k)_{k\in\mathbb{N}_0}$ are functions on L^2 defined by

$$R_0 := \mathbb{1}_{[0,1)},\quad R_1 := \mathbb{1}_{[0,\frac{1}{2})} - \mathbb{1}_{[\frac{1}{2},0)},\quad R_2 := \mathbb{1}_{[0,\frac{1}{4})} - \mathbb{1}_{[\frac{1}{4},\frac{1}{2})} + \mathbb{1}_{[\frac{1}{2},\frac{3}{4})} - \mathbb{1}_{[\frac{3}{4},1)}, \ldots .$$

The graphs of the first four Rademacher functions are shown in Fig. 28.3.

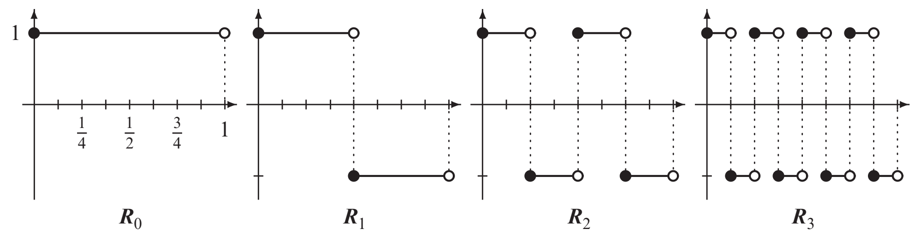


Fig. 28.3. The first four Rademacher functions.

In terms of Haar functions we have

$$R_0 = \chi_{0,0},\qquad R_{k+1} = \frac{1}{2^{k/2}} \sum_{j=1}^{2^k} \chi_{k,j},\qquad k \in \mathbb{N}_0. \tag{28.25}$$

Another equivalent definition of the Rademacher system is the following: expand each $x \in [0,1)$ as a binary series, $x = \sum_{n=1}^{\infty} \epsilon_n 2^{-n}$ with $\epsilon_n \in \{0,1\}$ – we exclude expansions terminating with a string of 1s to enforce uniqueness – and set

$$R_0(x) := \mathbb{1}_{[0,1)}(x),\qquad R_k(x) := 2\epsilon_k - 1.$$

Yet another way to think of the functions R_k is as right-continuous versions of sign changes: $R_k(x) \approx \operatorname{sgn}\sin(2^k \pi x)$, $k \in \mathbb{N}_0$.

Lemma 28.21 *The system of Rademacher functions $(R_k)_{k\in\mathbb{N}}$ is an ONS of independent*[4] *functions in $L^2([0,1), dx)$ which is NOT complete.*

[4] In the sense of Example 23.3(x) and Scholium 23.4.

Proof Orthonormality follows since we have $\int_{\{R_k=\pm1\}} R_\ell \, d\lambda = 0$ for all $k<\ell$, therefore $\int R_k R_\ell \, d\lambda = 0$ while $\int R_k^2 \, d\lambda = 1$ is obvious.

In very much the same way we deduce that $\int R_k R_1 R_2 \, d\lambda = 0$ for all $k \in \mathbb{N}_0$ which shows that the system $(R_k)_{k\in\mathbb{N}_0}$ is not complete.

Independence is a special case of Scholium 23.4 with $p = q = 1/2$. □

Although $(R_k)_{k\in\mathbb{N}_0}$ is not complete in L^2, it still has good a.e. convergence properties. The reason for this is formula (28.25) and independence.

Theorem 28.22 *The Rademacher series* $\sum_{k=1}^\infty c_k R_k$, $c_k \in \mathbb{R}$, *converges almost everywhere if, and only if,* $\sum_{k=0}^\infty c_k^2 < \infty$.

Proof Assume first that $\sum_{k=0}^\infty c_k^2 < \infty$. In view of (28.25) we set $c_{k,j} := 2^{-k/2} c_k$ and rearrange the absolutely convergent series as

$$\sum_{k=0}^\infty c_k^2 = \sum_{k=0}^\infty \sum_{j=1}^{2^k} c_{k,j}^2 < \infty.$$

We can now interpret the double sequence $(c_{k,j} : 1 \leqslant j \leqslant 2^k,\ k \in \mathbb{N}_0)$ as coefficients of the complete(!) Haar ONS $(\chi_{k,j} : 1 \leqslant j \leqslant 2^k,\ k \in \mathbb{N}_0)$. From Parseval's identity, Theorem 26.19(iv), we then conclude that the series

$$\sum_{k=0}^\infty c_k R_k = \sum_{k=0}^\infty \sum_{j=1}^{2^k} c_{k,j} \chi_{k,j}$$

converges almost everywhere and in L^2 to some element $u \in L^2$.

Conversely, assume that the series $\sum_{k=0}^\infty c_k R_k$ converges to a finite limit $s(x)$ for all $x \in E \in \mathscr{B}[0,1)$ such that $\lambda(E) > 0$. Writing s_N for the Nth partial sum of this series, we see that

$$A(N) := \bigcup_{j=N}^\infty \left\{ x \in E : |s_j(x) - s(x)| > \frac{1}{2} \right\} \quad \text{and} \quad \bigcap_{N\in\mathbb{N}} A(N) = \emptyset.$$

By virtue of the continuity of measures, Proposition 4.3, we find for every $\epsilon > 0$ some $N = N_\epsilon \in \mathbb{N}$ such that

$$\lambda(A(N)) < \epsilon < \frac{1}{2}\lambda(E) \quad \text{and} \quad \lambda(E \setminus A(N)) > 0.$$

In particular, if $E^* := E \setminus A(N)$,

$$|s_j(x) - s_k(x)| \leqslant |s_j(x) - s(x)| + |s(x) - s_k(x)| \leqslant 1 \quad \forall j,k > N,\ x \in E^*,$$

and an application of the Cauchy–Schwarz inequality for (double) series, see Example 13.11, shows that

$$
\begin{aligned}
\lambda(E^*) &\geqslant \int_{E^*} \underbrace{\left[\sum_{k=M+1}^{N} c_k R_k\right]^2}_{\leqslant 1} d\lambda \\
&= \lambda(E^*) \sum_{k=M+1}^{N} c_k^2 + 2 \sum_{M<j<k\leqslant N} c_j c_k \int_{E^*} R_j R_k \, d\lambda \\
&\geqslant \lambda(E^*) \sum_{k=M+1}^{N} c_k^2 - 2\Big[\sum_{M<j<k\leqslant N} c_j^2 c_k^2\Big]^{\frac{1}{2}} \Big[\sum_{M<j<k\leqslant N} \Big(\int_{E^*} R_j R_k \, d\lambda\Big)^2\Big]^{\frac{1}{2}} \\
&= \lambda(E^*) \sum_{k=M+1}^{N} c_k^2 - 2\Big[\sum_{M<k\leqslant N} c_k^2\Big] \Big[\sum_{M<j<k\leqslant N} \Big(\int_{E^*} R_j R_k \, d\lambda\Big)^2\Big]^{\frac{1}{2}}.
\end{aligned}
\tag{28.26}
$$

Consider now the system $(R_j R_k)_{0\leqslant j<k<\infty}$. Since for all $m > \ell \geqslant k \geqslant j$ the integral $\int_{\{R_j R_k R_\ell = \pm 1\}} R_m \, d\lambda = 0$, we see that

$$\int (R_j R_k)(R_\ell R_m) d\lambda = 0 \quad \text{if } (j,k) \neq (\ell, m).$$

This shows that $(R_j R_k)_{0\leqslant j<k<\infty}$ is itself an ONS in L^2, and by Bessel's inequality (Theorem 26.19(iii)) for this ONS and the function $u := \mathbb{1}_{E^*}$ we find

$$\sum_{j<k} \Big(\int \mathbb{1}_{E^*} R_j R_k \, d\lambda\Big)^2 \leqslant \|\mathbb{1}_{E^*}\|_2^2 = \lambda(E^*) \leqslant 1.$$

For sufficiently large values of $M \in \mathbb{N}$ we can thus achieve that

$$\sum_{M<j<k} \Big(\int_{E^*} R_j R_k \, d\lambda\Big)^2 \leqslant \Big(\frac{\lambda(E^*)}{4}\Big)^2,$$

and as $N \to \infty$ (28.26) becomes

$$\lambda(E^*) \geqslant \lambda(E^*) \sum_{k>M} c_k^2 - 2\Big(\sum_{k>M} c_k^2\Big) \frac{\lambda(E^*)}{4} = \frac{\lambda(E^*)}{2} \sum_{k>M} c_k^2,$$

which implies that $\sum_{k>M} c_k^2 \leqslant 2$, i.e. $\sum_{k=0}^{\infty} c_k^2 < \infty$, and we are done. □

It is possible to extend the Rademacher system explicitly to a complete ONS. This can be achieved by the following construction:

$$w_0 := R_0, \quad w_n := R_{j_1+1} \cdot R_{j_2+1} \cdot \dots \cdot R_{j_k+1}, \quad n \in \mathbb{N}, \tag{28.27}$$

where $n = 2^{j_1} + 2^{j_2} + \dots + 2^{j_k}$ is the unique dyadic representation of $n \in \mathbb{N}$, where $0 \leqslant j_1 < j_2 < \dots < j_k$. A similar argument to the one used in the second part of the proof of Theorem 28.22 shows that $(w_n)_{n\in\mathbb{N}_0}$ is indeed an ONS. Note that $R_{k+1} = w_{2^k}$, so that $(R_k)_{k\in\mathbb{N}_0} \subset (w_n)_{n\in\mathbb{N}_0}$.

Definition 28.23 The system (28.27) is called the *Walsh orthonormal system* (in Paley's ordering).

The Walsh system is a complete ONS, see Alexits [1, pp. 60–62] or Schipp *et al.* [46], and it is susceptible to a martingale treatment, see [46]. Again one considers the filtration of dyadic σ-algebras $(\mathscr{A}_n^\Delta)_{n\in\mathbb{N}_0}$ on $[0, 1)$ and the special partial sums

$$s_{2^n-1}(u) := \sum_{j=1}^{2^n-1} \langle u, w_j \rangle w_j.$$

Then $s_n(u) = \mathbb{E}^{\mathscr{A}_n^\Delta} u$ and we have the full martingale toolkit at our disposal. With the methods used so far it is possible to show that $s_{2^n-1}(u) \to u$ a.e. and in L^p, $1 \leqslant p < \infty$. The case of general partial sums $s_n(u)$ is somewhat harder to handle but it is still doable with some variations of the techniques presented here; see Schipp *et al.* [46, Chapters 4 and 6].

Well-Behaved Orthonormal Systems

For the Haar system and the Haar wavelet we can use martingale methods. A close inspection of our proofs reveals that the crucial input for getting martingales is that the ONS $(e_n)_{n\in\mathbb{N}_0}$ satisfies

$$\mathbb{E}^{\mathscr{A}_n} e_{n+1} = 0, \quad n \in \mathbb{N}_0, \tag{28.28}$$

where $\mathscr{A}_n = \sigma(e_0, e_1, \dots, e_n)$. This condition implies immediately that the partial sum $\sum_{k=0}^n c_k e_k$ is a martingale w.r.t. the filtration $(\mathscr{A}_n)_{n\in\mathbb{N}_0}$ generated by the ONS $(e_n)_{n\in\mathbb{N}}$.

Definition 28.24 Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and $1 \leqslant p < \infty$. A family of functions $(e_n)_{n\in\mathbb{N}_0} \subset L^p(X, \mathscr{A}, \mu)$ satisfying (28.28) is called a *system of martingale differences*.

For martingale differences orthogonality is not required. The archetype of martingale differences is sequences of independent[5] functions $(f_n)_{n\in\mathbb{N}_0} \subset L^2$ ($\subset L^1$, since μ is a probability measure) which are normalized such that $\int f_n \, d\mu = 0$ and $\int f_n^2 \, d\mu = 1$. Our methods used in connection with the Haar system and Haar wavelets still apply and yield.

Theorem 28.25 *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and let $(e_n)_{n\in\mathbb{N}_0}$ be an ONS of martingale differences in $L^2(X, \mathscr{A}, \mu)$. Then*

$$s_n(u; x) := \sum_{k=0}^{n} \langle u, e_k \rangle e_k(x), \quad n \in \mathbb{N},\ u \in L^1 \cap L^2,$$

is a martingale w.r.t. the filtration $\mathscr{A}_n := \sigma(e_0, e_1, \dots, e_n)$. For every $u \in L^2$ the sequence $(s_n(u))_{n\in\mathbb{N}}$ converges a.e. and satisfies the following maximal inequality:

$$\Big\| \sup_{n\in\mathbb{N}} s_n(u) \Big\|_2 \leqslant 2\|u\|_2, \quad u \in L^2.$$

Proof That the sequence of partial sums satisfies $\mathbb{E}^{\mathscr{A}_n} s_{n+1}(u) = s_n(u)$ for $u \in L^2$ and is, for $u \in L^1 \cap L^2$, a martingale is clear. Therefore, Corollary 27.20 shows that $(|s_n(u^\pm)|^2)_{n\in\mathbb{N}}$ are submartingales, and, from Bessel's inequality, see Theorem 26.19,

$$\sup_{n\in\mathbb{N}} \|s_n(u^\pm)\|_2 \leqslant \|u^\pm\|_2, \quad u \in L^2,$$

we conclude that $(|s_n(u^\pm)|^2)_{n\in\mathbb{N}}$ satisfy the conditions of the submartingale convergence theorem (Theorem 24.2). Thus $\lim_{n\to\infty} |s_n(u^\pm)|^2$ exists a.e. in $[0, \infty)$, and, since $s_n(u^\pm) \geqslant 0$, so does $\lim_{n\to\infty} s_n(u)$.

From Doob's maximal inequality theorem (Theorem 25.12) and Bessel's inequality (Theorem 26.19) we get

$$\Big\| \sup_{n\leqslant N} s_n(u) \Big\|_2 \leqslant 2\|s_N(u)\|_2 \leqslant 2\|u\|_2, \quad N \in \mathbb{N},$$

and the usual monotone convergence argument proves the maximal inequality as $N \to \infty$. ∎

In the situation of Theorem 28.25 we cannot say much more about the limit $\lim_{n\to\infty} s_n(u; x)$ apart from its mere existence. In particular, the partial sums $s_n(u)$ can converge to something completely different from u. Consider,

[5] In the sense of Example 23.3(x).

for example, the system of Rademacher functions $(R_n)_{n\in\mathbb{N}_0}$, which is clearly a system of martingale differences. If $u = R_1 R_2$ we get

$$\langle u, R_k\rangle = \langle R_1 R_2, R_k\rangle = \int_{[0,1)} R_1(x)R_2(x)R_k(x)dx = 0, \quad \forall k \in \mathbb{N}.$$

Thus $s_n(R_1R_2) \equiv 0$ is convergent, but $\lim_{n\to\infty} s_n(R_1R_2) \equiv 0 \neq R_1R_2$. The reason is that the Rademacher functions are not complete in L^2. This also means that we cannot hope to get L^p-convergence in Theorem 28.25.

Theorem 28.26 *Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space and $(e_n)_{n\in\mathbb{N}_0} \subset L^2(\mathscr{A})$ be an ONS of martingale differences. Denote by $s_n(u)$ the partial sum*

$$s_n(u;x) := \sum_{k=0}^{n} \langle u, e_k\rangle e_k(x), \quad u \in L^2,$$

and by $\mathscr{A}_n := \sigma(e_0, e_1, \ldots, e_n)$ the associated canonical filtration. Then the following assertions are equivalent.

(i) $(e_n)_{n\in\mathbb{N}_0}$ *is a complete ONS.*
(ii) $\int_A s_n(u)d\mu = \int_A u\,d\mu$ *for all $A \in \mathscr{A}_n$, $\mu(A) < \infty$, and $u \in L^2(\mu)$.*
(iii) $\mathbb{E}^{\mathscr{A}_n} u = s_n(u)$ *for all $u \in L^2(\mu)$.*
(iv) $\lim_{n\to\infty} \|s_n(u) - u\|_p = 0$ *for all $u \in L^p(\mu)$ and all $1 \leqslant p < \infty$.*

Proof (i)$\Rightarrow$(ii). Since $(e_j)_{j\in\mathbb{N}_0}$ is complete, we know from Theorem 26.21 that $\lim_{n\to\infty} \|s_n(u) - u\|_2 = 0$ for all $u \in L^2$. Using the Cauchy–Schwarz inequality we see for every $A \in \mathscr{A}_n$ with $\mu(A) < \infty$

$$\int_A |s_n(u) - u|d\mu \leqslant \|s_n(u) - u\|_2 \cdot \|\mathbb{1}_A\|_2 \xrightarrow[n\to\infty]{} 0.$$

Thus $\lim_{n\to\infty} \int_A s_n(u)d\mu = \int_A u\,d\mu$. Since $(e_j)_{j\in\mathbb{N}_0}$ is a system of martingale differences, we know that $\mathbb{E}^{\mathscr{A}_n} e_{n+k} = 0$, $k \in \mathbb{N}$, [✍] and by Theorem 27.12, applied to the function $\mathbb{1}_A e_{n+k} \in L^1$ and $A \in \mathscr{A}_n$,

$$\int_A e_{n+k}\,d\mu = \int_A \mathbb{1}_A e_{n+k}\,d\mu = 0.$$

Therefore $\int_A u\,d\mu = \lim_{j\to\infty} \int_A s_j(u)d\mu = \int_A s_n(u)d\mu$ holds for all $n \in \mathbb{N}$ and $A \in \mathscr{A}_n$ with $\mu(A) < \infty$.

(ii)$\Rightarrow$(iii). Since $u\mathbb{1}_A \in L^1$ for all $A \in \mathscr{A}_n$ with $\mu(A) < \infty$ and $u \in L^2$, Theorems 27.11(vii) and 27.12 show that

$$\begin{aligned}\int_A u\,d\mu = \int_A u\mathbb{1}_A\,d\mu = \int_A \mathbb{E}^{\mathscr{A}_n}(u\mathbb{1}_A)d\mu \\ = \int_A \mathbb{E}^{\mathscr{A}_n}(u)\mathbb{1}_A\,d\mu = \int_A \mathbb{E}^{\mathscr{A}_n}(u)d\mu.\end{aligned}$$

Together with the assumption this gives

$$\int_A \mathbb{E}^{\mathscr{A}_n}u\,d\mu = \int_A s_n(u)d\mu \quad \forall A \in \mathscr{A}_n,\ \mu(A) < \infty.$$

Choose, in particular, for every $k \in \mathbb{N}$ the set $\{s_n(u) > 1/k + \mathbb{E}^{\mathscr{A}_n}u\} \in \mathscr{A}_n$. With Markov's inequality (11.4) we see

$$\mu\left\{s_n(u) - \mathbb{E}^{\mathscr{A}_n}u > \frac{1}{k}\right\} \leqslant k^2\left\|s_n(u) - \mathbb{E}^{\mathscr{A}_n}u\right\|_2^2 < \infty,$$

so that the above equality becomes

$$0 = \int_{\left\{s_n(u) > \frac{1}{k} + \mathbb{E}^{\mathscr{A}_n}u\right\}} \left(s_n(u) - \mathbb{E}^{\mathscr{A}_n}u\right)d\mu \geqslant \frac{1}{k}\mu\left\{s_n(u) > \frac{1}{k} + \mathbb{E}^{\mathscr{A}_n}u\right\}.$$

This is possible only if $\mu\{s_n(u) > \frac{1}{k} + \mathbb{E}^{\mathscr{A}_n}u\} = 0$. A similar argument for the set $\{s_n(u) < \frac{1}{k} + \mathbb{E}^{\mathscr{A}_n}u\}$ finally shows

$$\begin{aligned}\mu\left\{s_n(u) \neq \mathbb{E}^{\mathscr{A}_n}u\right\} &= \mu\left(\bigcup_{k\in\mathbb{N}}\left\{|s_n(u) - \mathbb{E}^{\mathscr{A}_n}u| > \frac{1}{k}\right\}\right) \\ &\leqslant \sum_{k\in\mathbb{N}}\mu\left\{|s_n(u) - \mathbb{E}^{\mathscr{A}_n}u| > \frac{1}{k}\right\} = 0.\end{aligned}$$

Therefore, $s_n(u) = \mathbb{E}^{\mathscr{A}_n}u$ a.e.

(iii)$\Rightarrow$(iv). For $u \in L^1 \cap L^\infty$ Theorem 27.19 shows that $\left(\mathbb{E}^{\mathscr{A}_n}u\right)_{n\in\mathbb{N}}$ is a uniformly integrable martingale and that $\mathbb{E}^{\mathscr{A}_n}u \to u$ in L^1 and a.e. As in Step 4 of the proof of Theorem 28.17, we use the inequality

$$|a|^p - |b|^p \leqslant p|a-b|\max\{|a|^{p-1}, |b|^{p-1}\}, \quad a, b \in \mathbb{R},\ p > 1,$$

to deduce that

$$\pm\int\left(\left|\mathbb{E}^{\mathscr{A}_n}u\right|^p - |u|^p\right)d\mu \leqslant p\|\mathbb{E}^{\mathscr{A}_n}u - u\|_1 \cdot \|u\|_\infty^{p-1}$$

and, by Riesz's convergence theorem (Theorem 13.10), that $\mathbb{E}^{\mathscr{A}_n}u \to u$ in L^p.

If $u \in L^p$ is not bounded, we take an exhausting sequence $(A_k)_{k\in\mathbb{N}} \subset \mathscr{A}$ with $A_k \uparrow X$ and $\mu(A_k) < \infty$ and set $u_k := \big((-k) \vee u \wedge k\big)\mathbb{1}_{A_k}$. Clearly, $u_k \in L^1 \cap L^\infty$, and we see, using Theorem 27.11(ii) and (v), that

$$\begin{aligned}\left\|\mathbb{E}^{\mathscr{A}_n}u - u\right\|_p &\leqslant \left\|\mathbb{E}^{\mathscr{A}_n}u - \mathbb{E}^{\mathscr{A}_n}u_k\right\|_p + \left\|\mathbb{E}^{\mathscr{A}_n}u_k - u_k\right\|_p + \|u_k - u\|_p \\ &\leqslant \left\|\mathbb{E}^{\mathscr{A}_n}u_k - u_k\right\|_p + 2\|u_k - u\|_p.\end{aligned}$$

The claim follows if we let first $n \to \infty$ and then $k \to \infty$.

(iv) $\Rightarrow$ (i) is just $p = 2$ combined with Theorem 26.21. □

If we know that the elements of the ONS are independent, we obtain the following necessary and sufficient conditions for pointwise convergence which generalize Theorem 28.22.

Theorem 28.27 *Let $(X, \mathscr{A}, \mathbb{P})$ be a probability space and $(e_n)_{n\in\mathbb{N}_0} \subset L^2(\mathbb{P})$ be independent random variables such that*

$$\int e_n \, d\mathbb{P} = 0 \quad \textit{and} \quad \int e_n^2 \, d\mathbb{P} = 1$$

and let $(c_n)_{n\in\mathbb{N}_0} \subset \mathbb{R}$ be a sequence of real numbers. Then the following statements hold.

(i) *The family $(e_n)_{n\in\mathbb{N}_0}$ is an ONS of martingale differences.*
(ii) *If $\sum_{n=0}^\infty c_n^2 < \infty$, then $\sum_{n=0}^\infty c_n e_n$ converges in $L^2(\mathbb{P})$ and a.e.*
(iii) *If $\sup_{n\in\mathbb{N}_0} \|e_n\|_\infty \leqslant \kappa < \infty$ and if $\sum_{n=0}^\infty c_n e_n$ converges almost everywhere, then $\sum_{n=0}^\infty c_n^2 < \infty$.*

Proof (i) We set $\mathscr{A}_n := \sigma(e_0, e_1, \dots, e_n)$ and $u_n := \sum_{k=0}^n c_k e_k$. Since $\mathbb{P}$ is a probability measure, $u_j \in L^2(\mathbb{P}) \subset L^1(\mathbb{P})$ and under our assumptions it is clear that $(u_j, \mathscr{A}_j)_{j\in\mathbb{N}_0}$ is a martingale. [✍]

By independence we have

$$\int e_i e_k \, d\mathbb{P} = \begin{cases} \int e_i \, d\mathbb{P} \cdot \int e_k \, d\mathbb{P} = 0 & \text{if } i \neq k, \\ \int e_i^2 \, d\mathbb{P} = 1 & \text{if } i = k, \end{cases}$$

which entails

$$\int u_n^2 \, d\mathbb{P} = \sum_{i,k=0}^n c_i c_k \int e_i e_k \, d\mathbb{P} = \sum_{i=0}^n c_i^2 \tag{28.29}$$

and also

$$\int u_n u_{n+k} \, d\mathbb{P} = \int \mathbb{E}^{\mathscr{A}_n} u_n u_{n+k} \, d\mathbb{P} = \int u_n \mathbb{E}^{\mathscr{A}_n} u_{n+k} \, d\mathbb{P} = \sum_{i=0}^n c_i^2. \tag{28.30}$$

(ii) Because of (28.29) we see that

$$\|u_n\|_1^2 \leqslant \|u_n\|_2^2 = \sum_{j=0}^{n} c_j^2 \leqslant \sum_{j=0}^{\infty} c_j^2 < \infty,$$

and the martingale convergence theorem, Corollary 24.3, shows that almost everywhere $u_n \to u_\infty$. Using (28.30) we conclude that

$$\begin{aligned}\int (u_{n+k} - u_n)^2 \, d\mathbb{P} &= \int \left(u_{n+k}^2 - 2u_n u_{n+k} + u_n^2\right) d\mathbb{P} \\ &= \int u_{n+k}^2 \, d\mathbb{P} - \int u_n^2 \, d\mathbb{P} = \sum_{j=n+1}^{n+k} c_j^2 \leqslant \sum_{j=n+1}^{\infty} c_j^2.\end{aligned}$$

Thus, by Fatou's lemma (Theorem 9.11),

$$\int (u_\infty - u_n)^2 \, d\mathbb{P} \leqslant \liminf_{k\to\infty} \int (u_{n+k} - u_n)^2 \, d\mathbb{P} \leqslant \sum_{j=n+1}^{\infty} c_j^2 \xrightarrow[n\to\infty]{} 0,$$

and $u_n \to u_\infty$ follows in the L^2-sense.

(iii) Since e_n and $\mathscr{A}_{n-1}$ are independent, we find for all $A \in \mathscr{A}_{n-1}$

$$\int_A (u_n - u_{n-1})^2 \, d\mathbb{P} = \int_A c_n^2 e_n^2 \, d\mathbb{P} \overset{(23.6)}{=} c_n^2 \, \mathbb{P}(A) = c_n^2 \int_A d\mathbb{P}. \tag{28.31}$$

Essentially the same calculation that was used in (28.30) also yields

$$\begin{aligned}\int_A (u_n - u_{n-1})^2 \, d\mathbb{P} &= \int (u_n \mathbb{1}_A - u_{n-1}\mathbb{1}_A)^2 \, d\mathbb{P} \\ &= \int (u_n \mathbb{1}_A)^2 \, d\mathbb{P} - \int (u_{n-1}\mathbb{1}_A)^2 \, d\mathbb{P} \\ &= \int_A (u_n^2 - u_{n-1}^2) d\mathbb{P},\end{aligned}$$

which can be combined with (28.31) to give

$$\int_A \left(u_n^2 - \sum_{i=0}^{n} c_i^2\right) d\mathbb{P} = \int_A \left(u_{n-1}^2 - \sum_{i=0}^{n-1} c_i^2\right) d\mathbb{P} \quad \forall A \in \mathscr{A}_{n-1}.$$

This means, however, that $w_n := u_n^2 - \sum_{i=0}^{n} c_i^2$ is a martingale.

Consider the stopping time $\tau = \tau_\gamma := \inf\{n \in \mathbb{N}_0 : |u_n| > \gamma\}$, $\inf\emptyset = \infty$. Since the series $\sum_{i=0}^{\infty} c_i e_i$ converges a.e., we can choose $\gamma > 0$ in such a way that

$$\kappa^2 \mathbb{P}\{\tau < \infty\} < \frac{1}{2}\mathbb{P}\{\tau = \infty\}.$$

Without loss of generality we may also take $\gamma^2 > \left|\int w_0 \, d\mathbb{P}\right| + \left|\int u_0^2 \, d\mathbb{P}\right|$.

The optional sampling theorem (Theorem 23.8) proves that $(w_{\tau\wedge n})_{n\in\mathbb{N}}$ is again a martingale and, therefore,

$$\int w_0\,d\mathbb{P} = \int w_{\tau\wedge n}\,d\mathbb{P} = \int u_{\tau\wedge n}^2\,d\mathbb{P} - \int \sum_{i=0}^{\tau\wedge n} c_i^2\,d\mathbb{P}. \tag{28.32}$$

Taking into account the very definition of τ we find, furthermore,

$$\begin{aligned}\int u_{\tau\wedge n}^2\,d\mathbb{P} &= \int_{\{\tau>n\}} u_{\tau\wedge n}^2\,d\mathbb{P} + \int_{\{1\leqslant\tau\leqslant n\}} u_{\tau\wedge n}^2\,d\mathbb{P} + \int_{\{\tau=0\}} u_{\tau\wedge n}^2\,d\mathbb{P}\\ &\leqslant 2\gamma^2 + \int_{\{1\leqslant\tau\leqslant n\}} u_\tau^2\,d\mathbb{P}\\ &= 2\gamma^2 + \int_{\{1\leqslant\tau\leqslant n\}} (c_\tau e_\tau + u_{\tau-1})^2\,d\mathbb{P}\\ &\leqslant 2\gamma^2 + 2\int_{\{1\leqslant\tau\leqslant n\}} \left(c_\tau^2 e_\tau^2 + u_{\tau-1}^2\right)d\mathbb{P},\end{aligned}$$

where we use the elementary inequality $(a+b)^2 \leqslant 2a^2 + 2b^2$ in the last line. Since the e_i are uniformly bounded by κ and since $|u_{\tau-1}| \leqslant \gamma$, we get

$$\begin{aligned}\int u_{\tau\wedge n}^2\,d\mathbb{P} &\leqslant 4\gamma^2 + \kappa^2 \int_{\{\tau\leqslant n\}} c_\tau^2\,d\mathbb{P}\\ &\leqslant 4\gamma^2 + \kappa^2\mathbb{P}\{\tau\leqslant n\}\sum_{i=0}^n c_i^2\\ &\leqslant 4\gamma^2 + \tfrac12\mathbb{P}\{\tau=\infty\}\sum_{i=0}^n c_i^2,\end{aligned} \tag{28.33}$$

since, by construction, $\kappa^2\mathbb{P}\{\tau\leqslant n\} \leqslant \kappa^2\mathbb{P}\{\tau<\infty\} < \frac12\mathbb{P}\{\tau=\infty\}$.

By rearranging (28.32) and combining this with the above estimates, we obtain

$$\begin{aligned}\mathbb{P}\{\tau=\infty\}\sum_{i=0}^n c_i^2 = \int_{\{\tau=\infty\}} \sum_{i=0}^{\tau\wedge n} c_i^2\,d\mathbb{P} &\leqslant \int \sum_{i=0}^{\tau\wedge n} c_i^2\,d\mathbb{P}\\ &\overset{(28.32)}{=} \int u_{\tau\wedge n}^2\,d\mathbb{P} - \int w_0\,d\mathbb{P}\\ &\overset{(28.33)}{\leqslant} 4\gamma^2 + \tfrac12\mathbb{P}\{\tau=\infty\}\sum_{i=0}^n c_i^2 + \gamma^2,\end{aligned}$$

uniformly for all $n \in \mathbb{N}$. Since, by assumption, $\mathbb{P}\{\tau = \infty\} > 0$ for sufficiently large γ, we conclude that $\sum_{i=0}^{\infty} c_i^2 < \infty$. □

Theorem 28.27 has an astonishing corollary if we apply the Burkholder–Davis–Gundy (BDG) inequalities from p. 383 to the martingale

$$w_n := u_{n+k} - u_k = \sum_{i=k+1}^{n+k} c_i e_i$$

w.r.t. the filtration $\mathscr{F}_n := \mathscr{A}_{n+k} := \sigma(e_0, e_1, \dots, e_{n+k})$. The part of the inequalities which is important for our purposes reads

$$\kappa_p \|w_n\|_p \leqslant \kappa_p \Big\| \sup_{0 \leqslant i \leqslant n} |w_n| \Big\|_p \leqslant \Big\| \sqrt{[w_\bullet, w_\bullet]_n} \Big\|_p, \tag{28.34}$$

where $n \in \mathbb{N}_0$, $0 < p < \infty$, and the quadratic variation is given by

$$[w_\bullet, w_\bullet]_n = [u_{\bullet+k} - u_k, u_{\bullet+k} - u_k]_n = \sum_{i=0}^{n-1} |u_{i+k+1} - u_{i+k}|^2 = \sum_{i=k+1}^{n+k} c_i^2 e_i^2.$$

If we happen to know that $\sup_{i \in \mathbb{N}} \|e_i\|_\infty \leqslant \kappa < \infty$, we even find

$$\Big\| \sqrt{[w_\bullet, w_\bullet]_n} \Big\|_\infty \leqslant \kappa \bigg(\sum_{i=k+1}^{n+k} c_i^2 \bigg)^{1/2}$$

and we conclude from (28.34) that for all $k, n \in \mathbb{N}$ and $0 < p < \infty$

$$\begin{aligned} \kappa_p \|u_{n+k} - u_k\|_p &\leqslant \Big\| [u_{\bullet+k} - u_k, u_{\bullet+k} - u_k]_n^{1/2} \Big\|_p \\ &\leqslant \Big\| [u_{\bullet+k} - u_k, u_{\bullet+k} - u_k]_n^{1/2} \Big\|_\infty \leqslant \kappa \bigg(\sum_{i=k+1}^{n+k} c_i^2 \bigg)^{1/2} \end{aligned}$$

holds. This proves immediately the following corollary.

Corollary 28.28 *Let $(X, \mathscr{A}, \mathbb{P})$ be a probability space and let $(e_n)_{n \in \mathbb{N}_0}$ be a sequence of independent random variables such that*

$$\sup_{n \in \mathbb{N}_0} \|e_n\|_\infty < \infty, \qquad \int e_n \, d\mathbb{P} = 0 \quad \text{and} \quad \int e_n^2 \, d\mathbb{P} = 1.$$

Then $u_n := \sum_{k=0}^{n} c_k e_k$ converges in L^2 and a.e. to some $u \in L^2$ if, and only if, $\sum_{k=0}^{\infty} c_k^2 < \infty$. In this case, $u \in L^p$ and the convergence takes place in L^p-sense for all $0 < p < \infty$.

Unfortunately, many ONSs of martingale differences are incomplete and seem to behave more often like Rademacher functions rather than like Haar functions. More on this topic can be found in the paper by Gundy [20] and the book by Garsia [18].

Epilogue 28.29 The combination of martingale methods and orthogonal expansions opens up a whole new world. Let us illustrate this by a rapid construction of one of the most prominent stochastic processes: the *Wiener process* or *Brownian motion*.

Choose in Theorem 28.27 $(X, \mathscr{A}, \mathbb{P}) = ([0,1], \mathscr{B}[0,1], \lambda)$, where λ is one-dimensional Lebesgue measure on $[0,1]$; denoting points in $[0,1]$ by ω, we will often write $d\omega$ instead of $\lambda(d\omega)$. Assume that the independent, identically distributed random variables e_n are all standard normal Gaussian random variables, i.e.

$$\mathbb{P}(e_n \in B) = \frac{1}{\sqrt{2\pi}} \int_B e^{-x^2/2}\,dx, \quad B \in \mathscr{B}(\mathbb{R}),$$

and consider the series expansion

$$W_t(\omega) := \sum_{n=0}^{\infty} e_n(\omega) \langle \mathbb{1}_{[0,t]}, H_n \rangle, \quad \omega \in [0,1].$$

Here $t \in [0,1]$, $\langle u, v \rangle = \int_0^1 u(x)v(x)dx$, and H_n, $n = 2^k + j$, $0 \leqslant j < 2^k$, denote the lexicographically ordered Haar functions (28.16). A short calculation confirms for $n \geqslant 1$

$$\langle \mathbb{1}_{[0,t]}, H_n \rangle = \int_0^t H_n(x)dx = 2^{k/2} \int_0^t H_1(2^k x - j)dx = F_n(t),$$

where

$$F_1(t) = \int_0^t H_1(x)dx\, \mathbb{1}_{[0,1]}(t) = t\mathbb{1}_{[0,\frac{1}{2})}(t) - (t-1)\mathbb{1}_{[\frac{1}{2},1]}(t),$$
$$F_n(t) = 2^{-k/2} F_1(2^k t - j),$$

are tent-functions which have the same supports as the Haar functions. Since the Haar functions constitute a complete orthonormal system, see Theorem 28.17, we may apply Bessel's inequality, Theorem 26.19(iii), to get

$$\sum_{n=0}^{\infty} \langle \mathbb{1}_{[0,t]}, H_n \rangle^2 \leqslant \langle \mathbb{1}_{[0,t]}, \mathbb{1}_{[0,t]} \rangle = t \leqslant 1.$$

Thus, Theorem 28.27(ii) guarantees that $W_t(\omega)$ exists, for each $t \in [0,1]$, both in $L^2(d\omega)$-sense and $\lambda(d\omega)$-almost everywhere.

More is true. Since the e_n are independent Gaussian random variables, so are their finite linear combinations (e.g. Bauer [3, Section 24]) and, in particular, the expressions

$$S_N(t;\omega) := \sum_{n=0}^{N} e_n(\omega)\langle \mathbb{1}_{[0,t]}, H_n\rangle$$

and for $s<t$ resp. $s<t\leqslant u<v$ and $x,y\in\mathbb{R}$

$$S_N(t;\omega) - S_N(s,\omega) = \sum_{n=0}^{N} e_n(\omega)\langle \mathbb{1}_{(s,t]}, H_n\rangle,$$
$$x\big[S_N(t;\omega) - S_N(s;\omega)\big] + y\big[S_N(v;\omega) - S_N(u;\omega)\big]$$

are Gaussian. Gaussianity is preserved under L^2-limits;[6] we conclude that $W_t(\omega)$ has a Gaussian distribution for each t. The mean is given by

$$\int_0^1 W_t(\omega)d\omega = \sum_{n=0}^{\infty}\int_0^1 e_n(\omega)d\omega\ \langle \mathbb{1}_{[0,t]}, H_n\rangle = 0$$

(to change integration and summation we use that $L^2(d\omega)$-convergence entails $L^1(d\omega)$-convergence on a finite measure space). Since $\int e_n e_m\,d\omega = 0$ or 1 according to $n\neq m$ or $n=m$, we can calculate the variance for $0\leqslant s<t\leqslant 1$ by taking

$$\begin{aligned}
&\int_0^1 (W_t(\omega) - W_s(\omega))^2\,d\omega\\
&\quad = \sum_{n,m=0}^{\infty}\int_0^1 e_n(\omega)e_m(\omega)d\omega\ \langle \mathbb{1}_{[0,t]} - \mathbb{1}_{[0,s]}, H_n\rangle\langle \mathbb{1}_{[0,t]} - \mathbb{1}_{[0,s]}, H_m\rangle\\
&\quad = \sum_{n=0}^{\infty}\langle \mathbb{1}_{(s,t]}, H_n\rangle^2 \overset{28.17}{\underset{26.21}{=}} \langle \mathbb{1}_{(s,t]}, \mathbb{1}_{(s,t]}\rangle\ =\ t-s.
\end{aligned}$$

In particular, the increment $W_t - W_s$ has the same probability distribution as W_{t-s}. In the same vein, we find for $0\leqslant s<t\leqslant u<v\leqslant 1$ that

$$\int_0^1 (W_t(\omega) - W_s(\omega))(W_v(\omega) - W_u(\omega))d\omega = \langle \mathbb{1}_{(s,t]}, \mathbb{1}_{(u,v]}\rangle = 0.$$

[6] This can easily be seen with the Fourier transform: if ξ_n is normally distributed with mean 0 and variance σ_n^2, its probability density is given by $g_{\sigma_n^2}(x) = (2\pi\sigma_n^2)^{-1/2}e^{-x^2/2\sigma_n^2}$ and the Fourier transform is $(2\pi)^{-1/2}\int e^{-iu\xi_n}\,d\mathbb{P} = (2\pi)^{-1/2}e^{-\sigma_n^2u^2/2}$, see Example 19.2(iii). If $\xi_n\to\xi$ in L^2-sense, we have $\sigma_n^2\to\sigma^2$ and, by dominated convergence, $\int e^{-iu\xi}\,d\mathbb{P} = \lim_n\int e^{-iu\xi_n}\,d\mathbb{P} = \lim_n e^{-\sigma_n^2u^2/2} = e^{-\sigma^2u^2/2}$; the claim follows from the uniqueness of the Fourier transform, Corollary 19.8.

Since $W_t - W_s$ is Gaussian, this proves already the independence of the two increments $W_t - W_s$ and $W_v - W_u$, see [3, Section 24]. By induction, we conclude that

$$W_{t_n} - W_{t_{n-1}}, \dots, W_{t_1} - W_{t_0},$$

are independent for all $0 \leqslant t_0 \leqslant \cdots \leqslant t_n \leqslant 1$.

Let us finally turn to the dependence of $W_t(\omega)$ on t. Note that for $m < n$

$$\begin{aligned}
&\int_0^1 \sup_{t\in[0,1]} \left|S_{2^n-1}(t;\omega) - S_{2^m-1}(t;\omega)\right|^4 d\omega \\
&\quad = \int_0^1 \sup_{t\in[0,1]} \left(\sum_{k=m}^{n-1}\sum_{j=0}^{2^k-1} e_{2^k+j}(\omega)\langle \mathbb{1}_{[0,t]}, H_{2^k+j}\rangle\right)^4 d\omega \\
&\quad = \int_0^1 \sup_{t\in[0,1]} \left(\sum_{k=m}^{n-1} 2^{-\frac{k}{8}}\left[\sum_{j=0}^{2^k-1} 2^{\frac{k}{8}} e_{2^k+j}(\omega)\langle \mathbb{1}_{[0,t]}, H_{2^k+j}\rangle\right]\right)^4 d\omega \\
&\quad \leqslant \int_0^1 \sup_{t\in[0,1]} \underbrace{\left[\sum_{k=m}^{n-1} 2^{-\frac{k}{6}}\right]^3}_{\leqslant 10} \cdot \sum_{k=m}^{n-1}\left[\sum_{j=0}^{2^k-1} 2^{\frac{k}{8}} e_{2^k+j}(\omega)\langle \mathbb{1}_{[0,t]}, H_{2^k+j}\rangle\right]^4 d\omega,
\end{aligned}$$

where we use Hölder's inequality for the outer sum with $p = \frac{4}{3}$ and $q = 4$. Since the functions $F_{2^k+j}(t) = \langle \mathbb{1}_{[0,t]}, H_{2^k+j}\rangle$ with $0 \leqslant j < 2^k$ have disjoint supports and are bounded by $2^{-k/2}$, we find

$$\begin{aligned}
&\int_0^1 \sup_{t\in[0,1]} \left|S_{2^n-1}(t;\omega) - S_{2^m-1}(t;\omega)\right|^4 d\omega \\
&\qquad \leqslant 10 \int_0^1 \sup_{t\in[0,1]} \sum_{k=m}^{n-1}\sum_{j=0}^{2^k-1} 2^{\frac{k}{2}} e^4_{2^k+j}(\omega)\langle \mathbb{1}_{[0,t]}, H_{2^k+j}\rangle^4 d\omega \\
&\qquad \leqslant 10 \sum_{k=m}^{n-1}\sum_{j=0}^{2^k-1} 2^{\frac{k}{2}} \underbrace{\int_0^1 e^4_{2^k+j}(\omega) d\omega}_{= (2\pi)^{-1/2}\int_{\mathbb{R}} y^4 e^{-y^2/2}\,dy\ \forall j,k}\ 2^{-2k} \\
&\qquad = C\sum_{k=m}^{n-1} 2^{\frac{k}{2}} \cdot 2^k \cdot 2^{-2k} \leqslant 2C2^{-\frac{m}{2}},
\end{aligned}$$

which means that the partial sums $S_{2^n-1}(t;\omega)$ of $W_t(\omega)$ converge in $L^4(d\omega)$ uniformly for all $t \in [0, 1]$. By Corollary 13.8 we can extract a subsequence, which converges (uniformly in t) for $\lambda(d\omega)$-almost all ω to $W_t(\omega)$; since for fixed ω

the partial sums $t \mapsto S_{2^n-1}(t;\omega)$ are continuous functions of t, this property is inherited by the a.e. limit $W_t(\omega)$.

The above construction is a variation of a theme by Lévy [27, Chapter I.1, pp. 15–20] and Ciesielski [10]. In form one or another it can be found in many probability textbooks, e.g. Schilling and Partzsch [45, Chapter 3], where also a related construction due to Wiener (see Paley and Wiener [36, Chapter XI]) using random Fourier series is discussed.

Problems

28.1. Prove the orthogonality relation for the Jacobi polynomials 28.1.

28.2. Use the Gram–Schmidt orthonormalization procedure to verify the formulae for the Chebyshev, Legendre, Laguerre and Hermite polynomials given in items 28.1–28.5.

28.3. State and prove Theorem 28.6 and Corollary 28.8 for an arbitrary compact interval $[a,b]$.

28.4. Prove the orthogonality relations (28.4) for the trigonometric system.

28.5. **(i)** Show that for suitable constants $c_k, s_k, \sigma_k \in \mathbb{R}$ and all $n \in \mathbb{N}_0$

$$\cos^n x = \sum_{k=0}^{n} c_k \cos(kx), \ \ \sin^{2n+1} x = \sum_{k=1}^{n} s_k \sin(kx), \ \ \sin^{2n} x = \sum_{k=1}^{n-1} \sigma_k \cos(kx).$$

(ii) Show that for suitable constants $a_k, b_k \in \mathbb{R}$ and all $n \in \mathbb{N}$

$$\cos(nx) = \sum_{k=0}^{n} a_k \cos^{n-k} x \sin^k x \quad \text{and} \quad \sin(nx) = \sum_{k=1}^{n} b_k \cos^{n-k} x \sin^k x.$$

(iii) Deduce that every trigonometric polynomial $T_n(x)$ of order n can be written in the form

$$U_n(x) = \sum_{k,n=0}^{n} \gamma_{k,n} \cos^k x \sin^n x$$

and vice versa.

28.6. Use the formula $\sin a - \sin b = 2\cos((a+b)/2)\sin((a-b)/2)$ to show that $D_N(x)\sin(x/2) = \frac{1}{2}\sin\big(N+\frac{1}{2}\big)x$. This proves (28.11).

28.7. Find the Fourier series expansion for the function $|\sin x|$.

28.8. Let $u(x) = \mathbb{1}_{[0,1)}(x)$. Show that the Haar–Fourier series for u converges for all $1 \leqslant p < \infty$ in L^p-sense to u. Is this true also for the Haar wavelet expansion?

28.9. Show that the Haar–Fourier series for $u \in C_c$ converges uniformly for every x to $u(x)$. Show that this remains true for functions $u \in C_\infty$, i.e. the set of continuous functions such that $\lim_{|x|\to\infty} u(x) = 0$.
[Hint: use the fact that $u \in C_c$ is uniformly continuous. For $u \in C_\infty$ observe that $C_\infty = \overline{C_c}^{\|\cdot\|_\infty}$ (closure in sup-norm) and check that $|s_N(u;x)| \leqslant \|u\|_\infty$.]

28.10. Extend Problem 28.9 to the Haar wavelet expansion.
[Hint: use Problem 28.9 and show that $\|\mathbb{E}^{\mathscr{A}_{-N}^{\Delta}} u\|_\infty \to 0$ for all $u \in C_c(\mathbb{R})$.]

28.11. Let $u(x) = \mathbb{1}_{[0,1/3)}(x)$. Prove that the Haar–Fourier series diverges at $x = \frac{1}{3}$.
[Hint: verify that $\liminf_{N\to\infty} s_N(u, \frac{1}{3}) < \limsup_{N\to\infty} s_N(u, \frac{1}{3})$.]

Appendix A
lim inf and lim sup

For a sequence of real numbers $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ the *limes inferior* or *lower limit* is defined as

$$\liminf_{n\to\infty} a_n := \sup_{k\in\mathbb{N}} \inf_{n\geqslant k} a_n, \tag{A.1}$$

and the *limes superior* or *upper limit* is defined as

$$\limsup_{n\to\infty} a_n := \inf_{k\in\mathbb{N}} \sup_{n\geqslant k} a_n. \tag{A.2}$$

Lower and upper limits are always defined as numbers in $[-\infty, +\infty]$. This is due to the fact that $\big(\inf_{n\geqslant k} a_n\big)_{k\in\mathbb{N}} \subset [-\infty, +\infty)$ and $\big(\sup_{n\geqslant k} a_n\big)_{k\in\mathbb{N}} \subset (-\infty, +\infty]$ are increasing, resp. decreasing sequences, so that the $\sup_{k\in\mathbb{N}}$ and $\inf_{k\in\mathbb{N}}$ in (A.1) and (A.2) are actually (improper) limits $\lim_{k\to\infty}$.

Let us collect a few simple properties of lim inf and lim sup.

Properties A.1 (of lim inf and lim sup) Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences of real numbers.

(i) $\liminf\limits_{n\to\infty} a_n = \lim\limits_{k\to\infty} \inf\limits_{n\geqslant k} a_n$ and $\limsup\limits_{n\to\infty} a_n = \lim\limits_{k\to\infty} \sup\limits_{n\geqslant k} a_n$.

(ii) $\liminf\limits_{n\to\infty} a_n = -\limsup\limits_{n\to\infty}(-a_n)$.

(iii) $\liminf\limits_{n\to\infty} a_n \leqslant \limsup\limits_{n\to\infty} a_n$.

(iv) $\liminf\limits_{n\to\infty} a_n$ and $\limsup\limits_{n\to\infty} a_n$ are limits of subsequences of $(a_n)_{n\in\mathbb{N}}$ and all other limits L of subsequences of $(a_n)_{n\in\mathbb{N}}$ satisfy

$$\liminf_{n\to\infty} a_n \leqslant L \leqslant \limsup_{n\to\infty} a_n.$$

(v) $\lim\limits_{n\to\infty} a_n \in \mathbb{R}$ exists $\iff -\infty < \liminf\limits_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n < +\infty$.
In this case $\lim\limits_{n\to\infty} a_n = \liminf\limits_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n$.

(vi) $\liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n \leqslant \liminf_{n\to\infty}(a_n + b_n)$,
$\limsup_{n\to\infty}(a_n + b_n) \leqslant \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$.

(vii) For all bounded sequences $a_n, b_n \geqslant 0$, $n \in \mathbb{N}$, we have

$$\liminf_{n\to\infty} a_n \liminf_{n\to\infty} b_n \leqslant \liminf_{n\to\infty} a_n b_n,$$
$$\limsup_{n\to\infty} a_n b_n \leqslant \limsup_{n\to\infty} a_n \limsup_{n\to\infty} b_n.$$

(viii) $\liminf_{n\to\infty}(a_n + b_n) \leqslant \liminf_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \leqslant \limsup_{n\to\infty}(a_n + b_n)$.

(ix) For all bounded sequences $a_n, b_n \geqslant 0$, $n \in \mathbb{N}$, we have

$$\liminf_{n\to\infty} a_n b_n \leqslant \liminf_{n\to\infty} a_n \limsup_{n\to\infty} b_n \leqslant \limsup_{n\to\infty} a_n b_n.$$

(x) If the limit $\lim_{n\to\infty} a_n$ exists, then

$$\liminf_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \liminf_{n\to\infty} b_n,$$
$$\limsup_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \limsup_{n\to\infty} b_n.$$

(xi) If $a_n, b_n \geqslant 0$, $n \in \mathbb{N}$, such that $(b_n)_{n\in\mathbb{N}}$ is bounded and $\lim_{n\to\infty} a_n$ exists, then

$$\liminf_{n\to\infty} a_n b_n = \lim_{n\to\infty} a_n \liminf_{n\to\infty} b_n,$$
$$\limsup_{n\to\infty} a_n b_n = \lim_{n\to\infty} a_n \limsup_{n\to\infty} b_n.$$

(xii) $\limsup_{n\to\infty} |a_n| = 0 \implies \lim_{n\to\infty} a_n = 0$.

Proof (i) follows from the remark preceding A.1; (ii) is clear since

$$\inf_n a_n = -\sup_n(-a_n),$$

and (iii) follows from the inequality $\inf_{n\geqslant k} a_n \leqslant \sup_{n\geqslant k} a_n$, where we can pass to the limit $k \to \infty$ on both sides.

Notice that (ii) reduces any statement about lim sup to a dual statement for lim inf. This means that we need to show (iv)–(xi) for the lower limit only.

(iv) Let $(a_{n(k)})_{k\in\mathbb{N}} \subset (a_n)_{n\in\mathbb{N}}$ be some subsequence with the (improper) limit $L = \lim_{n\to\infty} a_{n(n)}$. Then

$$\inf_{n\geqslant k} a_n \leqslant \inf_{i\geqslant k} a_{n(i)} \leqslant L \implies \lim_{k\to\infty} \inf_{n\geqslant k} a_n \leqslant L,$$

i.e. $\liminf_{n\to\infty} a_n$ is smaller than any limit of any subsequence. Let us now construct a subsequence which has $L_* := \liminf_{n\to\infty} a_n > -\infty$ as its limit. By the very definition of L_* and the infimum we find for all $\epsilon > 0$ some $N_\epsilon \in \mathbb{N}$ such that

$$\left|L_* - \inf_{n \geqslant k} a_n\right| \leqslant \epsilon \quad \forall k \geqslant N_\epsilon.$$

Thus $\inf_{n \geqslant k} a_n > -\infty$, and by the definition of the infimum there is some $\ell = \ell_{\epsilon,k}$, $\ell \geqslant k \geqslant N_\epsilon$ and a_ℓ with

$$\left|a_\ell - \inf_{n \geqslant k} a_n\right| \leqslant \epsilon.$$

Specializing to $\epsilon = 1/n$, $n \in \mathbb{N}$, we obtain an infinite family of $a_{\ell(n)}$ from which we can extract a subsequence with limit L_*.

If $L_* = -\infty$, the sequence $(a_n)_{n \in \mathbb{N}}$ is unbounded from below and it is obvious that there must exist a subsequence tending to $-\infty$.

(v) If $\lim_{n\to\infty} a_n$ exists, then all subsequences converge and have the same limit, thus $\liminf_{n\to\infty} a_n = \lim_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$ by (iv).

Conversely, if $L = \liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$, we get for all $k \in \mathbb{N}$

$$0 \leqslant a_k - \inf_{n \geqslant k} a_n \leqslant \sup_{n \geqslant k} a_n - \inf_{n \geqslant k} a_n \xrightarrow[k\to\infty]{} 0,$$

and $\lim_{k\to\infty} a_k = \lim_{k\to\infty} \inf_{n \geqslant k} a_n = L$ follows from a sandwiching argument.

(vi) follows immediately from

$$\inf_{n \geqslant k} a_n + \inf_{n \geqslant k} b_n \leqslant a_\ell + b_\ell \quad \forall \ell \geqslant k \implies \inf_{n \geqslant k} a_n + \inf_{n \geqslant k} b_n \leqslant \inf_{\ell \geqslant k}(a_\ell + b_\ell)$$

if we pass to the limit $k \to \infty$ on both sides.

(vii) We have $0 \leqslant \inf_{n \geqslant k} b_n \leqslant b_\ell$ for all $\ell \geqslant k$ and so

$$\inf_{n \geqslant k} a_n \inf_{n \geqslant k} b_n \leqslant a_\ell b_\ell \quad \forall \ell \geqslant k \implies \inf_{n \geqslant k} a_n \inf_{n \geqslant k} b_n \leqslant \inf_{\ell \geqslant k} a_\ell b_\ell.$$

The assertion follows as we go to the limit $k \to \infty$ on both sides (at this point we need boundedness to exclude the case '$0 \cdot \infty$').

(viii) We have

$$\inf_{n \geqslant k}(a_n + b_n) \leqslant a_\ell + b_\ell \leqslant a_\ell + \sup_{n \geqslant k} b_n \quad \forall \ell \geqslant k,$$

so that $\inf_{n \geqslant k}(a_n + b_n) \leqslant \inf_{n \geqslant k} a_n + \sup_{n \geqslant k} b_n$, and the assertion follows as we go to the limit $k \to \infty$ on both sides.

(ix) is similar to (viii) taking into account the precautions set out in (vii).

(x) If $\lim_{n\to\infty} a_n$ exists, then $\lim_{n\to\infty} a_n = \liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$ according to (v). Thus

$$\begin{aligned}\lim_{n\to\infty} a_n + \liminf_{n\to\infty} b_n \overset{\text{(v)}}{=} \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n &\overset{\text{(vi)}}{\leqslant} \liminf_{n\to\infty}(a_n + b_n) \\ &\overset{\text{(viii)}}{\leqslant} \limsup_{n\to\infty} a_n + \liminf_{n\to\infty} b_n \\ &\overset{\text{(v)}}{\leqslant} \lim_{n\to\infty} a_n + \liminf_{n\to\infty} b_n.\end{aligned}$$

(xi) is similar to (x) using (v), (vii) and (ix).
(xii) Since $|a_n| \geqslant 0$,

$$0 \leqslant \liminf_{n\to\infty} |a_n| \overset{\text{(iii)}}{\leqslant} \limsup_{n\to\infty} |a_n| = 0,$$

and we conclude from (v) that

$$\lim_{n\to\infty} |a_n| = \liminf_{n\to\infty} |a_n| = \limsup_{n\to\infty} |a_n| = 0.$$

Thus $\lim_{n\to\infty} a_n = 0$. □

Sometimes the following definitions for upper and lower limits of a sequence of sets $(A_n)_{n\in\mathbb{N}}$, $A_n \subset X$, are used:

$$\liminf_{n\to\infty} A_n := \bigcup_{k\in\mathbb{N}} \bigcap_{n\geqslant k} A_n \quad \text{and} \quad \limsup_{n\to\infty} A_n := \bigcap_{k\in\mathbb{N}} \bigcup_{n\geqslant k} A_n. \tag{A.3}$$

Here is the connection between set-theoretic and numerical upper and lower limits.

Lemma A.2 *For all $x \in X$ we have*

$$\liminf_{n\to\infty} \mathbb{1}_{A_n}(x) = \mathbb{1}_{\liminf_{n\to\infty} A_n}(x), \tag{A.4}$$

$$\limsup_{n\to\infty} \mathbb{1}_{A_n}(x) = \mathbb{1}_{\limsup_{n\to\infty} A_n}(x). \tag{A.5}$$

Proof Note that

$$\mathbb{1}_{\bigcap_{n\in\mathbb{N}} B_n} = \inf_{n\in\mathbb{N}} \mathbb{1}_{B_n} \quad \text{and} \quad \mathbb{1}_{\bigcup_{n\in\mathbb{N}} B_n} = \sup_{n\in\mathbb{N}} \mathbb{1}_{B_n},$$

which follows from

$$\begin{aligned}\mathbb{1}_{\bigcap_{n\in\mathbb{N}} B_n}(x)=1 &\iff x\in\bigcap_{n\in\mathbb{N}} B_n\\ &\iff \forall n\in\mathbb{N}: x\in B_n\\ &\iff \forall n\in\mathbb{N}: \mathbb{1}_{B_n}(x)=1\\ &\iff \inf_{n\in\mathbb{N}} \mathbb{1}_{B_n}(x)=1.\end{aligned}$$

A similar argument proves the assertion for $\sup_{n\in\mathbb{N}} \mathbb{1}_{B_n}$. Hence,

$$\mathbb{1}_{\liminf\limits_{n\to\infty} A_n}=\mathbb{1}_{\bigcup_{k\in\mathbb{N}}\bigcap_{n\geqslant k} A_n}=\sup_{k\in\mathbb{N}} \mathbb{1}_{\bigcap_{n\geqslant k} A_n}=\sup_{k\in\mathbb{N}}\inf_{n\geqslant k} \mathbb{1}_{A_n}=\liminf_{n\to\infty} \mathbb{1}_{A_n},$$

and (A.5) is proved analogously. □

Appendix B
Some Facts from Topology

Continuity in Euclidean Spaces

Recall that a function $f: \mathbb{R}^m \to \mathbb{R}^n$ is *continuous at the point* $x \in \mathbb{R}^m$ if

$$\forall \epsilon > 0 \, \exists \delta = \delta(\epsilon, x) > 0 \, \forall y, |x - y| < \delta \; : \; |f(x) - f(y)| < \epsilon. \tag{B.1}$$

The function is *continuous*, if it is continuous at all $x \in \mathbb{R}^m$.

Lemma B.1 *Let $f: \mathbb{R}^m \to \mathbb{R}^n$ be a function.*

(i) *f is continuous at $x \in \mathbb{R}^m$ if, and only if,*

$$\forall \epsilon > 0 \, \exists \delta = \delta(\epsilon, x) > 0 \; : \; f(B_\delta(x)) \subset B_\epsilon(f(x)). \tag{B.2}$$

(ii) *f is continuous if, and only if, for every open set $V \in \mathscr{O}_{\mathbb{R}^n}$ the pre-image $f^{-1}(V) \in \mathscr{O}_{\mathbb{R}^m}$ is open in $\mathbb{R}^m$.*

(ii) *f is continuous if, and only if, for every closed set $F \in \mathscr{C}_{\mathbb{R}^n}$ the pre-image $f^{-1}(F) \in \mathscr{C}_{\mathbb{R}^m}$ is closed in $\mathbb{R}^m$.*

Proof (i) Let f be continuous at x. Then

$$\begin{aligned}
&\forall \epsilon > 0 \quad \exists \delta = \delta(\epsilon, x) > 0 \quad \forall y, |x - y| < \delta \; : \; |f(x) - f(y)| < \epsilon \\
\Longleftrightarrow &\forall \epsilon > 0 \quad \exists \delta = \delta(\epsilon, x) > 0 \quad \forall y \in B_\delta(x) \; : \; f(y) \in B_\epsilon(f(x)) \\
\Longleftrightarrow &\forall \epsilon > 0 \quad \exists \delta = \delta(\epsilon, x) > 0 \; : \; f(B_\delta(x)) \subset B_\epsilon(f(x)).
\end{aligned}$$

(ii) '$\Rightarrow$'. Assume that f is continuous. Pick any non-void open set $V \in \mathscr{O}_{\mathbb{R}^n}$ and set $U := f^{-1}(V) \subset \mathbb{R}^m$. For $x \in U$ we have $f(x) \in V$ and, by continuity, for any neighbourhood $B_\epsilon(f(x)) \subset V$ there is some $B_\delta(x)$ such that $f(B_\delta(x)) \subset B_\epsilon(f(x))$. Thus,

$$B_\delta(x) \subset f^{-1}(f(B_\delta(x))) \subset f^{-1}(B_\epsilon(f(x))) \subset f^{-1}(V) \stackrel{\text{def}}{=} U.$$

This shows that $x \in U$ has an open neighbourhood in U, i.e. U is open.

'$\Leftarrow$'. Assume that the pre-image $f^{-1}(V)$ of any open set $V \subset \mathbb{R}^n$ is open. Fix $x \in \mathbb{R}^m$, take any open ball $B_\epsilon(f(x)) \subset \mathbb{R}^n$ and set $U := f^{-1}(B_\epsilon(f(x)))$. By assumption, this is an open set and $x \in U$. Thus, there is some $\delta > 0$ such that $B_\delta(x) \subset U$ and we get

$$f(B_\delta(x)) \subset f(U) \stackrel{\text{def}}{=} f(f^{-1}(B_\epsilon(f(x))) \subset B_\epsilon(f(x)).$$

By part (i) this shows that f is continuous at x, and, since x is arbitrary, we get the continuity of f.

(iii) Let $F \subset \mathbb{R}^n$ be a closed set. Then $V := F^c$ is open, and, since set-operations and inverse images interchange, see (2.6), we get

$$f^{-1}(F) = f^{-1}(V^c) = \left(f^{-1}(V)\right)^c.$$

This shows that the conditions in (ii) and (iii) are equivalent.

□

Metric Spaces

A *metric space* is a set X where we can measure the distance between any two points $x, y \in X$ with the help of a *metric*, i.e. a function $d: X \times X \to [0, \infty]$ such that for all $x, y, z \in X$

$$d(x, y) = 0 \iff x = y \qquad \text{(definiteness)};$$
$$d(x, y) = d(y, x) \qquad \text{(symmetry)};$$
$$d(x, z) \leqslant d(x, y) + d(y, z) \qquad \text{(triangle inequality)}.$$

Every norm $\|x\|$ induces a metric $d(x, y) := \|x - y\|$; in particular $\mathbb{R}$ and $\mathbb{R}^n$ are metric spaces if we use $d(x, y) = |x - y|$ and $d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^2\right)^{1/2}$, respectively.

A metric induces a natural topology. An *open ball* with radius $r > 0$ and centre $x \in X$ is defined by $B_r(x) := \{y \in X : d(x, y) < r\}$. A set $U \subset X$ is *open* (notation: $U \in \mathscr{O}$), if for every $x \in U$ there is some $r = r(x) > 0$ such that $B_r(x) \subset U$. Using the triangle inequality it is not hard to see that open balls are indeed open sets. This topology is, equivalently, described in terms of convergence of sequences. A sequence $(x_n)_{n \in \mathbb{N}} \subset X$ converges in X to $x \in X$ if, and only if, $\lim_{n\to\infty} d(x_n, x) = 0$. Since we have a topology, we can also speak of Borel sets $\mathscr{B}(X) = \sigma(\mathscr{O})$.

A set $A \subset X$ is *relatively compact*, if its closure $\overline{A}$ is compact. A metric space (X, d) is *locally compact*, if each x has a relatively compact open neighbourhood

$V(x)$; without loss of generality we may assume that $V(x) = B_r(x)$ for some small $r = r(x)$.

The distance between a $x \in X$ and $A \subset X$ is given by $d(x, A) := \inf_{a \in A} d(x, a)$. From

$$d(x, A) = \inf_{a \in A} d(x, a) \leqslant \inf_{a \in A} (d(x, y) + d(y, a)) = d(x, y) + d(y, A)$$

we conclude that

$$|d(x, A) - d(y, A)| \leqslant d(x, y) \quad \forall x, y \in X,$$

i.e. $x \mapsto d(x, A)$ is (Lipschitz) continuous. In particular, the Urysohn functions

$$f_{K,U}(x) := \frac{d(x, U^c)}{d(x, U^c) + d(x, K)}, \quad K \subset U \subset X,$$

are continuous and satisfy $\mathbb{1}_K(x) \leqslant f_{K,U}(x) \leqslant \mathbb{1}_U(x)$. The following result is usually called *Urysohn's lemma*.

Lemma B.2 (Urysohn) *Let (X, d) be a locally compact metric space.*

(i) *For every compact set $K \subset X$ there are functions $u_\epsilon \in C_c(X)$ and relatively compact open sets $U_\epsilon \supset K$ such that ${}_K \leqslant u_\epsilon \leqslant 1_{U_\epsilon}$ and $u_\epsilon \downarrow {}_K$ as $\epsilon \to 0$.*

(ii) *For every relatively compact open set $U \subset X$ there are functions $w_\epsilon \in C_c(X)$ such that $0 \leqslant w_\epsilon \leqslant {}_U$ and $w_\epsilon \uparrow {}_U$ as $\epsilon \to 0$.*

Proof (i) Fix a compact set K and $\epsilon > 0$. We cover K with relatively compact, open balls $B_{\epsilon(x)}(x)$ such that $\epsilon(x) \leqslant \epsilon$; at this point we use local compactness of X. Since K is compact, there are finitely many points $x_1, \ldots, x_n \in K$ such that

$$K \subset \bigcup_{i=1}^{n} B_{\epsilon(x_i)}(x_i) =: U_\epsilon.$$

Obviously, $K_\epsilon := \bigcup_{i=1}^{n} \overline{B_{\epsilon(x_i)}(x_i)}$ is compact, $U_\epsilon \subset K_\epsilon$ and $U_\epsilon \to K$. Therefore, the Urysohn functions $u_\epsilon(x) := f_{K,U_\epsilon}(x)$, $\epsilon > 0$, will do the job.

(ii) Let $U \subset X$ be relatively compact and open. Since $x \mapsto d(x, U^c)$ is continuous, the sets $U_\epsilon := \{x \in U : d(x, U^c) > \epsilon\}$ are open and $K_\epsilon := \overline{U_\epsilon} \subset U$ is compact. Again, the Urysohn functions $w_\epsilon(x) = f_{K_\epsilon,U}(x)$ will do the job.

$\square$

Denote by $\widehat{\mathscr{O}}$ the family of relatively compact open sets U and write $\chi \prec \mathbb{1}_U$ if $\chi \leqslant \mathbb{1}_U$ and $\operatorname{supp} \chi \subset U$.

Lemma B.3 (partition of unity) *Let (X,d) be a locally compact metric space and $K\subset X$ a compact set which is covered by finitely many open sets: $K\subset\bigcup_{i=1}^n U_i$. There exist functions $\chi_1,\dots,\chi_n\in C_c(X)$ such that*

$$0\leqslant\chi_i\prec_{U_i}\quad\textit{and}\quad\sum_{i=1}^n\chi_i(x)=1\quad\forall x\in K.$$

Proof Since X is locally compact and $K\subset\bigcup_{i=1}^n U_i$, we find for all $x\in K$ some $i(x)\in\{1,\dots,n\}$ and a relatively compact open set $V(x)\in\widehat{\mathscr{O}}$ such that

$$x\in V(x)\subset\overline{V(x)}\subset U_{i(x)}.$$

Thus, we have $K\subset\bigcup_{x\in K}V(x)$ and, by compactness, there exists a finite subcover $K\subset V(x_1)\cup\dots\cup V(x_m)$. Define

$$W_i:=\bigcup_{k:\,V(x_k)\subset U_i}V(x_k)\implies\overline{W_i}=\bigcup_{k:\,V(x_k)\subset U_i}\overline{V(x_k)}\subset U_i.$$

Since the $V(x)$ are relatively compact, $\overline{W}_i$ is compact, and the sets $\overline{W}_i$ and U_i^c have strictly positive distance $\inf_{x\in\overline{W}_i,y\in U_i^c}d(x,y)>0$. Therefore, Urysohn's lemma (Lemma B.2) shows that there is a function

$$\phi_i\in C_c(X)\quad\text{such that}\quad\mathbb{1}_{\overline{W_i}}\leqslant\phi_i\prec\mathbb{1}_{U_i}.$$

Set

$$\chi_i=(1-\phi_1)\cdots(1-\phi_{i-1})\phi_i,\quad i=1,\dots,n,$$

and observe that $0\leqslant\chi_i\prec\mathbb{1}_{U_i}$ as well as

$$\begin{aligned}(1-\phi_1)\cdots(1-\phi_{n-1})(1-\phi_n)&=(1-\phi_1)\cdots(1-\phi_{n-1})-\chi_n\\&=(1-\phi_1)\cdots(1-\phi_{n-2})-\chi_{n-1}-\chi_n\\&=\dots=1-\chi_1-\chi_2-\dots-\chi_n.\end{aligned}$$

Since $K\subset\bigcup_{i=1}^n W_i$, we have $(1-\phi_1(x))\cdots(1-\phi_n(x))=0$ for any $x\in K$. □

A metric space is called σ-compact if there exists a sequence of compact sets $K_n\uparrow X$. The next lemma shows that we can change the K_ns in such a way that we can 'catch' every fixed compact set $K\subset X$.

Lemma B.4 (absorption property) *Let (X,d) be a locally compact and σ-compact metric space. There is a sequence of compact sets $L_n\subset L_{n+1}^\circ\subset L_{n+1}\uparrow X$. In particular, every compact set $K\subset X$ is contained in some L_N, $N=N(K)\in\mathbb{N}$.*

Proof Since X is σ-compact, there is a sequence of compact sets $K_n \uparrow X$. Using Urysohn's lemma (Theorem B.2(i)) we can construct functions $\chi_n \in C_c(X)$ such that $\mathbb{1}_{K_n} \leqslant \chi_n \leqslant 1$; if needed, we can use $\max\{\chi_1, \dots, \chi_n\}$ to make sure that $\chi_n \uparrow 1$.

Define $L_n := \{\chi_n \geqslant 1/n\}$. Since χ_n is continuous, L_n is closed and, being contained in the compact set $\operatorname{supp} \chi_n$, even compact. From $\chi_n \leqslant \chi_{n+1}$ we get

$$L_n = \left\{\chi_n \geqslant \frac{1}{n}\right\} \subset \left\{\chi_n > \frac{1}{n+1}\right\} \subset \underbrace{\left\{\chi_{n+1} > \frac{1}{n+1}\right\}}_{=:U_n \text{ open as } \chi_n \text{ is continuous}} \subset \left\{\chi_{n+1} \geqslant \frac{1}{n+1}\right\} = L_{n+1}.$$

Since U_n is open and L_{n+1}° is by definition the largest open set inside L_{n+1}, we have $L_n \subset U_n \subset L_{n+1}^\circ \subset L_{n+1}$ and $\chi_n \uparrow 1$ entails $L_n^\circ \uparrow X$.

Let $K \subset X$ be a compact set. Clearly, $K \subset \bigcup_{n \in \mathbb{N}} L_n^\circ$ and, because of compactness, there is a finite sub-cover, i.e. there is some index $N = N(K)$ such that $K \subset L_1 \cup \cdots \cup L_N = L_N$. □

Let (X, d) be a locally compact metric space. We have seen that the open balls $B_r(x)$ are the building blocks for the topology on X. In fact, we can represent every open set U as a union of open balls: $U = \bigcup_{x \in U,\, B_r(x) \subset U} B_r(x)$. If X is separable, i.e. contains a countable dense subset D, then we can even write U as a *countable* union of balls. The family $\{B_r(x) : r \in \mathbb{Q}^+, x \in D\}$ is called a *countable basis* of the topology.

Theorem B.5 *If (X, d) is a locally compact, separable metric space, then $C_c(X)$ is separable, i.e. it contains a uniformly dense subset.*

Proof Let $\mathscr{G}$ be a countable basis of $\mathscr{O}$ and $\mathscr{J} := \{(a, b) : a < b,\ a, b \in \mathbb{Q}\}$. Let $G_1, \dots, G_n \in \mathscr{G}$ and $I_1, \dots, I_n \in \mathscr{J}$. We call a function $f \in C_c(X)$ satisfying

$$f(G_i) \subset I_i \quad (i = 1, \dots, n) \quad \text{and} \quad \operatorname{supp} f \subset G_1 \cup \cdots \cup G_n$$

$(G_i, I_i)_{i=1,\dots,n}$-adapted. For any such tuple we fix some adapted function (provided there is one) and we denote by $\mathcal{F}$ the collection of all adapted functions.

Since $\bigcup_{n \in \mathbb{N}} \mathscr{G}^n \times \mathscr{J}^n$ is countable, the family $\mathcal{F}$ is at most countable. For every $u \in C_c(X)$ we have

$$\forall x \in X \quad \forall \epsilon > 0 \quad \exists U(x) \in \mathscr{G},\ x \in U(x) \quad \forall y \in U(x) : |u(x) - u(y)| < \epsilon.$$

Since $\operatorname{supp} u$ is compact, $\operatorname{supp} u$ is covered by finitely many $U(x_1) \cup \cdots \cup U(x_n)$. On the other hand, we have for all $i = 1, \dots, n$

$$\sup_{x, y \in U(x_i)} |u(x) - u(y)| < 2\epsilon \Longrightarrow \exists J_i \in \mathscr{J},\ \lambda^1(J_i) < 3\epsilon : u(U(x_i)) \subset J_i.$$

This shows that u is adapted for $(U(x_i), J_i)_{i=1,\dots,n}$. Assume that $f \in \mathcal{F}$ is also $(U(x_i), J_i)_{i=1,\dots,n}$-adapted. Then

$$\sup_{x \in U(x_i)} |u(x) - f(x)| < 3\epsilon \quad \forall i = 1, \dots, n \quad \text{and} \quad f = u = 0 \text{ if } x \notin \bigcup_{i=1}^{n} U(x_i).$$

This shows that $\|f - u\|_\infty < 3\epsilon$, i.e. $\mathcal{F}$ is a countable dense subset of $C_c(X)$. $\square$

Appendix C
The Volume of a Parallelepiped

In this appendix we give a simple derivation for the volume of the parallelepiped

$$A\big([0,1)^n\big) := \big\{Ax \in \mathbb{R}^n : x \in [0,1)^n\big\}, \quad A \in \mathbb{R}^{n\times n},\ \det A \neq 0,$$

determined by a non-degenerate $n \times n$ matrix $A \in \mathbb{R}^{n\times n}$.

Theorem C.1 $\lambda^n\big[A([0,1)^n)\big] = |\det A|$ *for all invertible matrices* $A \in \mathbb{R}^{n\times n}$.

The proof of Theorem C.1 requires two auxiliary results.

Lemma C.2 *If* $D = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$, $\lambda_i > 0$, *is a diagonal* $n \times n$ *matrix, then* $\lambda^n(D(B)) = \det D\lambda^n(B)$ *for all Borel sets* $B \in \mathscr{B}(\mathbb{R}^n)$.

Proof Since D and D^{-1} are continuous maps, $D(B)$ is a Borel set if $B \in \mathscr{B}(\mathbb{R}^n)$, see Example 7.3. In view of the uniqueness theorem for measures (Theorem 5.7) it suffices to prove the lemma for half-open rectangles $[\![a, b)\!)$, $a, b \in \mathbb{R}^n$. Obviously,

$$D\,[\![a,b)\!) = \bigtimes_{i=1}^{n} [\lambda_i a_i, \lambda_i b_i),$$

and

$$\lambda^n\big(D\,[\![a,b)\!)\big) = \prod_{i=1}^{n}(\lambda_i b_i - \lambda_i a_i) = \lambda_1 \cdot \dots \cdot \lambda_n \prod_{i=1}^{n}(b_i - a_i)$$
$$= \det D\lambda^n\,[\![a,b)\!). \qquad \square$$

Lemma C.3 *Every invertible matrix* $A \in \mathbb{R}^{n\times n}$ *can be written as* $A = SDT$, *where* S, T *are orthogonal* $n \times n$ *matrices and* $D = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$ *is a diagonal matrix with positive entries* $\lambda_i > 0$.

Proof The matrix $A^\top A$ is symmetric and so we can find some orthogonal matrix[1] $U \in \mathbb{R}^{n\times n}$ such that

$$U^\top (A^\top A) U = \tilde{D} = \operatorname{diag}[\mu_1, \dots, \mu_n].$$

Since for $e_i := (\underbrace{0, \dots, 0, 1,}_{i} 0 \dots, 0)$ and the Euclidean norm $\|\cdot\|$

$$\mu_i = e_i^\top \tilde{D} e_i = (e_i^\top U^\top A^\top)(AUe_i) = \|AUe_i\|^2 > 0,$$

we can define $D := \sqrt{\tilde{D}} = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$, where $\lambda_i := \sqrt{\mu_i}$. Thus,

$$D^{-1} U^\top A^\top A U D^{-1} = \operatorname{id}_n,$$

and this proves that $S := AUD^{-1}$ is an orthogonal matrix. Since $T := U^\top$ is orthogonal, we easily see that

$$SDT = (AUD^{-1}) DU^\top = A. \qquad \square$$

Proof of Theorem C.1 We have for an invertible matrix $A \in \mathbb{R}^{n\times n}$

$$\begin{aligned} \lambda^n\big[A([0,1)^n)\big] &\overset{\text{C.3}}{=} \lambda^n\big[SDT([0,1)^n)\big] \\ &\overset{7.9}{=} \lambda^n\big[DT([0,1)^n)\big] \\ &\overset{\text{C.2}}{=} \det D \lambda^n\big[T([0,1)^n)\big] \\ &\overset{\text{C.3}}{=} \det D \lambda^n\big([0,1)^n\big). \end{aligned}$$

Since S, T are orthogonal matrices, their determinants are either $+1$ or -1, and we conclude that $|\det A| = |\det(SDT)| = |\det S| \cdot |\det D| \cdot |\det T| = \det D$. $\square$

[1] Recall that U is orthogonal, if $U^\top = U^{-1}$. In particular, $|\det U| = 1$.

Appendix D
The Integral of Complex-Valued Functions

It is often necessary to consider complex-valued functions $u : X \to \mathbb{C}$ on a measurable space $(X, \mathscr{A})$. Since $\mathbb{C}$ is a normed space, we have a natural topology on $\mathbb{C}$ and we may consider the Borel σ-algebra $\mathscr{B}(\mathbb{C})$ on $\mathbb{C}$. Recall that

$$z = x + iy, \quad \bar{z} = x - iy, \quad x = \operatorname{Re} z = \tfrac{1}{2}(z + \bar{z}), \quad y = \operatorname{Im} z = \tfrac{1}{2i}(z - \bar{z}),$$

which shows that the maps $\phi : \mathbb{R}^2 \to \mathbb{C}$ and $\psi : \mathbb{C} \to \mathbb{R}^2$ given by

$$\phi(x, y) := x + iy \quad \text{and} \quad \psi(z) := (\operatorname{Re} z, \operatorname{Im} z)$$

are continuous and inverses to each other. In particular, we can identify open sets in $\mathbb{C}$ with open sets in $\mathbb{R}^2$, and vice versa. Since continuous functions are Borel measurable, this identification extends to the Borel σ-algebras. Consequently,

$$\left.\begin{array}{l} f : X \to \mathbb{C} \quad \text{is} \\ \mathscr{A}/\mathscr{B}(\mathbb{C})\text{-measurable} \end{array}\right\} \iff \left\{\begin{array}{l} \operatorname{Re} f,\ \operatorname{Im} f : X \to \mathbb{R} \quad \text{are} \\ \mathscr{A}/\mathscr{B}(\mathbb{R})\text{-measurable.} \end{array}\right. \tag{D.1}$$

To see '$\Rightarrow$' note that the maps $\operatorname{Re} : z \mapsto \frac{1}{2}(z + \bar{z})$ and $\operatorname{Im} : z \mapsto \frac{1}{2i}(z - \bar{z})$ are continuous, and hence measurable, and so are, by Theorem 7.4, the compositions $\operatorname{Re} \circ f$ and $\operatorname{Im} \circ f$.

Conversely, '$\Leftarrow$' follows – if we write $f = u + iv$ – from the formula

$$f^{-1}[\![z, w)\!) = \underbrace{u^{-1}\big([\operatorname{Re} z, \operatorname{Re} w)\big)}_{\in \mathscr{A}} \cap \underbrace{v^{-1}\big([\operatorname{Im} z, \operatorname{Im} w)\big)}_{\in \mathscr{A}} \in \mathscr{A}$$

and the fact that the rectangles

$$[\![z, w)\!) := \{x + iy : \operatorname{Re} z \leqslant x < \operatorname{Re} w,\ \operatorname{Im} z \leqslant y < \operatorname{Im} w\}$$

generate $\mathscr{B}(\mathbb{C})$.

This means that we can define the integral of a $\mathbb{C}$-valued measurable function by linearity

$$\int f\,d\mu := \int \mathrm{Re} f\,d\mu + i\int \mathrm{Im} f\,d\mu; \tag{D.2}$$

we call $f: X \to \mathbb{C}$ *integrable* and write $f \in \mathcal{L}^1_{\mathbb{C}}(\mu)$ if $\mathrm{Re} f, \mathrm{Im} f: X \to \mathbb{R}$ are integrable in the usual sense. The following rules for $f \in \mathcal{L}^1_{\mathbb{C}}(\mu)$ are readily checked:

$$\mathrm{Re}\int f\,d\mu = \int \mathrm{Re} f\,d\mu, \quad \mathrm{Im}\int f\,d\mu = \int \mathrm{Im} f\,d\mu, \quad \overline{\int f\,d\mu} = \int \bar{f}\,d\mu, \tag{D.3}$$

$$f \in \mathcal{L}^1_{\mathbb{C}}(\mu) \iff f \in \mathcal{M}(\mathscr{B}(\mathbb{C})) \text{ and } |f| \in \mathcal{L}^1_{\mathbb{R}}(\mu). \tag{D.4}$$

In (D.4) the direction '⇒' follows since $|f| = \big((\mathrm{Re} f)^2 + (\mathrm{Im} f)^2\big)^{1/2}$ is measurable and $|f| \leqslant |\mathrm{Re} f| + |\mathrm{Im} f|$, while '⇐' is implied by $|\mathrm{Re} f|, |\mathrm{Im} f| \leqslant |f|$.

The equivalence (D.4) can be used to show that $\mathcal{L}^1_{\mathbb{C}}(\mu)$ is a $\mathbb{C}$-vector space: for $f, g \in \mathcal{L}^1_{\mathbb{C}}(\mu)$ and $\alpha, \beta \in \mathbb{C}$ we have $\alpha f + \beta g \in \mathcal{L}^1_{\mathbb{C}}(\mu)$, in which case

$$\int (\alpha f + \beta g)\,d\mu = \alpha\int f\,d\mu + \beta\int g\,d\mu; \tag{D.5}$$

moreover, we have the following *triangle estimate*:

$$\left|\int f\,d\mu\right| \leqslant \int |f|\,d\mu. \tag{D.6}$$

Only (D.6) is not entirely straightforward. Since $\int f\,d\mu \in \mathbb{C}$, we can find some $\theta \in [0, 2\pi)$ such that

$$\begin{aligned} \left|\int f\,d\mu\right| = e^{i\theta}\int f\,d\mu \;&=\; \mathrm{Re}\left(e^{i\theta}\int f\,d\mu\right) \\ &\overset{\text{(D.3)}}{\underset{\text{(D.5)}}{=}} \int \mathrm{Re}\big(e^{i\theta} f\big)\,d\mu \\ &\leqslant\; \int \left|e^{i\theta} f\right| d\mu = \int |f|\,d\mu. \end{aligned}$$

The spaces $\mathcal{L}^p_{\mathbb{C}}(\mu)$, $1 < p \leqslant \infty$, can now be defined by

$$\mathcal{L}^p_{\mathbb{C}}(\mu) := \big\{f \in \mathcal{M}(\mathscr{B}(\mathbb{C})) : |f| \in \mathcal{L}^p_{\mathbb{R}}(\mu)\big\}, \tag{D.7}$$

and it is obvious that all of the assertions from Chapter 13 remain valid. In particular, $L^p_{\mathbb{C}}(\mu)$ stands for the set of all equivalence classes of $\mathcal{L}^p_{\mathbb{C}}(\mu)$-functions if we identify functions which coincide outside some μ-null set. Note also that most of our results on $\mathbb{R}$-valued integrands carry over to $\mathbb{C}$-valued functions on considering real and imaginary parts separately.

Appendix E

Measurability of the Continuity Points of a Function

Let (X, d) be a metric space and write

$$B_r(x) := \{y \in X : d(x, y) < r\}$$

for the open ball with radius $r > 0$ and centre $x \in X$. For any – not necessarily measurable – function $u : X \to \mathbb{R}$ we define the following modulus of continuity:

$$w^u(x) := \inf_{r>0} (\operatorname{diam} u\,(B_r(x))) = \inf_{r>0} \left(\sup_{z \in B_r(x)} u(z) - \inf_{z \in B_r(x)} u(z) \right),$$

where $\operatorname{diam} B = \sup_{x,y \in X} |x - y|$ denotes the diameter of the set $B \subset \mathbb{R}$. Since $r \mapsto \operatorname{diam} u(B_r(x))$ is a decreasing function, we may replace $\inf_{r>0}$ in the above definition by $\inf_{0<r<\delta}$.

Lemma E.1 *The function u is continuous at x if, and only if,* $w^u(x) = 0$.

Proof '$\Rightarrow$'. Assume that u is continuous at the point x. By the definition of continuity, there is for every $\epsilon > 0$ some $r(\epsilon) > 0$ such that

$$\left(\sup_{z \in B_r(x)} u(z) - u(x) \right) + \left(u(x) - \inf_{z \in B_r(x)} u(z) \right) < 2\epsilon \quad \forall r < r(\epsilon).$$

Therefore,

$$w^u(x) \leqslant \sup_{z \in B_r(x)} u(z) - \inf_{z \in B_r(x)} u(z) < 2\epsilon \xrightarrow[\epsilon \to 0]{} 0.$$

'$\Leftarrow$'. For all $r > 0$ and x, x' such that $d(x, x') < r$ we find

$$u(x) - u(x') \leqslant \sup_{z \in B_r(x)} u(z) - \inf_{z \in B_r(x)} u(z),$$

and since x and x' play symmetric roles, we conclude that

$$|u(x) - u(x')| \leqslant \sup_{z \in B_r(x)} u(z) - \inf_{z \in B_r(x)} u(z).$$

Assume that $w^u(x) = 0$. This means that we can find for every $\epsilon > 0$ some $r(\epsilon)$ such that

$$|u(x) - u(x')| \leqslant \sup_{z \in B_r(x)} u(z) - \inf_{z \in B_r(x)} u(z) \leqslant \epsilon + w^u(x) = \epsilon$$

holds for all $r < r_\epsilon$ and $x' \in B_r(x)$. We conclude that u is continuous at x. □

Lemma E.2 *w^u is upper semi-continuous, i.e. the set $\{w^u < \alpha\}$ is open for every $\alpha > 0$.*

Proof For every $x_0 \in \{w^u < \alpha\}$ there is some $r = r(\alpha) > 0$ such that

$$\sup_{z \in B_r(x_0)} u(z) - \inf_{z \in B_r(x_0)} u(z) < \alpha.$$

Pick $y \in B_{r/3}(x_0)$. Since $B_{r/3}(y) \subset B_r(x_0)$, we get

$$w^u(y) \leqslant \sup_{z \in B_{r/3}(y)} u(z) - \inf_{z \in B_{r/3}(y)} u(z) \leqslant \sup_{z \in B_r(x_0)} u(z) - \inf_{z \in B_r(x_0)} u(z) < \alpha.$$

This shows that $y \in \{w^u < \alpha\}$, and so $B_{r/3}(x_0) \subset \{w^u < \alpha\}$. □

Theorem E.3 *The set of continuity points $C^u := \{x : u$ is continuous at $x\}$ of an arbitrary function $u : X \to \mathbb{R}$ is a Borel set.*

Proof Lemma E.1 shows that

$$C^u = \bigcap_{\delta > 0} \{w^u < \delta\} = \bigcap_{n \in \mathbb{N}} \left\{ w^u < \frac{1}{n} \right\}.$$

By Lemma E.2, the sets $\left\{w^u < \frac{1}{n}\right\}$ are open. Thus, C^u is the countable intersection of Borel sets, and hence a Borel set. In fact, C^u is even a G_δ-set: the countable intersection of open sets. □

Appendix F
Vitali's Covering Theorem

Consider Lebesgue λ^n on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and write $K_r(x) := \overline{B_r(x)}$ for the closed ball with centre $x \in \mathbb{R}^n$ and radius $r > 0$. If $K = K_r(x)$, then $\widehat{K} = K_{5r}(x)$ denotes the concentric closed ball whose radius is five times the radius of K. Vitali's covering theorem shows that we can approximate any open set $U \subset \mathbb{R}^n$ from the inside by countably many mutually disjoint closed balls K such that the difference of U and the union of the balls becomes a Lebesgue null set. The proof which follows is an adaptation from Evans and Gariepy [16, Section 1.5.1]. Here we denote by $\operatorname{diam} B = \sup_{x,y \in B} |x - y|$ the diameter of the set B.

Theorem F.1 (Vitali) *Let $U \subset \mathbb{R}^n$ be an open set. For every $\delta > 0$ there is a countable family $\mathscr{G}$ of mutually disjoint closed balls $K \subset U$ such that* $\operatorname{diam} K \leqslant \delta$ *and* $\lambda^n \left(U \setminus \dot{\bigcup}_{K \in \mathscr{G}} K\right) = 0$.

Proof Fix $\delta > 0$ and denote by $\mathscr{F}_U = \{K_r(x) : K_r(x) \subset U,\ 2r \leqslant \delta\}$ the family of all closed balls in U with diameter less or equal to δ.

Step 1. Claim. There exists a countable family $\mathscr{G}_U \subset \mathscr{F}_U$ of disjoint balls such that $\bigcup_{K \in \mathscr{F}_U} K \subset \bigcup_{K \in \mathscr{G}_U} \widehat{K}$. Set $\mathscr{F}_k = \{K \in \mathscr{F}_U : \delta 2^{-k} < \operatorname{diam} K \leqslant \delta 2^{-k+1}\}$, $k \in \mathbb{N}$, and define the families $\mathscr{G}_k$ recursively.

- $\mathscr{G}_1$ is any maximal collection of disjoint closed balls from $\mathscr{F}_1$. Since each $K \in \mathscr{G}_1$ contains at least one element $q \in \mathbb{Q}^n$, $\mathscr{G}_1$ is at most countable.
- Assume that $\mathscr{G}_1, \ldots, \mathscr{G}_{k-1}$ have been constructed. $\mathscr{G}_k$ is any maximal collection of disjoint closed balls from the family

$$\left\{K \in \mathscr{F}_k : K \cap K' = \emptyset \quad \forall K' \in \mathscr{G}_1 \cup \cdots \cup \mathscr{G}_{k-1}\right\}.$$

With the same argument as for $\mathscr{G}_1$ we see that $\mathscr{G}_k$ is at most countable, too.

By construction, $\mathscr{G}_U := \bigcup_{k \in \mathbb{N}} \mathscr{G}_k$ is a countable family of mutually disjoint closed balls inside U and of diameter $\operatorname{diam} K \leqslant \delta$. Moreover, for every $K \in \mathscr{F}_U$

there is some $K' \in \mathscr{G}_U$ such that $K \cap K' \neq \emptyset$ and $K \subset \widehat{K'}$. *Indeed:* $K \in \mathscr{G}_k$ for some k, and there is some $K' \in \mathscr{G}_1 \cup \cdots \cup \mathscr{G}_k$ with $K \cap K' \neq \emptyset$ – otherwise the family $\mathscr{G}_k$ would not be maximal as we could add K. Moreover,

$$\operatorname{diam} K \leqslant \delta 2^{-k+1} = 2\delta 2^{-k} \leqslant 2 \operatorname{diam} K' \implies K \subset \widehat{K'}.$$

Step 2. Assume that $\lambda^n(U) < \infty$ and construct $\mathscr{G}_U$ for $\mathscr{F}_U$ as in Step 1. We have

$$\lambda^n(U) \leqslant \sum_{K \in \mathscr{G}_U} \lambda^n\big(\widehat{K}\big) = 5^n \sum_{K \in \mathscr{G}_U} \lambda^n(K) = 5^n \lambda^n\bigg(\dot{\bigcup_{K \in \mathscr{G}_U}} K \bigg).$$

This means that

$$\lambda^n\bigg(\dot{\bigcup_{K \in \mathscr{G}_U}} K \bigg) \geqslant 5^{-n} \lambda^n(U) \quad \text{and so} \quad \lambda^n\bigg(U \setminus \dot{\bigcup_{K \in \mathscr{G}_U}} K \bigg) \leqslant (1 - 5^{-n}) \lambda^n(U).$$

Upon enlarging the factor a bit, say $1 > \theta > 1 - 5^{-n}$, we can use the continuity of measures to find some $M_1 \in \mathbb{N}$ such that

$$\lambda^n\bigg(U \setminus \dot{\bigcup_{i=1}^{M_1}} K_i \bigg) \leqslant \theta \lambda^n(U).$$

Step 3. Consider the open set $U_2 := U \setminus \dot{\bigcup}_{i=1}^{M_1} K_i$ and the family $\mathscr{F}_{U_2}$. Using Step 1 again, we get a countable family $\mathscr{G}_{U_2}$ and, as in Step 2, an integer $M_2 > M_1$ such that

$$\lambda^n\bigg(U \setminus \dot{\bigcup_{i=1}^{M_2}} K_i \bigg) = \lambda^n\bigg(U_2 \setminus \dot{\bigcup_{i=M_1+1}^{M_2}} K_i \bigg) \leqslant \theta \lambda^n(U_2) \leqslant \theta^2 \lambda^n(U).$$

Step 4. By iterating Step 3 we get indices $M_k > M_{k-1} > \cdots > M_1$ with

$$\lambda^n\bigg(U \setminus \dot{\bigcup_{i=1}^{M_k}} K_i \bigg) \leqslant \theta^k \lambda^n(U) \xrightarrow[k \to \infty]{(\theta < 1)} 0.$$

Step 5. If $\lambda^n(U) = \infty$, we define open sets $U_i = U \cap \{i - 1 < |x| < i\}$, $i \in \mathbb{N}$. Since $\lambda^n(U_i) < \infty$, for each $i \in \mathbb{N}$ there is a countable family of disjoint closed balls $\mathscr{G}_i$ approximating U_i. Observe that $\lambda^n\left(U \setminus \dot{\bigcup}_{i \in \mathbb{N}} U_i\right) = 0$ and $\mathscr{G} := \bigcup_{i \in \mathbb{N}} \mathscr{G}_i$ is again countable. Therefore,

$$\lambda^n\bigg(U \setminus \dot{\bigcup_{K \in \mathscr{G}}} K \bigg) = \lambda^n\bigg(\dot{\bigcup_{i \in \mathbb{N}}} U_i \setminus \dot{\bigcup_{i \in \mathbb{N}}} \bigcup_{K \in \mathscr{G}_i} K \bigg) \leqslant \sum_{i=1}^{\infty} \underbrace{\lambda^n\bigg(U_i \setminus \bigcup_{K \in \mathscr{G}_i} K \bigg)}_{=0} = 0. \quad \square$$

Appendix G
Non-measurable Sets

Let $(X, \mathscr{A}, \mu)$ be a measure space and denote by $(X, \overline{\mathscr{A}}, \bar{\mu})$ its completion, see Problem 4.15 for the definition and Problems 6.4, 11.5, 11.6, 14.15 and 16.3 for various properties. Here we need only that

$$\overline{\mathscr{A}} = \{A \cup N : A \in \mathscr{A},\ N \text{ is a subset of some } \mathscr{A}\text{-measurable } \mu\text{-null set}\}$$

is the completion of $\mathscr{A}$ with respect to the measure μ. It is natural to ask how big $\mathscr{A}$ and $\overline{\mathscr{A}}$ are and whether $\mathscr{A} \subset \overline{\mathscr{A}} \subset \mathscr{P}(X)$ are proper inclusions.

Sometimes, see Problems 6.13 and 6.14, these questions are easy to answer. For the Borel σ-algebra $\mathscr{A} = \mathscr{B}(\mathbb{R}^n)$ and Lebesgue measure $\mu = \lambda^n$ this is more difficult. The following definition helps to distinguish between sets in $\mathscr{B}(\mathbb{R}^n)$ and the completion $\overline{\mathscr{B}}(\mathbb{R}^n)$ w.r.t. Lebesgue measure.

Definition G.1 The *Lebesgue σ-algebra* is the completion $\overline{\mathscr{B}}(\mathbb{R}^n)$ of the Borel σ-algebra w.r.t. Lebesgue measure λ^n. A set $B \in \overline{\mathscr{B}}(\mathbb{R}^n)$ is called *Lebesgue measurable*.

The next theorem shows that there are 'as many' Lebesgue measurable sets as there are subsets of $\mathbb{R}^n$.

Theorem G.2 *We have* $\#\overline{\mathscr{B}}(\mathbb{R}^n) = \#\mathscr{P}(\mathbb{R}^n)$ *for all* $n \in \mathbb{N}$.

Proof Since $\overline{\mathscr{B}}(\mathbb{R}^n) \subset \mathscr{P}(\mathbb{R}^n)$ we have that $\#\overline{\mathscr{B}}(\mathbb{R}^n) \leqslant \#\mathscr{P}(\mathbb{R}^n)$. On the other hand, we have seen in Problem 7.12 that the Cantor ternary set C is an uncountable Borel measurable λ^1-null set of cardinality $\#\mathbb{R} = \mathfrak{c}$. Consequently, $\mathbb{R}^{n-1} \times C$ is a λ^n-null set. By definition of the Lebesgue σ-algebra, all sets in $\mathscr{P}(\mathbb{R}^{n-1} \times C)$ are Lebesgue measurable (null) sets, i.e. $\mathscr{P}(\mathbb{R}^{n-1} \times C) \subset \overline{\mathscr{B}}(\mathbb{R}^n)$, and therefore $\#\mathscr{P}(\mathbb{R}^{n-1} \times C) \leqslant \#\overline{\mathscr{B}}(\mathbb{R}^n)$. Using the fact that there is a bijection between C and $\mathbb{R}$ we also get $\#\mathscr{P}(\mathbb{R}^n) \leqslant \#\mathscr{P}(\mathbb{R}^{n-1} \times C) \leqslant \#\overline{\mathscr{B}}(\mathbb{R}^n)$, and the Cantor–Bernstein theorem (Theorem 2.7) proves that $\#\mathscr{P}(\mathbb{R}^n) = \#\overline{\mathscr{B}}(\mathbb{R}^n)$. □

Unfortunately, we cannot use Theorem G.2 to decide whether there are sets which are not Lebesgue measurable. To answer this question we need the axiom of choice.

Axiom of choice G.3 (AC) *Let $\{M_i : i \in I\}$ be a collection of non-empty and mutually disjoint subsets of X. Then there exists a set $L \subset \bigcup_{i\in I} M_i$ which contains exactly one element from each set M_i, $i \in I$.*

Note that AC only asserts the existence of the set L but does not tell us how or even whether the set L can be constructed. This problem is at the heart of the controversy over whether one should or should not accept AC.

Theorem G.4 *Assuming the axiom of choice, there exist non-Lebesgue measurable sets in $\mathbb{R}$ with strictly positive outer Lebesgue measure.*

Proof We will construct a non-Lebesgue measurable subset of $\mathbb{J} = [0, 1)$. We call any two $x, y \in \mathbb{J}$ equivalent, $x \sim y$, if $x - y \in \mathbb{Q}$. The equivalence class containing x is denoted by

$$[x] = \{y \in \mathbb{J} : x - y \in \mathbb{Q}\} = (x + \mathbb{Q}) \cap \mathbb{J}.$$

By construction, $\mathbb{J}$ is partitioned by a family of mutually disjoint equivalence classes $[x_i]$, $i \in I$. The axiom of choice[1] shows that there exists a set L which contains exactly one element, say m_i, from each of the classes $[x_i]$, $i \in I$. We will prove that L cannot be Lebesgue measurable.

Assume L is Lebesgue measurable. Since $[x] \cap L$, $x \in \mathbb{J}$, contains exactly one element, say m_{i_0} for some $i_0 = i_0(x) \in I$, we can then find some $q \in \mathbb{Q}$ such that $x = m_{i_0} + q$. Obviously, $-1 < q < 1$. Thus

$$\mathbb{J} \subset L + \big(\mathbb{Q} \cap (-1, 1)\big) \subset \mathbb{J} + (-1, 1) = [-1, 2),$$

which we can rewrite as

$$[0, 1) \subset \bigcup_{q\in\mathbb{Q}\cap(-1,1)} (q + L) \subset [-1, 2).$$

Moreover, $(r + L) \cap (q + L) = \emptyset$ for all $r \neq q$, $r, q \in \mathbb{Q}$. Otherwise $r + x = q + y$ for $x, y \in L$, so that $x \sim y$, which is impossible since L contains only one representative of each equivalence class. Therefore we can use the σ sub-additivity of the measure $\bar{\lambda}^1$ to find

$$1 = \bar{\lambda}^1[0, 1) \leqslant \sum_{q\in\mathbb{Q}\cap(-1,1)} \bar{\lambda}^1(q + L) \leqslant \bar{\lambda}^1[-1, 2) = 3.$$

[1] We have to use the axiom of choice since I is uncountable. This follows from the observation that the uncountable set $\mathbb{J} = \dot{\bigcup}_{i\in I}[x_i]$ is the disjoint union of countable sets $[x_i] = (x + \mathbb{Q}) \cap \mathbb{J}$. It is known that all proofs for Theorem G.4 must use the axiom of choice or some equivalent statement, see Solovay [49].

Since $\bar{\lambda}^1$ is invariant under translations, we then have $\bar{\lambda}^1(q+L)=\bar{\lambda}^1(L)$ for all $q\in\mathbb{Q}\cap(-1,1)$. We conclude that

$$1\leqslant \sum_{q\in\mathbb{Q}\cap(-1,1)} \bar{\lambda}^1(L)\leqslant 3,$$

which is not possible. This proves that L cannot be Lebesgue measurable. The same calculation works for the outer measure – see Theorem 6.1 with $\mathcal{S}=\mathcal{J}$ in (6.1) – and reveals that $(\lambda^1)^*(L)>0$. $\square$

Corollary G.5 *Assuming the axiom of choice, every Borel set $B\in\mathscr{B}(\mathbb{R}^n)$ with $\lambda^n(B)>0$ contains a non-Lebesgue measurable set with strictly positive outer Lebesgue measure.*

Proof First, we assume that $n=1$ and $B\in\mathscr{B}(\mathbb{R})$, $\lambda^1(B)>0$. Without loss of generality, we may assume that $B\subset(-r,r)$ for some $r>0$. If we repeat literally the proof of Theorem G.4 with $\mathbb{J}\rightsquigarrow B$, $(-1,1)\rightsquigarrow(-r,r)$ and $[-1,2)\rightsquigarrow(-3r,3r)$, we find some $L\subset B$ with $(\lambda^1)^*(L)>0$.

Now it is clear how the n-dimensional version works: in the above argument replace $\mathbb{Q}\rightsquigarrow\mathbb{Q}^n$, $(-r,r)\rightsquigarrow B_r(0)$ and $(-3r,3r)\rightsquigarrow B_{3r}(0)$. $\square$

The question of whether there are Lebesgue measurable sets which are not Borel measurable can be answered constructively. Since this is quite tedious, we content ourselves with the fact that there are ‘fewer’ Borel than Lebesgue measurable sets.

Theorem G.6 *We have $\#\mathscr{B}(\mathbb{R}^n)=\mathfrak{c}$.*

Corollary G.7 *There are Lebesgue measurable sets which are not Borel sets.*

Proof of G.7 We know from Theorem G.2 that $\#\overline{\mathscr{B}}(\mathbb{R}^n)=\#\mathscr{P}(\mathbb{R}^n)$ and we know from Theorem G.6 that $\#\mathscr{B}(\mathbb{R}^n)=\mathfrak{c}$. Since, by Theorem 2.9 and Problem 2.17, $\#\mathscr{P}(\mathbb{R}^n)>\#\mathbb{R}^n=\mathfrak{c}$, we conclude that $\mathscr{B}(\mathbb{R}^n)\subsetneq\overline{\mathscr{B}}(\mathbb{R}^n)$. $\square$

To prove Theorem G.6 we show that the Borel sets are contained in a family of sets which has cardinality $\mathfrak{c}$. Let $\mathbb{F}:=\bigcup_{k=1}^{\infty}\mathbb{N}^k$ be the set of all *finite* sequences of natural numbers and write $\mathscr{C}$ for the family of open balls $B_r(x)\subset\mathbb{R}^n$ with radius $r\in\mathbb{Q}^+$ and centre $x\in\mathbb{Q}^n$. We saw in Problems 2.19 and 2.9 that

$$\#\mathbb{F}=\#\mathbb{N}\quad\text{and}\quad\#\mathscr{C}=\#\left(\mathbb{Q}^+\times\mathbb{Q}^n\right)=\#\mathbb{N}.$$

Therefore, the collection of all *Souslin schemes*

$$\mathfrak{s}:\mathbb{F}\to\mathscr{C},\ (i_1,i_2,\ldots,i_k)\mapsto C_{i_1i_2\cdots i_k}$$

has cardinality $\#\mathscr{C}^{\mathbb{F}} = \#\mathbb{N}^{\mathbb{N}} = \mathfrak{c}$, see Problem 2.18. With each Souslin scheme $\mathfrak{s}$ we can associate a set $A \subset \mathbb{R}^n$ in the following way: take any sequence $(i_j)_{j\in\mathbb{N}}$ of natural numbers and consider the sequence of finite tuples (i_1), (i_1, i_2), $(i_1, i_2, i_3), \dots$, $(i_1, i_2, \dots, i_k), \dots$ formed by the first $1, 2, 3, \dots$ members of the sequence $(i_j)_{j\in\mathbb{N}}$. Using the Souslin scheme $\mathfrak{s}$ we pick for each tuple $(i_1, i_2, \dots, i_k)$ the corresponding set $C_{i_1 i_2 \dots i_k} \in \mathscr{C}$ to get a sequence of sets $C_{i_1}, C_{i_1 i_2}, C_{i_1 i_2 i_3}, \dots, C_{i_1 i_2 \dots i_k}, \dots$ from $\mathscr{C}$. Finally, we form the intersection $C_{i_1} \cap C_{i_1 i_2} \cap C_{i_1 i_2 i_3} \cap \dots \cap C_{i_1 i_2 \cdots i_k} \cap \cdots$ of all these sets and consider the union over all possible sequences $(i_j)_{j\in\mathbb{N}}$ of natural numbers:

$$A := A(\mathfrak{s}) := \bigcup_{(i_j : j\in\mathbb{N})\in\mathbb{N}^{\mathbb{N}}} \bigcap_{k=1}^{\infty} C_{i_1 i_2 \dots i_k}.$$

Note that this union is uncountable, so that A is not necessarily a Borel set.

It is often helpful to visualize this construction as a tree, see Fig. G.1, where the $C_{i_1}, C_{i_1 i_2}, C_{i_1 i_2 i_3}, \dots \in \mathscr{C}$ are the sets of the first, second, third, etc. generation. We will call $C_{i_1 i_2}$ or $C_{i_1 i_2 i_3}$ children or grandchildren of C_{i_1}.

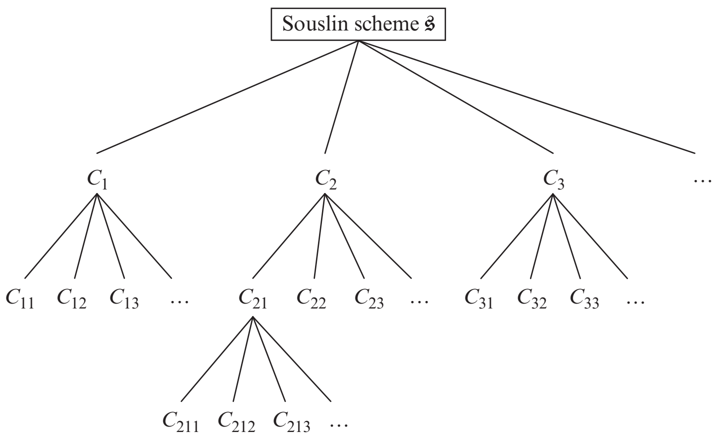


Fig. G.1. The Souslin scheme in the form of a family tree.

Definition G.8 (Souslin) Let $\mathbb{F} = \bigcup_{k=1}^{\infty} \mathbb{N}^k$, $\mathscr{C}$ be the family of all open balls in $\mathbb{R}^n$ with rational centres and radii, with $\mathfrak{s}$ a Souslin scheme and $A(\mathfrak{s})$ as above. The sets in $\alpha(\mathscr{C}) := \{A(\mathfrak{s}) : \mathfrak{s} \in \mathscr{C}^{\mathbb{F}}\}$ are called *analytic* or *Souslin sets* (generated by $\mathscr{C}$).

Lemma G.9 *Let $\mathbb{F}, \mathscr{C}$ and $\mathfrak{s}$ be as in Definition G.8.*

(i) $\alpha(\mathscr{C})$ *is stable under countable unions and countable intersections.*
(ii) $\alpha(\mathscr{C})$ *contains all open and all closed subsets of* $\mathbb{R}^n$.

(iii) $\mathscr{B}(\mathbb{R}^n) = \sigma(\mathscr{C}) \subset \alpha(\mathscr{C})$.
(iv) $\#\alpha(\mathscr{C}) \leqslant \mathfrak{c}$.

Proof (i) Let $A_\ell \in \alpha(\mathscr{C})$, $\ell \in \mathbb{N}$, be a sequence of analytic sets

$$A_\ell = \bigcup_{(i_j\,:\,j\in\mathbb{N})\in\mathbb{N}^\mathbb{N}} \bigcap_{k=1}^{\infty} C^\ell_{i_1 i_2 \cdots i_k}.$$

Since

$$A := \bigcup_{\ell\in\mathbb{N}} A_\ell = \bigcup_{\ell\in\mathbb{N}} \bigcup_{(i_j\,:\,j\in\mathbb{N})\in\mathbb{N}^\mathbb{N}} \bigcap_{k=1}^{\infty} C^\ell_{i_1 i_2 \cdots i_k}$$

it is obvious that A can be obtained from a Souslin scheme $\mathfrak{s}$ which arises from the juxtaposition of the Souslin schemes belonging to the A_ℓ: arrange the double sequence $C^\ell_{i_1}$, $(i_1,\ell)\in\mathbb{N}\times\mathbb{N}$, in one sequence – e.g. using the counting scheme of Example 2.5(iv) – to get the first generation of sets, while all other generations follow suit in genealogical order. Thus $A \in \alpha(\mathscr{C})$.

For the countable intersection of the A_ℓ we observe first that

$$B := \bigcap_{\ell\in\mathbb{N}} A_\ell = \bigcap_{\ell\in\mathbb{N}} \bigcup_{(i_j\,:\,j\in\mathbb{N})\in\mathbb{N}^\mathbb{N}} \bigcap_{k=1}^{\infty} C^\ell_{i_1 i_2 \ldots i_k} \overset{[\text{✍}]}{=} \bigcup_{\substack{(i^m_j\,:\,j\in\mathbb{N})\in\mathbb{N}^\mathbb{N}\\ m=1,2,3,\ldots}} \bigcap_{\ell=1}^{\infty} \bigcap_{k=1}^{\infty} C^\ell_{i^\ell_1 i^\ell_2 \ldots i^\ell_k}$$

and then we merge the two infinite intersections indexed by $(\ell,k)\in\mathbb{N}\times\mathbb{N}$ into a single infinite intersection. Once again this can be achieved through the counting scheme of Example 2.5(iv):

$$\begin{array}{ccccccccccccc}
C^1_{i^1_1} & \cap & C^1_{i^1_1 i^1_2} & \cap & C^2_{i^2_1} & \cap & C^1_{i^1_1 i^1_2 i^1_3} & \cap & C^2_{i^2_1 i^2_2} & \cap & C^3_{i^3_1} & \cap & \ldots \\
\updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \\
(1,1) & \to & (1,2) & \to & (2,1) & \to & (1,3) & \to & (2,2) & \to & (3,1) & \to & \ldots
\end{array}$$

and so

$$B = \bigcup_{\substack{(i^m_j\,:\,j\in\mathbb{N})\in\mathbb{N}^\mathbb{N}\\ m=1,2,3,\ldots}} \left(C^1_{i^1_1} \cap C^1_{i^1_1 i^1_2} \cap C^2_{i^2_1} \cap C^1_{i^1_1 i^1_2 i^1_3} \cap C^2_{i^2_1 i^2_2} \cap C^3_{i^3_1} \cap \ldots \right).$$

We will now construct a Souslin scheme which produces B by arranging the sets $C^j_{k\ell m\ldots}$ in a tree as follows.

- The first generation are the sets $C^1_{i^1_1}$, $i^1_1 \in \mathbb{N}$.
- The second generation are the sets $C^1_{i^1_1 i^1_2}$, $i^1_2 \in \mathbb{N}$, such that they are, for fixed i^1_1, the children of $C^1_{i^1_1}$.

- Each $C^1_{i^1_1 i^1_2}$ has the same offspring, namely the sets $C^2_{i^2_1}$, $i^2_1 \in \mathbb{N}$, which jointly form the third generation.
- The fourth generation are the sets $C^1_{i^1_1 i^1_2 i^1_3}$, $i^1_3 \in \mathbb{N}$, such that they are, for fixed i^1_1, i^1_2, the grandchildren of $C^1_{i^1_1 i^1_2}$.
- The fifth generation are the sets $C^2_{i^2_1 i^2_2}$, $i^2_2 \in \mathbb{N}$, such that they are, for fixed i^2_1, the grandchildren of $C^2_{i^2_1}$.
- Each $C^2_{i^2_1 i^2_2}$ has the same offspring, namely the sets $C^3_{i^3_1}$, $i^3_1 \in \mathbb{N}$, which jointly form the sixth generation.
- …

This shows that $B \in \alpha(\mathscr{C})$.

(ii) Every open set can be written as a countable union of $\mathscr{C}$-sets

$$U = \bigcup_{\substack{B_r(x) \subset U \\ B_r(x) \in \mathscr{C}}} B_r(x).$$

Indeed, the inclusion '$\supset$' is obvious, for '$\subset$' fixes $x \in U$. Then there exists some $r \in \mathbb{Q}^+$ with $B_r(x) \subset U$. Since $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$, $x \in B_{r/2}(y)$ for some $y \in \mathbb{Q}^n$ with $|x-y| < r/4$, so that $x \in B_{r/2}(y) \subset U$. Since there are only countably many sets in $\mathscr{C}$, the union is *a fortiori* countable. By part (i) we then get that $U \in \alpha(\mathscr{C})$, i.e. $\alpha(\mathscr{C})$ contains all open sets.

For a closed set F we know that

$$F = \bigcap_{k \in \mathbb{N}} U_k, \quad \text{where} \;\; U_k = F + B_{1/k}(0) = \left\{ y : x \in F,\ |x-y| < \frac{1}{k} \right\}$$

is a countable intersection of open [✍] sets U_k. Since open sets are analytic, part (i) implies that $F \in \alpha(\mathscr{C})$.

(iii) Consider the system $\Sigma := \{A \in \alpha(\mathscr{C}) : A^c \in \alpha(\mathscr{C})\}$. We claim that Σ is a σ-algebra. Obviously, Σ satisfies conditions (Σ_1), (Σ_2) – i.e. it contains $\mathbb{R}^n$ and is stable under complementation. To see (Σ_3), we take a sequence $(A_k)_{k \in \mathbb{N}} \subset \Sigma$ and observe that, by part (i),

$$\bigcup_{k \in \mathbb{N}} A_k \in \alpha(\mathscr{C}) \quad \text{and} \quad \left(\bigcup_{k \in \mathbb{N}} A_k \right)^c = \bigcap_{k \in \mathbb{N}} \underbrace{A_k^c}_{\in \alpha(\mathscr{C})} \in \alpha(\mathscr{C}),$$

so that $\bigcup_k A_k \in \Sigma$.

Because of (ii) we have $\mathscr{C} \subset \Sigma \subset \alpha(\mathscr{C})$ and this implies $\sigma(\mathscr{C}) \subset \alpha(\mathscr{C})$. Since, by (ii), all open sets are countable unions of sets from $\mathscr{C}$, we get $\mathscr{C} \subset \mathscr{O} \subset \sigma(\mathscr{C})$ ($\mathscr{O}$ denotes the family of open sets) or $\sigma(\mathscr{C}) = \sigma(\mathscr{O}) \stackrel{\text{def}}{=} \mathscr{B}(\mathbb{R}^n)$.

(iv) follows immediately from the fact that there are $\#\mathscr{C}^{\mathbb{F}} = \#\mathbb{N}^{\mathbb{N}} = \mathfrak{c}$ Souslin schemes, see Definition G.8. □

The *Proof of Theorem G.6* is now easy By Lemma G.9 there are at most $\mathfrak{c}$ analytic sets. Since each singleton $\{x\}$, $x \in \mathbb{R}^n$, is a Borel set, there are at least $\mathfrak{c}$ Borel sets (use Problem 2.17 to see that $\#\mathbb{R}^n = \mathfrak{c}$). So,

$$\mathfrak{c} \leqslant \#\mathscr{B}(\mathbb{R}^n) \leqslant \#\alpha(\mathscr{C}) \leqslant \mathfrak{c}$$

and an application of Theorem 2.7 finishes the proof. □

Remark G.10 Our approach to analytic sets follows the original construction of Souslin [47], which makes it easy to determine the cardinality of $\alpha(\mathscr{C})$. This, however, comes at a price: if one wants to work with this definition, things become messy, as we have seen in the proof of Lemma G.9(i). Nowadays analytic sets are often introduced by one of the following equivalent properties. *A set $A \subset \mathbb{R}^n$ is analytic if, and only if, one of the following equivalent conditions holds:*

(i) $A = f(\mathbb{R})$ *for some left-continuous function* $f : \mathbb{R} \to \mathbb{R}^n$*;*
(ii) $A = g(\mathbb{N}^{\mathbb{N}})$ *for some Borel measurable function* $g : \mathbb{N}^{\mathbb{N}} \to \mathbb{R}^n$*;*
(iii) $A = h(B)$ *for some Borel set* $B \in \mathscr{B}(X)$*, some Polish space*[2] X *and some Borel measurable function* $h : B \to \mathbb{R}^n$*;*
(iv) $A = \pi_2(B)$ *where* $\pi_2 : Y \times \mathbb{R}^n \to \mathbb{R}^n$ *is the coordinate projection onto* $\mathbb{R}^n$*,* Y *is a compact Hausdorff space*[3] *and* $B \subset Y \times \mathbb{R}^n$ *is a* $\mathscr{K}_{\sigma\delta}$*-set, i.e.* B *can be written as a countable intersection ('*δ*') of countable unions ('*σ*') of compact subsets ('*$\mathscr{K}$*') of* $Y \times \mathbb{R}^n$*.*

For a proof we refer to Srivastava [48] which is also our main reference for analytic sets. The *Souslin operation* α can be applied to other systems of sets than $\mathscr{C}$. Without proof we mention the following facts:

$$\alpha(\mathscr{C}) = \alpha(\{\text{open sets}\}) = \alpha(\{\text{closed sets}\}) = \alpha(\{\text{compact sets}\})$$

and also

$$\alpha(\mathscr{C}) = \alpha(\alpha(\mathscr{C})) \quad \text{and} \quad \mathscr{B}(\mathbb{R}^n) \subsetneq \alpha(\mathscr{C}) \subsetneq \overline{\mathscr{B}}(\mathbb{R}^n).$$

Most constructions of sets which are not Borel but still Lebesgue measurable are actually constructions of non-Borel analytic sets, see Dudley [14, Section 13.2].

[2] A space which can be endowed with a metric for which it is complete and has a countable dense subset.
[3] This is a topological space where distinct points have disjoint neighbourhoods.

Appendix H
Regularity of Measures

Let (X,d) be a metric space and denote by $\mathscr{O}$ the open (with respect to the metric d), by $\mathscr{C}$ the closed and by $\mathscr{B}(X)=\sigma(\mathscr{O})$ the Borel sets of X.

Definition H.1 A measure μ on $(X,d,\mathscr{B}(X))$ is called *outer regular*, if

$$\mu(B)=\inf\{\mu(U):U\supset B,\ U\text{ open}\}\quad\forall B\in\mathscr{B}(X)\tag{H.1}$$

and *inner regular*, if $\mu(K)<\infty$ for all compact sets $K\subset X$ and

$$\mu(U)=\sup\{\mu(K):K\subset U,\ K\text{ compact}\}\quad\forall U\in\mathscr{O}.\tag{H.2}$$

A measure which is both outer and inner regular is called *regular*. We write $\mathfrak{M}^+_{\mathrm{r}}(X)$ for the family of regular measures on $(X,\mathscr{B}(X))$.

Regularity is essentially a topological property. The space X is called σ-*compact* if there exists a sequence of compact sets $K_n\uparrow X$. A typical example of a σ-compact space is a locally compact, separable metric space.

Theorem H.2 *Let (X,d) be a metric space. Every finite measure μ on $(X,\mathscr{B}(X))$ is outer regular. If X is σ-compact, then μ is also inner regular, hence regular.*

Proof Define

$$\Sigma:=\{A\subset X\mid\forall\epsilon>0\,\exists F\in\mathscr{C},\ U\in\mathscr{O},\ F\subset A\subset U:\ \mu(U\setminus F)<\epsilon\}.$$

Step 1. Σ is a σ-algebra.

Σ_1 $\emptyset\in\Sigma$ is obvious.

Σ_2 Let $A\in\Sigma$ and fix $\epsilon>0$. By definition, there is an open and closed set $F_\epsilon\subset A\subset U_\epsilon$ such that $\mu(U_\epsilon\setminus F_\epsilon)<\epsilon$. Since

$$U^c_\epsilon\in\mathscr{C},\quad F^c_\epsilon\in\mathscr{O},\quad U^c_\epsilon\subset A^c\subset F^c_\epsilon\quad\text{and}\quad F^c_\epsilon\setminus U^c_\epsilon=U_\epsilon\setminus F_\epsilon$$

we see $\mu(F^c_\epsilon\setminus U^c_\epsilon)=\mu(U_\epsilon\setminus F_\epsilon)<\epsilon$, hence $A^c\in\Sigma$.

Σ_3 Let $A_n \in \Sigma$, $n \in \mathbb{N}$, and $\epsilon > 0$. For every n there are

$$F_n \in \mathscr{C},\ U_n \in \mathscr{O},\ F_n \subset A_n \subset U_n \ \text{ with } \ \mu(U_n \setminus F_n) < \epsilon/2^n.$$

Thus,

$$\overbrace{\Phi_n := F_1 \cup \cdots \cup F_n}^{\text{closed}} \subset A_1 \cup \cdots \cup A_n \subset \overbrace{U := \bigcup_{i\in\mathbb{N}} U_i}^{\text{open}}.$$

Because of the inclusion

$$\bigcap_{n\in\mathbb{N}} U \setminus \Phi_n = \bigcup_{i\in\mathbb{N}} U_i \setminus \bigcup_{k\in\mathbb{N}} F_k = \bigcup_{i\in\mathbb{N}} \Big(U_i \setminus \bigcup_{k\in\mathbb{N}} F_k \Big) \subset \bigcup_{i\in\mathbb{N}} (U_i \setminus F_i)$$

we can use the continuity and σ-subadditivity of μ to get

$$\lim_{n\to\infty} \mu(U \setminus \Phi_n) = \mu\Big(\bigcup_{i\in\mathbb{N}} U_i \setminus \bigcup_{k\in\mathbb{N}} F_k \Big) \leqslant \sum_{i\in\mathbb{N}} \mu(U_i \setminus F_i) \leqslant \sum_{i\in\mathbb{N}} \frac{\epsilon}{2^i} = \epsilon.$$

Consequently, $\mu(U \setminus \Phi_n) < 2\epsilon$ for all $n > n(\epsilon)$ and we have $\bigcup_n A_n \in \Sigma$.

Step 2. $\mathscr{C} \subset \Sigma$ and $\mathscr{B}(X) \subset \Sigma$. For a closed set $F \subset X$ we have

$$U_n := \bigcup_{x\in F} B_{1/n}(x) \in \mathscr{O} \quad \text{and} \quad \bigcap_n U_n = F.$$

Because of the continuity of measures we get $\lim_{n\to\infty} \mu(U_n) = \mu(F)$, i.e. $F \subset U_n$ and $\mu(U_n \setminus F) < \epsilon$ for large $n > N(\epsilon)$. This proves $F \in \Sigma$ as well as

$$\mathscr{B}(X) = \sigma(\mathscr{C}) \subset \sigma(\Sigma) = \Sigma.$$

Step 3. μ is outer regular. Let $B \in \mathscr{B}(X)$. Step 2 ensures the existence of $(U_n)_n \subset \mathscr{O}$ and $(F_n)_n \subset \mathscr{C}$ such that $F_n \subset B \subset U_n$ and $\lim_{n\to\infty} \mu(U_n \setminus F_n) = 0$. Therefore,

$$\mu(B \setminus F_n) + \mu(U_n \setminus B) \leqslant 2\,\mu(U_n \setminus F_n) \xrightarrow[n\to\infty]{} 0,$$

which shows that the infimum in (H.1) is attained.

Step 4. If X is σ-compact, then μ is inner regular. Let $B \in \mathscr{B}(X)$, pick a sequence $L_m \uparrow X$ of compact sets and F_n as in Step 3. Define

$$K_{n,m} := F_n \cap L_m.$$

Because of the continuity of measures, $\lim_{m\to\infty} \mu(F_n \setminus K_{n,m}) = 0$, and we see

$$\mu(B \setminus K_{n,m}) \leqslant \mu(B \setminus F_n) + \mu(F_n \setminus K_{n,m}) \xrightarrow[m\to\infty]{} \mu(B \setminus F_n) \xrightarrow[n\to\infty]{} 0.$$

This proves (H.2) – even for arbitrary sets $U = B \in \mathscr{B}(X)$. $\square$

We can extend Theorem H.2 to σ-finite measures.

Theorem H.3 *Let (X, d) be a metric space and μ a measure on $(X, \mathscr{B}(X))$ such that $\mu(K) < \infty$ for all compact sets $K \subset X$.*

(i) *If X is σ-compact, then μ is inner regular.*
(ii) *If there exists a sequence $G_n \in \mathscr{O}$, $G_n \uparrow X$ such that $\mu(G_n) < \infty$, then μ is outer regular.*

Proof (i) By assumption, there exists a sequence of compact sets $K_n \uparrow X$. We define a family of finite measures $\mu_n(B) := \mu(B \cap K_n)$, $B \in \mathscr{B}(X)$, $n \in \mathbb{N}$. Using the continuity of measures and Theorem H.2 we get for every $B \in \mathscr{B}(X)$

$$\mu(B) \overset{4.3(\text{vi})}{=} \sup_n \mu_n(B) \overset{\text{H.2}}{=} \sup_n \sup_{K \subset B, \text{cpt}} \mu_n(K) = \sup_{K \subset B, \text{cpt.}} \sup_n \mu_n(K) = \sup_{K \subset B, \text{cpt.}} \mu(K).$$

(ii) Pick some $B \in \mathscr{B}(X)$ and $\epsilon > 0$. From the proof of Theorem H.2 (definition of the family Σ for the measures $\mu(\cdot \cap G_n)$) we know that

$$\forall n \in \mathbb{N} \quad \exists U_n \in \mathscr{O},\ B \subset U_n : \mu((U_n \setminus B) \cap G_n) < \epsilon 2^{-n}. \tag{H.3}$$

If $n = 1$, we can use this as the start of the induction for the following assertion:

$$\mu\left(\bigcup_{i=1}^{n} U_i \cap G_i\right) \leqslant \mu(B \cap G_n) + \sum_{i=1}^{n} \epsilon 2^{-i}. \tag{H.4}$$

In order to see that (H.4) remains valid for $n \rightsquigarrow n+1$, we use the strong additivity of measures

$$\begin{aligned}
&\mu\left(\bigcup_{i=1}^{n+1} U_i \cap G_i\right) \\
&= \mu\left((U_{n+1} \cap G_{n+1}) \cup \bigcup_{i=1}^{n} U_i \cap G_i\right) \\
&\overset{4.3(\text{iv})}{=} \mu(U_{n+1} \cap G_{n+1}) + \mu\left(\bigcup_{i=1}^{n} U_i \cap G_i\right) - \mu\Bigg(\underbrace{(U_{n+1} \cap G_{n+1})}_{\supset B \cap G_{n+1} \supset B \cap G_n} \cap \underbrace{\bigcup_{i=1}^{n} U_i \cap G_i}_{\supset B \cap G_n}\Bigg) \\
&\underset{(\text{H.4})}{\overset{(\text{H.3})}{\leqslant}} \mu(B \cap G_{n+1}) + \epsilon 2^{-n-1} + \mu(B \cap G_n) + \sum_{i=1}^{n} \epsilon 2^{-i} - \mu(B \cap G_n) \\
&= \mu(B \cap G_{n+1}) + \sum_{i=1}^{n+1} \epsilon 2^{-i}.
\end{aligned}$$

Since $B = \bigcup_{i=1}^{\infty}(B \cap G_i) \subset \bigcup_{i=1}^{\infty}(U_i \cap G_i) \in \mathscr{O}$, we find

$$\mu(B) \leqslant \mu\left(\bigcup_{i=1}^{\infty} U_i \cap G_i\right) \overset{4.3(\text{vi})}{=} \sup_{n\in\mathbb{N}} \mu\left(\bigcup_{i=1}^{n} U_i \cap G_i\right)$$

$$\underset{4.3(\text{vi})}{\overset{(\text{H.4})}{\leqslant}} \sup_{n\in\mathbb{N}}\left(\mu(B \cap G_n) + \sum_{i=1}^{n} \epsilon 2^{-i}\right) \leqslant \mu(B) + \epsilon.$$

Finally, letting $\epsilon \downarrow 0$, we get outer regularity. $\square$

Appendix I

A Summary of the Riemann Integral

In this appendix we give an outline of the Riemann integral on the real line. The notion of integration was well known for a long time and, ever since the creation of differential calculus by Newton and Leibniz, integration was perceived as anti-derivative. Several attempts to make this precise were made, but the problem with these approaches was partly that the notion of an integral was implicit – i.e. axiomatic rather than given constructively – partly that the choice of possible integrands was rather limited and partly that some fundamental points were unclear.

Out of the need to overcome these insufficiencies and to have a sound foundation, Bernhard Riemann asked in his Habilitationsschrift *Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe*[1] the question *Also zuerst: Was hat man unter $\int_a^b f(x)dx$ zu verstehen?*[2] [39, p. 239] and proposed a general way to define an integral which is constructive, which is (at least for continuous integrands) the anti-derivative, and which can deal with a wider range of integrands than all its predecessors.

We do not follow Riemann's original approach but use the Darboux technique of upper and lower integrals. Riemann's original definition will be recovered in Theorem I.5(iv).

The (Proper) Riemann Integral

Riemann integrals are defined only for *bounded* functions on *compact* intervals $[a,b] \subset \mathbb{R}$; this avoids all sorts of complications arising when either the domain or the range of the integrand is infinite. Both cases can be dealt with

[1] On the representability of a function by a trigonometric series.

[2] First of all: what is the meaning of $\int_a^b f(x)dx$?

by various extensions of the Riemann integral, one of which – the so-called *improper* Riemann integral – we will discuss later on.

A *partition* π of the interval $[a,b]$ consists of finitely many points satisfying

$$\pi = \{a = t_0 < t_1 < \cdots < t_{k-1} < t_k = b\}, \qquad k = k(\pi).$$

We call $\operatorname{mesh}(\pi) := \max_{1\leqslant i\leqslant k(\pi)}(t_i - t_{i-1})$ the *mesh* or *fineness* of the partition.

Given a partition π and a bounded function $u : [a,b] \to \mathbb{R}$, we define

$$m_i := \inf_{x\in[t_{i-1},t_i]} u(x) \quad \text{and} \quad M_i := \sup_{x\in[t_{i-1},t_i]} u(x),$$

for all $i = 1, 2, \ldots, k(\pi)$, and introduce the *lower*, resp. *upper, Darboux sums*

$$S_\pi[u] := \sum_{i=1}^{k(\pi)} m_i(t_i - t_{i-1}) \quad \text{resp.} \quad S^\pi[u] := \sum_{i=1}^{k(\pi)} M_i(t_i - t_{i-1}).$$

Obviously, $S_\pi[\,\cdot\,], S^\pi[\,\cdot\,]$ are linear, and, if $|u(x)| \leqslant M$, they satisfy

$$|S_\pi[u]| \leqslant S_\pi[|u|] \leqslant M(b-a), \qquad |S^\pi[u]| \leqslant S^\pi[|u|] \leqslant M(b-a). \tag{I.1}$$

Lemma I.1 *Let π be a partition of $[a,b]$ and $\pi' \supset \pi$ be a refinement of π. Then*

$$S_\pi[u] \leqslant S_{\pi'}[u] \leqslant S^{\pi'}[u] \leqslant S^\pi[u]$$

holds for all bounded functions $u : [a,b] \to \mathbb{R}$.

Proof Since $S_\pi[u] = -S^\pi[-u]$ and since $S_{\pi'}[u] \leqslant S^{\pi'}[u]$ is trivially fulfilled, it suffices to show that $S_\pi[u] \leqslant S_{\pi'}[u]$. The partitions π, π' contain only finitely many points and we may assume that $\pi' = \pi \cup \{\tau\}$, where $t_{i_0-1} < \tau < t_{i_0}$ for some index $1 \leqslant i_0 \leqslant k(\pi)$. The rest follows by iteration. Clearly,

$$\begin{aligned}
S_\pi[u] &= \sum_{i\neq i_0} m_i(t_i - t_{i-1}) + m_{i_0}(t_{i_0} - \tau) + m_{i_0}(\tau - t_{i_0-1}) \\
&\leqslant \sum_{i\neq i_0} m_i(t_i - t_{i-1}) + \inf_{x\in[\tau,t_{i_0}]} u(x)(t_{i_0} - \tau) + \inf_{x\in[t_{i_0-1},\tau]} u(x)(\tau - t_{i_0-1}) \\
&= S_{\pi'}[u].
\end{aligned}$$

□

Lemma I.1 shows that the following definition makes sense.

Definition I.2 Let $u : [a,b] \to \mathbb{R}$ be a bounded function. The *lower* and *upper integrals* of u are given by

$$\underline{\int_a^b} u := \sup_\pi S_\pi[u] \quad \text{and} \quad \overline{\int_a^b} u := \inf_\pi S_\pi[u],$$

where $\sup_\pi$ and $\inf_\pi$ range over all finite partitions of $[a,b]$.

Lemma I.3 $\underline{\int_a^b} u \leqslant \overline{\int_a^b} u$ *and* $\underline{\int_a^b} u = -\overline{\int_a^b}(-u)$.

Definition I.4 A bounded function $u : [a, b] \to \mathbb{R}$ is said to be (*Riemann*) *integrable*, if the upper and lower integrals coincide. Their common value

$$\int_a^b u(x)dx := \underline{\int_a^b} u = \overline{\int_a^b} u$$

is called the (*Riemann*) *integral* of u. The collection of all Riemann integrable functions in $[a, b]$ is denoted by $\mathcal{R}[a, b]$.

Theorem I.5 (characterization of $\mathcal{R}[a, b]$) *Let* $u : [a, b] \to \mathbb{R}$ *be a bounded function. Then the following assertions are equivalent.*

(i) $u \in \mathcal{R}[a, b]$.
(ii) *For every* $\epsilon > 0$ *there is some partition* π *such that* $S^\pi[u] - S_\pi[u] \leqslant \epsilon$.
(iii) *For every* $\epsilon > 0$ *there is some* $\delta > 0$ *such that* $S^\pi[u] - S_\pi[u] \leqslant \epsilon$ *for all partitions* π *with* $\operatorname{mesh}\pi < \delta$.
(iv) *The limit*

$$I = \lim_{\operatorname{mesh}\pi \to 0} \sum_{i : t_i \in \pi} u(\xi_i)(t_i - t_{i-1})$$

exists for every choice of intermediate values $t_{i-1} \leqslant \xi_i \leqslant t_i$*; this means that for all* $\epsilon > 0$ *there exists a* $\delta > 0$ *such that for all partitions* π *with* $\operatorname{mesh}\pi < \delta$

$$\left| I - \sum_{i : t_i \in \pi} u(\xi_i)(t_i - t_{i-1}) \right| \leqslant \epsilon$$

independently of the intermediate points.

If the limit exists, $I = \overline{\int_a^b} u = \underline{\int_a^b} u$.

Proof We show the implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (i).

(i) $\Rightarrow$ (ii). By the very definition and the lower and upper integrals in terms of sup and inf, we find for every $\epsilon > 0$ partitions π' and π'' such that

$$\underline{\int_a^b} u - S_{\pi'}[u] \leqslant \frac{\epsilon}{2} \quad \text{and} \quad S^{\pi''}[u] - \overline{\int_a^b} u \leqslant \frac{\epsilon}{2}.$$

Using the common refinement $\pi = \pi' \cup \pi''$, we get from Lemma I.1 and the integrability of u

$$S^\pi[u] - S_\pi[u] \leqslant S^{\pi''}[u] - S_{\pi'}[u] = \left(S^{\pi''}[u] - \overline{\int_a^b} u \right) + \left(\underline{\int_a^b} u - S_{\pi'}[u] \right) \leqslant \epsilon.$$

(ii) $\Rightarrow$ (iii). This is the most intricate step in the proof. Fix $\epsilon > 0$ and denote by $\pi_\epsilon := \{a = t_0^\epsilon < t_1^\epsilon < \cdots < t_k^\epsilon = b\}$ the partition in (ii). We choose $\delta > 0$ in such a way that

$$\delta < \frac{1}{2} \min_{1 \leqslant i \leqslant k} \left(t_i^\epsilon - t_{i-1}^\epsilon\right) \quad \text{and} \quad \delta < \frac{\epsilon}{4k\|u\|_\infty}.$$

If $\pi := \{a = t_0 < t_1 < \cdots < t_N = b\}$ is any partition with mesh $\pi < \delta$ we find

$$\begin{aligned} S^\pi[u] - S_\pi[u] \quad = \quad & \sum_{i\,:\,\pi_\epsilon \cap [t_{i-1}, t_i] \neq \emptyset} (M_i^\pi - m_i^\pi)(t_i - t_{i-1}) \\ & + \sum_{i\,:\,\pi_\epsilon \cap [t_{i-1}, t_i] = \emptyset} (M_i^\pi - m_i^\pi)(t_i - t_{i-1}), \end{aligned} \tag{I.2}$$

where M_i^π, m_i^π indicates that the supremum, resp. infimum, is taken w.r.t. intervals defined by the partition π. The first sum has at most $2k$ terms as $\pi_\epsilon \cap [a, t_1] = \{a\}$, $\pi_\epsilon \cap [t_{N-1}, b] = \{b\}$ and since all other t_i^ϵ, $1 \leqslant i \leqslant k-1$, appear in exactly one or two intervals defined by π. Thus

$$\sum_{i\,:\,\pi_\epsilon \cap [t_{i-1}, t_i] \neq \emptyset} (M_i^\pi - m_i^\pi)(t_i - t_{i-1}) \leqslant 2k \cdot 2\|u\|_\infty \cdot \delta \leqslant \epsilon. \tag{I.3}$$

The second sum in (I.2) can be written as a double sum

$$\begin{aligned} & \sum_{i\,:\,\pi_\epsilon \cap [t_{i-1}, t_i] = \emptyset} (M_i^\pi - m_i^\pi)(t_i - t_{i-1}) \\ & \qquad = \sum_{i=1}^k \left[\sum_{\ell\,:\,[t_{\ell-1}, t_\ell] \subset [t_{i-1}^\epsilon, t_i^\epsilon]} (M_\ell^\pi - m_\ell^\pi)(t_\ell - t_{\ell-1}) \right] \\ & \qquad \leqslant \sum_{i=1}^k \left[\sum_{\ell\,:\,[t_{\ell-1}, t_\ell] \subset [t_{i-1}^\epsilon, t_i^\epsilon]} (M_i^{\pi_\epsilon} - m_i^{\pi_\epsilon})(t_\ell - t_{\ell-1}) \right] \\ & \qquad \leqslant \sum_{i=1}^k (M_i^{\pi_\epsilon} - m_i^{\pi_\epsilon})(t_i^\epsilon - t_{i-1}^\epsilon) \\ & \qquad = S^{\pi_\epsilon}[u] - S_{\pi_\epsilon}[u] \leqslant \epsilon. \end{aligned} \tag{I.4}$$

Together (I.2)–(I.4) show that $S^\pi[u] - S_\pi[u] \leqslant 2\epsilon$ for any partition π with mesh $\pi < \delta$.

(iii) $\Rightarrow$ (iv). Fix $\epsilon > 0$ and choose $\delta > 0$ as in (iii). Then we have, for any partition $\pi = \{a = t_0 < \cdots < t_{k(\pi)} = b\}$ with mesh $\pi < \delta$ and any choice of intermediate points $\xi_i \in [t_{i-1}, t_i]$,

$$S^\pi[u] - \epsilon \leqslant S_\pi[u] \leqslant \sum_{i=1}^{k(\pi)} u(\xi_i)(t_i - t_{i-1}) \leqslant S^\pi[u] \leqslant S_\pi[u] + \epsilon.$$

This implies

$$\overline{\int_a^b} u - \epsilon \leqslant \sum_{i=1}^{k(\pi)} u(\xi_i)(t_i - t_{i-1}) \leqslant \overline{\int_a^b} u$$

and

$$\underline{\int_a^b} u \leqslant \sum_{i=1}^{k(\pi)} u(\xi_i)(t_i - t_{i-1}) \leqslant \underline{\int_a^b} u + \epsilon,$$

which means that $\sum_{i=1}^{k(\pi)} u(\xi_i)(t_i - t_{i-1}) \to I = \underline{\int_a^b} u = \overline{\int_a^b} u$ as mesh $\pi \to 0$.

(vi) $\Rightarrow$ (i). Assume that $\sum_{i=1}^{k(\pi)} u(\xi_i)(t_i - t_{i-1}) \to I$, mesh $\pi \to 0$, exists for any choice of intermediate values. We have to show that $I = \overline{\int_a^b} u = \underline{\int_a^b} u$. By definition of the limit, there is some $\epsilon > 0$ and some partition π with mesh $\pi < \delta$ such that

$$I - \epsilon \leqslant \sum_{i=1}^{k(\pi)} u(\xi_i)(t_i - t_{i-1}) \leqslant I + \epsilon.$$

Since this must hold uniformly for any choice of intermediate values, we can pass to the infimum and supremum of these values and get

$$I - \epsilon \leqslant \sum_{i=1}^{k(\pi)} \inf_{\xi \in [t_{i-1}, t_i]} u(\xi)(t_i - t_{i-1}) \leqslant \sum_{i=1}^{k(\pi)} \sup_{\xi \in [t_{i-1}, t_i]} u(\xi)(t_i - t_{i-1}) \leqslant I + \epsilon.$$

Thus $I - \epsilon < S_\pi[u] \leqslant S^\pi[u] \leqslant I + \epsilon$, and

$$I - \epsilon < S_\pi[u] \leqslant \underline{\int_a^b} u \leqslant \overline{\int_a^b} u \leqslant S^\pi[u] \leqslant I + \epsilon. \qquad \square$$

Once we know that u is Riemann integrable, we can work out the value of the integral by taking particular Riemann sums.

Corollary I.6 *If $u: [a,b] \to \mathbb{R}$ is Riemann integrable, then the integral is the limit of Riemann sums*

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} u\big(\xi_i^{(n)}\big)\big(t_i^{(n)} - t_{i-1}^{(n)}\big),$$

where $\pi_n = \big\{a = t_0^{(n)} < t_1^{(n)} < \cdots < t_{k_n}^{(n)} = b\big\}$ is any sequence of partitions with mesh $\pi_n \to 0$ *and where $\xi_i^{(n)} \in \big[t_{i-1}^{(n)}, t_i^{(n)}\big]$ are some intermediate points.*

The existence of the limit of Riemann sums for some particular sequence of partitions does not guarantee integrability.

Example I.7 The Dirichlet jump function $u(x) := \mathbb{1}_{[0,1]\cap\mathbb{Q}}(x)$ on $[0,1]$ is not Riemann integrable, since for each partition π of $[0,1]$ we have $M_i = 1$ and $m_i = 0$, so that $\underline{\int_0^1} u = S_\pi[u] = 0$, while $\overline{\int_0^1} u = S^\pi[u] = 1$.

On the other hand, the equidistant Riemann sum

$$\sum_{i=1}^{k} u(\xi_i)\left(\frac{i}{k} - \frac{i-1}{k}\right) = \frac{1}{k}\sum_{i=1}^{k} u(\xi_i)$$

takes the value n/k, $0 \leqslant n \leqslant k$ if we choose $\xi_1, \ldots, \xi_n$ rational and $\xi_{n+1}, \ldots, \xi_k$ irrational. This allows us to construct sequences of Riemann sums which converge to any value in $[0,1]$.

Let us now find concrete functions which are Riemann integrable. A *step function* on $[a,b]$ is a function $f: [a,b] \to \mathbb{R}$ of the form

$$f(x) = \sum_{i=1}^{N} y_i \mathbb{1}_{I_i}(x),$$

where $N \in \mathbb{N}$, $y_i \in \mathbb{R}$ and I_i are (open, half-open, closed, even degenerate) adjacent intervals such that $I_1 \cup \cdots \cup I_N = [a,b]$ and $I_i \cap I_k$, $i \neq k$, intersect in at most one point. We denote by $\mathcal{T}[a,b]$ the family of all step functions on $[a,b]$.

Theorem I.8 *Continuous functions, monotone functions and step functions on $[a,b]$ are Riemann integrable.*

Proof Notice that the functions from all three classes are bounded on $[a,b]$.

Continuous functions. Let $u: [a,b] \to \mathbb{R}$ be continuous. Since $[a,b]$ is compact, u is uniformly continuous and we find for all $\epsilon > 0$ some $\delta > 0$ such that

$$|u(x) - u(y)| \leqslant \epsilon \quad \forall x, y \in [a,b],\ |x-y| < \delta.$$

If π is a partition of $[a, b]$ with mesh $\pi < \delta$ we find

$$S^{\pi}[u] - S_{\pi}[u] = \sum_{t_i \in \pi} (M_i - m_i)(t_i - t_{i-1}) \leqslant \epsilon \sum_{t_i \in \pi} (t_i - t_{i-1}) = \epsilon(b - a),$$

since, by uniform continuity,

$$M_i - m_i = \sup u([t_{i-1}, t_i]) - \inf u([t_{i-1}, t_i]) = \sup_{\xi, \eta \in [t_{i-1}, t_i]} \big(u(\xi) - u(\eta)\big) \leqslant \epsilon.$$

Thus $u \in \mathcal{R}[a, b]$ by Theorem I.5(iii).

Monotone functions. We can safely assume that $u : [a, b] \to \mathbb{R}$ is monotone increasing, otherwise we would consider $-u$. For the equidistant partition π_k with points $t_i = a + i((b - a)/k)$, $0 \leqslant i \leqslant k$, we get

$$\begin{aligned} S^{\pi_k}[u] - S_{\pi_k}[u] &= \sum_{i=1}^{k} (u(t_i) - u(t_{i-1}))(t_i - t_{i-1}) \\ &= \frac{b - a}{k} \sum_{i=1}^{k} (u(t_i) - u(t_{i-1})) = \frac{b - a}{k} (u(b) - u(a)), \end{aligned}$$

where we use that $\sup u([t_{i-1}, t_i]) = u(t_i)$ and $\inf u([t_{i-1}, t_i]) = u(t_{i-1})$ because of monotonicity. Since $((b - a)/k)(u(b) - u(a))$ can be made arbitrarily small, we have that $u \in \mathcal{R}[a, b]$ by Theorem I.5(ii).

Step functions. Let u be a step function which has value y_i on the interval I_i, $i = 1, \dots, k$. The endpoints of the non-degenerate intervals form a partition of $[a, b]$, $\pi = \{a = t_0 < t_1 < \cdots < t_N = b\}$, $N \leqslant k$, and we set for every $\epsilon > 0$

$$\pi_\epsilon := \big\{a = s'_0 < s_1 < s'_1 < s_2 < \cdots < s_{N-1} < s'_{N-1} < s_N = b\big\},$$

where $s_i < t_i < s'_i$, $1 \leqslant i \leqslant N - 1$, and $s'_i - s_i < \epsilon/(2N\|u\|_\infty)$. Since u is constant with value y_i on each interval $[s'_{i-1}, s_i]$, we find

$$\begin{aligned} &S^{\pi_\epsilon}[u] - S_{\pi_\epsilon}[u] \\ &\quad = \sum_{i=1}^{N} (y_i - y_i)(s_i - s'_{i-1}) + \sum_{i=1}^{N-1} \Big[\sup u([s_i, s'_i]) - \inf u([s_i, s'_i])\Big](s'_i - s_i) \\ &\quad \leqslant \sum_{i=1}^{N-1} 2\|u\|_\infty \frac{\epsilon}{2N\|u\|_\infty} \leqslant \epsilon. \end{aligned}$$

Therefore Theorem I.5(ii) proves that $u \in \mathcal{R}[a, b]$. $\square$

With somewhat more effort one can prove the following general theorem.

Theorem I.9 *Any bounded function $u: [a,b] \to \mathbb{R}$ with at most countably many points of discontinuity is Riemann integrable.*

An elementary proof of this, which is based on a compactness argument, can be found in Strichartz [52, Section 6.2.3], but, since Theorem 12.9 supersedes this result anyway, we do not include a proof here.

A combination of Theorems I.8 and I.5 yields the following quite useful criterion for integrability.

Corollary I.10 *$u \in \mathcal{R}[a,b]$ if, and only if, for every $\epsilon > 0$ there are $f, g \in \mathcal{T}[a,b]$ such that $f \leqslant u \leqslant g$ and $\int_a^b (g-f)dt \leqslant \epsilon$.*

Theorem I.11 *The Riemann integral is a positive linear form on the vector lattice $\mathcal{R}[a,b]$, that is, for all $\alpha, \beta \in \mathbb{R}$ and $u, w \in \mathcal{R}[a,b]$ one has the following.*

(i) $\alpha u + \beta w \in \mathcal{R}[a,b]$ *and* $\displaystyle\int_a^b (\alpha u + \beta w) dt = \alpha \int_a^b u\,dt + \beta \int_a^b w\,dt.$

(ii) $u \leqslant w \implies \displaystyle\int_a^b u\,dt \leqslant \int_a^b w\,dt.$

(iii) $u \vee w,\ u \wedge w,\ u^+,\ u^-,\ |u| \in \mathcal{R}[a,b]$ *and* $\displaystyle\left|\int_a^b u\,dt\right| \leqslant \int_a^b |u|dt.$

(iv) $|u|^p,\ uw \in \mathcal{R}[a,b],\ 1 \leqslant p < \infty.$

(v) *If $\phi: \mathbb{R}^2 \to \mathbb{R}$ is a Lipschitz continuous function*[3] *then $\phi \circ (u,w) \in \mathcal{R}[a,b]$.*

Proof (i) follows from the linearity of the criterion in Theorem I.5(iv).

(ii) In view of (i) it suffices to show that $v := w - u \geqslant 0$ entails $\int_a^b v\,dt \geqslant 0$. This, however, is clear, since $v \in \mathcal{R}[a,b]$ and

$$0 \leqslant \underline{\int_a^b} v = \int_a^b v\,dt.$$

(iii) and (iv) follow immediately from (v) since the functions $(x,y) \mapsto x \vee y$, $(x,y) \mapsto x \wedge y$, $x \mapsto x \vee 0$, $x \mapsto (-x) \vee 0$ and $x \mapsto |x|^p$, $p \geqslant 1$, are Lipschitz continuous.

The estimate in part (iii) can be derived from parts (i) and (ii) since $\pm u \leqslant |u|$ entails $\pm \int_a^b u\,dt \leqslant \int_a^b |u|dt$, which implies $\left|\int_a^b u\,dt\right| \leqslant \int_a^b |u|dt$.

(v) Let $u, w \in \mathcal{R}[a,b]$. Since any Riemann integrable function is, by definition, bounded, there is some $r < \infty$ such that $-r \leqslant u(x), w(x) \leqslant r$; write $L = L(r)$ for

[3] That is, for every ball $B_r(0) \subset \mathbb{R}^2$ there is a (so-called *Lipschitz*) constant $L = L(r) < \infty$ such that for all $(x,x'), (y,y') \in B_r(0)$

$$|\phi(x,x') - \phi(y,y')| \leqslant L\big(|x-y| + |x'-y'|\big).$$

the corresponding Lipschitz constant of $\phi|_{B_r(0)}$. Observe that for any partition $\pi = \{a = t_0 < t_1 < \cdots < t_k = b\}$ of $[a, b]$ we have

$$\begin{aligned} S^\pi&[\phi \circ (u, w)] - S_\pi[\phi \circ (u, w)] \\ &= \sum_{i=1}^{k} \Big[\sup_{s \in [t_{i-1}, t_i]} \phi(u(s), w(s)) - \inf_{t \in [t_{i-1}, t_i]} \phi(u(t), w(t)) \Big] (t_i - t_{i-1}) \\ &= \sum_{i=1}^{k} \sup_{s,t \in [t_{i-1}, t_i]} \big[\phi(u(s), w(s)) - \phi(u(t), w(t)) \big] (t_i - t_{i-1}) \\ &\leqslant L \sum_{i=1}^{k} \sup_{s,t \in [t_{i-1}, t_i]} \big[|u(s) - u(t)| + |w(s) - w(t)| \big] (t_i - t_{i-1}). \end{aligned}$$

Because of the symmetric rôles played by s and t,

$$\sup_{s,t} |u(s) - u(t)| = \sup_{s,t} \max\{u(s) - u(t), u(t) - u(s)\} = \sup_{s,t} (u(s) - u(t)),$$

and this proves

$$\begin{aligned} S^\pi&[\phi \circ (u, w)] - S_\pi[\phi \circ (u, w)] \\ &\leqslant L \sum_{i=1}^{k} \sup_{s,t \in [t_{i-1}, t_i]} \big(u(s) - u(t) \big) (t_i - t_{i-1}) \\ &\quad + L \sum_{i=1}^{k} \sup_{s,t \in [t_{i-1}, t_i]} \big(w(s) - w(t) \big) (t_i - t_{i-1}) \\ &= L \big(S^\pi[u] - S_\pi[u] \big) + L \big(S^\pi[w] - S_\pi[w] \big). \end{aligned}$$

Thus, if both u and w are Riemann integrable, so is $\phi \circ (u, w)$. □

Note that Theorem I.11(iii) has no converse: $|u| \in \mathcal{R}[a, b]$ does not imply that $u \in \mathcal{R}[a, b]$ (as is the case for the Lebesgue integral, see Theorem 10.3). This can be seen from the modified Dirichlet jump function $u := \mathbb{1}_{[0,1] \cap \mathbb{Q}} - \mathbb{1}_{[0,1] \setminus \mathbb{Q}}$, which is not Riemann integrable but whose modulus $|u| = \mathbb{1}_{[0,1]}$ is Riemann integrable.

Corollary I.12 (mean value theorem for integrals) *Let $u \in \mathcal{R}[a, b]$ be either positive or negative and let $v \in C[a, b]$. Then there exists some $\xi \in (a, b)$ such that*

$$\int_a^b u(t) v(t) dt = v(\xi) \int_a^b u(t) dt. \tag{I.5}$$

Proof The case $u \leqslant 0$ being similar, we may assume that $u \geqslant 0$. By Theorem I.8 and I.11(iv), uv is integrable and because of I.11(ii) we have

$$\inf v([a,b]) \int_a^b u(t)dt \leqslant \int_a^b u(t)v(t)dt \leqslant \sup v([a,b]) \int_a^b u(t)dt.$$

Since v is continuous on $[a,b]$, the intermediate value theorem guarantees the existence of some $\xi \in (a,b)$ such that (I.5) holds. □

Theorem I.13 *Let $[c,d] \subset [a,b]$. Then $\mathcal{R}[a,b] \subset \mathcal{R}[c,d]$ in the sense that every $u \in \mathcal{R}[a,b]$ satisfies $u|_{[c,d]} \in \mathcal{R}[c,d]$. Moreover, for any $u \in \mathcal{R}[a,b]$*

$$\int_a^b u\,dt = \int_a^c u\,dt + \int_c^b u\,dt.$$

Proof By Theorems I.8 and I.11 we find that $u\mathbb{1}_{[c,d]} \in \mathcal{R}[a,b]$. Since we can always add the points c and d to any of the partitions appearing in one of the criteria of Theorem I.5, we see that $u|_{[c,d]} = (u\mathbb{1}_{[c,d]})|_{[c,d]} \in \mathcal{R}[c,d]$ and

$$\int_a^b u\mathbb{1}_{[c,d]}\,dt = \int_c^d u\,dt.$$

Considering $u = u\mathbb{1}_{[a,c]} + u\mathbb{1}_{[c,b]}$ proves also the formula in the statement of the theorem. □

The Fundamental Theorem of Integral Calculus

Since by Theorem I.13 $\mathcal{R}[a,x] \subset \mathcal{R}[a,b]$, we can treat $\int_a^x u(t)dt$, $u \in \mathcal{R}[a,b]$, as a function of its upper limit $x \in [a,b]$.

Lemma I.14 *For every $u \in \mathcal{R}[a,b]$ the function $U(x) := \int_a^x u(t)dt$ is continuous for all $x \in [a,b]$.*

Proof Since u is bounded, $M := \sup_{x\in[a,b]} |u(x)| < \infty$. For all $x, y \in [a,b]$, $x < y$, we have by Theorems I.11 and I.13

$$\begin{aligned}|U(y) - U(x)| &= \left| \int_a^y u(t)dt - \int_a^x u(t)dt \right| \\ &= \left| \int_x^y u(t)dt \right| \leqslant \int_x^y |u(t)|dt \leqslant M(y-x) \xrightarrow[x-y\to 0]{} 0,\end{aligned}$$

showing even uniform continuity. □

We can now discuss the connection between differentiation and integration. Let us begin with a few examples.

Example I.15 (i) Let $[0,1] \subsetneq [a,b]$. Then $u(x) = \mathbb{1}_{[0,1]}(x)$ is an integrable function and

$$U(x) := \int_a^x u(t)\,dt = \left\{ \begin{array}{ll} 0, & \text{if } x \leqslant 0, \\ x, & \text{if } 0 < x < 1, \\ 1, & \text{if } x \geqslant 1, \end{array} \right\} = x^+ \wedge 1.$$

Note that $U'(x)$ does not exist at $x=0$ or $x=1$, so $u(x)$ cannot be the derivative of any function (at every point).

(ii) Let $[a,b]=[0,1]$ and take an enumeration $(q_i)_{i\in\mathbb{N}}$ of $[0,1]\cap\mathbb{Q}$. Then the function

$$u(x) := \sum_{i\,:\,q_i \leqslant x} 2^{-i} = \sum_{i=1}^{\infty} 2^{-i}\mathbb{1}_{[q_i,1]}(x), \qquad x \in [0,1],$$

is increasing and satisfies $0 \leqslant u \leqslant 1$, and its discontinuities are jumps at the points q_i of height $u(q_i+) - u(q_i-) = 2^{-i}$ – this is as bad as it can get for a monotone function, see Lemma 14.14. By Theorem I.8 u is integrable, and, since $(q_i)_{i\in\mathbb{N}}$ is dense, there is no interval $[c,d]\subset[0,1]$ such that $U'(x)=u(x)$ for all $x\in(c,d)$ for any function $U(x)$.

(iii) Consider on $[-1,1]$ the function

$$u(x) := \begin{cases} x^2 \sin\left(\dfrac{1}{x^2}\right), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

It is an elementary exercise to show that $u'(x)$ exists on $(-1,1)$ and

$$u'(x) = \begin{cases} 2x\sin\left(\dfrac{1}{x^2}\right) - \dfrac{2}{x}\cos\left(\dfrac{1}{x^2}\right), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

Thus u' exists everywhere, but it is not Riemann integrable in any neighbourhood of $x=0$ since u' is unbounded.

(iv) Let $(q_n)_{n\in\mathbb{N}}$ be an enumeration of $(0,1)\cap\mathbb{Q}$. The function

$$u(x) := \begin{cases} 2^{-n}, & \text{if } x = q_n,\ n\in\mathbb{N}, \\ 0, & \text{if } x \in \big([0,1]\setminus\mathbb{Q}\big)\cup\{0,1\}, \end{cases}$$

is discontinuous for every $x\in(0,1)\cap\mathbb{Q}$ and continuous otherwise. Moreover, we have $u\in\mathcal{R}[0,1]$, which follows from Theorem I.9 or directly from the following argument: fix $\epsilon>0$ and $n\in\mathbb{N}$ such that $2^{-n}<\epsilon$. Then choose a partition $\pi=\{0=t_0<t_1<\cdots<t_N=1\}$ with $\operatorname{mesh}\pi=\delta<\epsilon/n$ in such a way that each

q_k from $Q_n := \{q_1, q_2, \ldots, q_n\}$ is the midpoint of some $[t_{i-1}, t_i]$, $i=1,2,\ldots,N$. Therefore, if M_i denotes $\sup u([t_{i-1}, t_i])$,

$$
\begin{aligned}
0 \leqslant S_\pi[u] \leqslant S^\pi[u] &= \sum_{i=1}^{N} M_i(t_i - t_{i-1}) \\
&= \sum_{i\,:\,[t_{i-1},t_i]\cap Q_n \neq \emptyset} M_i(t_i - t_{i-1}) + \sum_{i\,:\,[t_{i-1},t_i]\cap Q_n = \emptyset} M_i(t_i - t_{i-1}) \\
&\leqslant n\frac{\epsilon}{n} + 2^{-n} \sum_{i\,:\,[t_{i-1},t_i]\cap Q_n = \emptyset} (t_i - t_{i-1}) \\
&\leqslant \epsilon + \epsilon \sum_{i=1}^{N} (t_i - t_{i-1}) = 2\epsilon.
\end{aligned}
$$

This proves $u \in \mathcal{R}[0,1]$ and $0 \leqslant \int_0^x u(t)dt \leqslant \int_0^1 u(t)dt = 0$. Thus $u'(x) = 0 \neq u(x)$ for all x from a dense subset.

The above examples show that the Riemann integral is not always the antiderivative, nor is the antiderivative an extension of the Riemann integral. The two concepts, however, coincide on a large class of functions.

Definition I.16 Let $u : [a,b] \to \mathbb{R}$ be a bounded function. Every function $U \in C[a,b]$ such that $U'(x) = u(x)$ for all but possibly finitely many $x \in (a,b)$ is called a *primitive* of u.

Obviously, primitives are unique only up to constants: for every constant c, $U + c$ is again a primitive of u. On the other hand, if U, W are two primitives of u, we have $U' - W' = 0$ at all but finitely many points $a = x_0 < x_1 < \cdots < x_n = b$. Thus the mean value theorem of differential calculus shows that $U = W +$ const. (see Rudin [42, Theorem 5.11]), first on each interval (x_{i-1}, x_i), $i = 1, 2, \ldots, n$, and then, by continuity, on the whole interval $[a,b]$.

Proposition I.17 *Every $u \in C[a,b]$ has $U(x) := \int_a^x u(t)dt$ as a primitive. Moreover,*

$$U(b) - U(a) = \int_a^b u(t)dt.$$

Proof Since continuous functions are integrable, $U(x)$ is well-defined by Theorem I.13 and continuous by Lemma I.14. For $a < x < x + h < b$ and sufficiently small h we find

$$\begin{aligned}\left|U(x+h)-U(x)-hu(x)\right| &= \left|\int_a^{x+h} u(t)dt - \int_a^x u(t)dt - \int_x^{x+h} u(x)dt\right| \\ &= \left|\int_x^{x+h} \big(u(t)-u(x)\big)dt\right| \\ &\leqslant \int_x^{x+h} |u(t)-u(x)|\,dt \\ &\leqslant \int_x^{x+h} \epsilon\,dt = \epsilon h,\end{aligned}$$

where we use that $u(t)$ is continuous at $t=x$. With a similar calculation we get

$$|U(x)-U(x-h)-hu(x)| \leqslant \epsilon h,$$

and a combination of these two inequalities shows that

$$\lim_{y\to x}\frac{U(y)-U(x)}{y-x}=u(x).$$

The formula $U(b)-U(a)=\int_a^b u(t)dt$ follows from the fact that $U(a)=0$. $\square$

Theorem I.18 (fundamental theorem of calculus) *Assume that U is a primitive of $u\in\mathcal{R}[a,b]$. Then*

$$U(b)-U(a)=\int_a^b u(t)dt.$$

Proof Let C be some finite set such that $U'(x)=u(x)$ if $x\in(a,b)\setminus C$. Fix $\epsilon>0$. Since u is integrable, we find by Theorem I.5(ii) a partition π of $[a,b]$ such that $S^\pi[u]-S_\pi[u]\leqslant\epsilon$. Because of Lemma I.1 this inequality still holds for the partition $\pi':=\pi\cup C$, whose points we denote by $a=t_0<t_1<\cdots<t_k=b$. Since

$$U(b)-U(a)=\sum_{i=1}^k (U(t_i)-U(t_{i-1}))$$

and since U is differentiable in each segment (t_{i-1},t_i) and continuous on $[a,b]$, we can use the mean value theorem of differential calculus to find points $\xi_i\in(t_{i-1},t_i)$ with

$$U(t_i)-U(t_{i-1})=U'(\xi_i)(t_i-t_{i-1})=u(\xi_i)(t_i-t_{i-1}),\qquad 1\leqslant i\leqslant k.$$

Using $m_i=\inf u([t_{i-1},t_i])\leqslant u(\xi_i)\leqslant\sup u([t_{i-1},t_i])=M_i$ we can sum the above equality over $i=1,\ldots,k$ and get

$$S^{\pi'}[u]-\epsilon\leqslant S_{\pi'}[u]\leqslant U(b)-U(a)\leqslant S^{\pi'}[u]\leqslant S_{\pi'}[u]+\epsilon.$$

By integrability, $S_{\pi'}[u] \leqslant \int_a^b u\,dt \leqslant S^{\pi'}[u]$, and this shows

$$\int_a^b u\,dt - \epsilon \leqslant U(b) - U(a) \leqslant \int_a^b u\,dt + \epsilon \qquad \forall \epsilon > 0,$$

which proves our claim. □

Remark I.19 There is not much room to improve the fundamental theorem. On the one hand, Example I.15(ii) shows that an integrable function need not have a primitive and Example I.15(iv) gives an example where $\int_a^x u\,dt$ exists, but is not a primitive in any interval; on the other hand, Example I.15(iii) provides an example of a function u' which has a primitive u but is not itself Riemann integrable since it is *unbounded*. Volterra even constructed an example of a *bounded* but not Riemann integrable function with a primitive, see Sz.-Nagy [32, pp. 154–156].

The desire to overcome this phenomenon was one of the motivations for Lebesgue when he introduced the *Lebesgue integral*. In fact, *every bounded function f on the interval $[a, b]$ with a primitive F is Lebesgue integrable*: indeed, since F is continuous, it is measurable in the sense of Chapter 8 and so is the limit $f(x) = \lim_{n\to\infty}(F(x + \frac{1}{n}) - F(x))/\frac{1}{n}$, see Corollary 8.10. The finitely many points where the limit does not exist are a Lebesgue null set and pose no problem. Since $|f|$ is dominated by the (Lebesgue) integrable function $M\mathbb{1}_{[a,b]}$, $M := \sup f([a, b])$, we conclude that $f \in \mathcal{L}^1[a, b]$.

An immediate consequence of the integral as antiderivative are the following integration formulae which are easily proved by 'integrating up' the corresponding differentiation rules.

Theorem I.20 (integration by parts) *Let u' and v' be integrable functions on $[a, b]$ with primitives u and v. Then uv is a primitive of $u'v + uv'$ and, in particular,*

$$\int_a^b u'(t)v(t)dt = u(b)v(b) - u(a)v(a) - \int_a^b u(t)v'(t)dt.$$

Theorem I.21 (integration by substitution) *Let $\phi : [c, d] \to [a, b]$ be a strictly increasing differentiable function such that $\phi(c) = a$ and $\phi(d) = b$ and $u \in \mathcal{R}[a, b]$. If $u \circ \phi$, $\phi' \in \mathcal{R}[c, d]$ and u has a primitive U, then $U \circ \phi$ is a primitive of $u \circ \phi \cdot \phi'$ as well as*

$$\int_a^b u(t)dt = \int_c^d u(\phi(s))\phi'(s)ds = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} u(\phi(s))\phi'(s)ds.$$

Corollary I.22 (Bonnet's mean value theorem[4]) *Let $u, v \in \mathcal{R}[a,b]$ have primitives U and V. If $u \leqslant 0$ [resp. $u \geqslant 0$] and $U \geqslant 0$, then there exists some $\xi \in (a,b)$ such that*

$$\int_a^b U(t)v(t)dt = U(a)\int_a^\xi v(t)dt. \tag{I.6}$$

$$\left[resp.\ \int_a^b U(t)v(t)dt = U(b)\int_\xi^b v(t)dt.\right] \tag{I.6$'$}$$

Proof On subtracting a suitable constant from V we may assume that $V(a) = 0$ and, by the fundamental theorem, $V(a) = \int_a^x v(t)dt$. Integration by parts shows that

$$\int_a^b U(t)v(t)dt = U(b)V(b) - \int_a^b u(t)V(t)dt.$$

Since $u \leqslant 0$ we get

$$\begin{aligned}\int_a^b U(t)v(t)dt &\leqslant U(b)V(b) - \sup V([a,b])\int_a^b u(t)dt\\ &= U(b)V(b) - \sup V([a,b])(U(b) - U(a))\\ &= U(b)\big(V(b) - \sup V([a,b])\big) + \sup V([a,b])U(a)\\ &\leqslant \sup V([a,b])U(a),\end{aligned}$$

and a similar calculation yields the lower inequality:

$$\inf V([a,b])U(a) \leqslant \int_a^b U(t)v(t)dt \leqslant \sup V([a,b])U(a).$$

Applying the intermediate value theorem to the continuous function V furnishes some $\xi \in (a,b)$ such that (I.6) holds. □

Integrals and Limits

One of the strengths of Lebesgue integration is the fact that we have fairly general theorems that allow us to interchange pointwise limits and Lebesgue integrals.

Similar results for the Riemann integral regularly require uniform convergence. Recall that a sequence of functions $(u_n(\cdot))_{n\in\mathbb{N}}$ on $[a,b]$ *converges uniformly* (in x) to u, if

$$\forall \epsilon > 0\ \exists N_\epsilon \in \mathbb{N}\ \forall x \in [a,b],\ \forall n \geqslant N_\epsilon : |u_n(x) - u(x)| \leqslant \epsilon.$$

The basic convergence result for the Riemann integral is the following.

[4] This is also known as the second mean value theorem of integral calculus.

Theorem I.23 *Let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{R}[a,b]$ be a sequence which converges uniformly to a function u. Then $u \in \mathcal{R}[a,b]$ and*

$$\lim_{n\to\infty} \int_a^b u_n\,dt = \int_a^b \lim_{n\to\infty} u_n\,dt = \int_a^b u\,dt.$$

Proof Let π be a partition of $[a,b]$ and let $\epsilon > 0$ be given. Since $u_n \to u$ uniformly, we can find some $N_\epsilon \in \mathbb{N}$ such that $|u(x) - u_n(x)| \leqslant \epsilon/(b-a)$ uniformly in $x \in [a,b]$ for all $n \geqslant N_\epsilon$. Because of (I.1) we find for all $n \geqslant N_\epsilon$

$$\begin{aligned} S^\pi[u] - S_\pi[u] &= S^\pi[u-u_n] + S^\pi[u_n] - S_\pi[u_n] - S_\pi[u-u_n] \\ &\leqslant 2\epsilon + S^\pi[u_n] - S_\pi[u_n], \end{aligned}$$

thus

$$\overline{\int_a^b} u - \underline{\int_a^b} u \leqslant 2\epsilon + S^\pi[u_n] - S_\pi[u_n] \quad \forall n \geqslant N_\epsilon.$$

On fixing some $n_0 \geqslant N_\epsilon$ we can use that u_{n_0} is integrable and choose π in such a way that $S^\pi[u_{n_0}] - S_\pi[u_{n_0}] \leqslant \epsilon$. This shows that $\overline{\int_a^b} u - \underline{\int_a^b} u \leqslant 3\epsilon$ and $u \in \mathcal{R}[a,b]$.

Once u is known to be integrable, we get for all $n \geqslant N_\epsilon$

$$\left| \int_a^b (u-u_n)dt \right| \leqslant \int_a^b |u-u_n|dt \leqslant \epsilon(b-a) \xrightarrow[\epsilon\to 0]{} 0. \qquad \square$$

We can now consider Riemann integrals which depend on a parameter.

Theorem I.24 (continuity theorem) *Let $u : [a,b] \times \mathbb{R} \to \mathbb{R}$ be a continuous function. Then*

$$w(y) := \int_a^b u(t,y)dt$$

is continuous for all $y \in \mathbb{R}$.

Proof Since $u(\,\cdot\,,y)$ is continuous, the above Riemann integral exists. Fix $y \in \mathbb{R}$ and consider any sequence $(y_n)_{n\in\mathbb{N}}$ with limit y. Without loss of generality we can assume that $(y_n)_{n\in\mathbb{N}} \subset I := [y-1, y+1]$. Since $[a,b] \times I$ is compact, $u|_{[a,b]\times I}$ is uniformly continuous, and we can find for all $\epsilon > 0$ some $\delta > 0$ such that

$$\sqrt{(t-\tau)^2 + (y-\eta)^2} < \delta \implies |u(t,y) - u(\tau,\eta)| < \epsilon.$$

As $y_n \to y$, there is some $N_\epsilon \in \mathbb{N}$ with

$$|u(t,y_n) - u(t,y)| < \epsilon \quad \forall t \in [a,b],\ \forall n \geqslant N_\epsilon,$$

i.e. $u(y_n, t) \to u(y, t)$ uniformly in $t \in [a, b]$. Theorem I.23 and the continuity of $u(t, \cdot)$ therefore show that

$$\lim_{n\to\infty} w(y_n) = \lim_{n\to\infty} \int_a^b u(t, y_n) dt = \int_a^b \lim_{n\to\infty} u(t, y_n) dt = \int_a^b u(t, y) dt = w(y)$$

which is merely the continuity of w at y. □

Theorem I.25 (differentiation theorem) *Let $u\colon [a, b] \times \mathbb{R} \to \mathbb{R}$ be a continuous function with continuous partial derivative $\frac{\partial}{\partial y} u(t, y)$. Then*

$$w(y) := \int_a^b u(t, y) dt$$

is continuously differentiable and

$$w'(y) = \frac{d}{dy} \int_a^b u(t, y) dt = \int_a^b \frac{\partial}{\partial y} u(t, y) dt.$$

Proof Since $u(\cdot\,, y)$ and $\frac{\partial}{\partial y} u(\cdot\,, y)$ are continuous, the above integrals do exist. Fix $y \in \mathbb{R}$ and consider any sequence $(y_n)_{n\in\mathbb{N}}$ with limit y. Without loss of generality we can assume that $(y_n)_{n\in\mathbb{N}} \subset I := [y - 1, y + 1]$.

We introduce the following auxiliary function:

$$h(t, z) := u(t, z) - u(t, y) - \frac{\partial}{\partial y} u(t, y)(z - y).$$

Clearly, $h(t, y) = 0$ and we have that $\frac{\partial}{\partial z} h(t, z) = \frac{\partial}{\partial z} u(t, z) - \frac{\partial}{\partial y} u(t, y)$ is continuous and uniformly continuous on $[a, b] \times I$, i.e. for all $\epsilon > 0$ there is some $\delta > 0$ such that

$$\sqrt{(t - \tau)^2 + (z - \zeta)^2} < \delta \Longrightarrow \left| \frac{\partial}{\partial z} h(t, z) - \frac{\partial}{\partial \zeta} h(\tau, \zeta) \right| < \epsilon.$$

From the mean value theorem of differential calculus we infer that for some ζ between z and y

$$\begin{aligned} |h(t, z)| = |h(t, z) - h(t, y)| &= \left| \frac{\partial}{\partial \zeta} h(\tau, \zeta) \right| \cdot |z - y| \\ &= \left| \frac{\partial}{\partial \zeta} h(\tau, \zeta) - \frac{\partial}{\partial y} h(t, y) \right| \cdot |z - y| \\ &\leqslant \epsilon |z - y| \end{aligned}$$

whenever $z, y \in I$ and $|z - y| < \delta$. This shows that for some $N_\epsilon \in \mathbb{N}$

$$\left| u(t, y_n) - u(t, y) - \frac{\partial}{\partial y} u(t, y)(y_n - y) \right| \leqslant \epsilon |y_n - y| \quad \forall t \in [a, b],\ \forall n \geqslant N_\epsilon.$$

Theorem I.23 now shows that

$$w'(y) = \lim_{n\to\infty} \frac{w(y_n) - w(y)}{y_n - y} = \lim_{n\to\infty} \int_a^b \frac{u(t, y_n) - u(t, y)}{y_n - y}\, dt$$
$$= \int_a^b \lim_{n\to\infty} \frac{u(t, y_n) - u(t, y)}{y_n - y}\, dt = \int_a^b \frac{\partial}{\partial y} u(t, y) dt. \qquad \square$$

Improper Riemann Integrals

Let us finally have a glance at various extensions of the Riemann integral to unbounded intervals and/or unbounded integrands. The following cases can occur:

(A) the interval of integration is $[a, +\infty)$ or $(-\infty, b]$;
(B) the interval of integration is $[a, b)$ or $(a, b]$, and the integrand $u(t)$ is unbounded as $t \uparrow b$, resp. $t \downarrow a$;
(C) the interval of integration is (a, b) with $-\infty \leqslant a < b \leqslant +\infty$ and the integrand may or may not be unbounded.

(A) *Improper Riemann integrals of the type $\int_a^\infty u\, dt$ and $\int_{-\infty}^b u\, dt$*

Definition I.26 If $u \in \mathcal{R}[a, b]$ for all $b \in (a, \infty)$ [resp. $a \in (-\infty, b)$] and, if the limit

$$\lim_{b\to\infty} \int_a^b u\, dt \qquad \left[\text{resp. } \lim_{a\to-\infty} \int_a^b u\, dt\right]$$

exists and is finite, we call u *improperly Riemann integrable* and write $u \in \mathcal{R}[a, \infty)$ [resp. $u \in \mathcal{R}(-\infty, b]$]. The value of the limit is called the (*improper Riemann*) *integral* and denoted by $\int_a^\infty u\, dt$ $\left[\text{resp. } \int_{-\infty}^b u\, dt\right]$.

The *typical examples* of improper integrals of this kind are expressions of the type $\int_1^\infty t^\lambda\, dt$ if $\lambda < 0$. In fact, if $\lambda \neq -1$,

$$\int_1^\infty t^\lambda = \lim_{b\to\infty} \int_1^b t^\lambda\, dt = \lim_{b\to\infty} \frac{1}{\lambda+1}(b^{\lambda+1} - 1) = \begin{cases} \dfrac{-1}{\lambda+1} & \text{if } \lambda < -1, \\ \infty & \text{if } \lambda > -1, \end{cases}$$

and a similar calculation confirms that $\int_1^\infty t^{-1}\, dt = \infty$. Thus $t^\lambda \in \mathcal{R}[1, \infty)$ if, and only if, $\lambda < -1$.

From now on we will consider only integrals of the type $\int_a^\infty u\, dt$; the case of a finite upper and infinite lower limit is very similar. The following Cauchy criterion for improper integrals is quite useful.

Lemma I.27 *$u \in \mathcal{R}[a,\infty)$ if, and only if, $u \in \mathcal{R}[a,b]$ for all $b \in (a,\infty)$ and $\lim_{x,y\to\infty} \int_x^y u\,dt = 0$ ($x, y \to \infty$ simultaneously).*

Proof Use Cauchy's convergence criterion for $U(z) = \int_a^z u(t)dt$ as $z \to \infty$. □

It is not hard to see that Lemma I.27 implies, in particular, that

- $\mathcal{R}[a,\infty)$ is a vector space, i.e. for all $\alpha, \beta \in \mathbb{R}$ and $u, w \in \mathcal{R}[a,\infty)$
$$\int_a^\infty (\alpha u + \beta w)dt = \alpha \int_a^\infty u\,dt + \beta \int_a^\infty w\,dt;$$
- $u \in \mathcal{R}[a,\infty)$ if, and only if, $\int_b^\infty u\,dt$ exists for all $b > a$.

Corollary I.28 *Let $u, w : [a,\infty) \to \mathbb{R}$ be two functions such that $|u| \leqslant w$. If $w \in \mathcal{R}[a,\infty)$ and if $u \in \mathcal{R}[a,b]$ for all $b > a$, then $u,\ |u| \in \mathcal{R}[a,\infty)$. In particular, $|u| \in \mathcal{R}[a,\infty)$ implies that $u \in \mathcal{R}[a,\infty)$.*

Proof For all $y > x > a$ we find using Theorem I.11 and Lemma I.27 that
$$\left| \int_x^y u\,dt \right| \leqslant \int_x^y |u|\,dt \leqslant \int_x^y w\,dt \xrightarrow[x,y\to\infty]{} 0$$
which shows, again by Lemma I.27, that $u,\ |u| \in \mathcal{R}[a,\infty)$. □

Unlike Lebesgue integrals, improper Riemann integrals are not *absolute integrals* since improper integrability of u does NOT imply improper integrability of $|u|$, see e.g. Example 12.12, where $\int_0^\infty \sin t/t\,dt$ is discussed. This means that the following convergence theorems for improper Riemann integrals are not necessarily covered by Lebesgue's theory.

Theorem I.29 *Let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{R}[a,\infty)$. If for some $u : [a,\infty) \to \mathbb{R}$*

- $\lim_{n\to\infty} \sup_{a\leqslant t\leqslant b} |u_n(t) - u(t)| = 0$ *for every $b > a$,*
- $\lim_{b\to\infty} \int_a^b u_n\,dt$ *exists uniformly for all $n \in \mathbb{N}$, i.e. for every $\epsilon > 0$ there is some $N_\epsilon \in \mathbb{N}$ such that*
$$\sup_{n\in\mathbb{N}} \left| \int_x^y u_n\,dt \right| < \epsilon \quad \forall y > x > N_\epsilon,$$

then $u \in \mathcal{R}[a,\infty)$ and
$$\lim_{n\to\infty} \int_a^\infty u_n\,dt = \int_a^\infty \lim_{n\to\infty} u_n\,dt = \int_a^\infty u\,dt.$$

Proof That $u \in \mathcal{R}[a,b]$ for all $b > a$ follows from Theorem I.23. Fix $\epsilon > 0$ and choose N_ϵ as in the above statement. For all $y > x > N_\epsilon$

$$\left| \int_x^y u\,dt \right| \leqslant \left| \int_x^y (u - u_n)dt \right| + \left| \int_x^y u_n\,dt \right| \leqslant (y - x) \sup_{t \in [x,y]} |u(t) - u_n(t)| + \epsilon,$$

and as $n \to \infty$ we find $\left| \int_x^y u\,dt \right| \leqslant \epsilon$ for all $y > x > N_\epsilon$, and hence $u \in \mathcal{R}[a, \infty)$ by Lemma I.27. □

In pretty much the same way as we derived Theorems I.24 and I.25 from the basic convergence result Theorem I.23, we get now from Theorem I.29 the following continuity and differentiability theorems for improper integrals.

Theorem I.30 *Let $I \subset \mathbb{R}$ be an open interval and $u : [a, \infty) \times I \to \mathbb{R}$ be continuous such that $u(\,\cdot\,, y) \in \mathcal{R}[a, \infty)$ for all $y \in I$ and*

$$\lim_{b \to \infty} \int_a^b u(t,y)dt \quad \textit{exists uniformly for all} \quad y \in [c,d] \subset I.$$

Then $U(y) := \int_a^\infty u(t,y)dt$ is continuous for all $y \in (c,d)$.

Sketch of the proof Fix $y \in (c,d)$ and choose any sequence $(y_n)_{n \in \mathbb{N}} \subset (c,d)$ with limit y. By the assumptions $u_n(t) := u(t, y_n) \to u(t,y)$ uniformly for all $t \in [a,b]$. Now the basic convergence theorem for improper integrals, Theorem I.29, applies and shows that $U(y_n) \to U(y)$. □

Theorem I.31 *Let $I \subset \mathbb{R}$ be an open interval and $u : [a, \infty) \times I \to \mathbb{R}$ be continuous with continuous partial derivative $\frac{\partial}{\partial y} u(t,y)$. If $u(\,\cdot\,, y)$, $\frac{\partial}{\partial y} u(t,y) \in \mathcal{R}[a, \infty)$ for all $y \in I$, and if*

$$\lim_{b \to \infty} \int_a^b u(t,y)dt \quad \textit{and} \quad \lim_{b \to \infty} \int_a^b \frac{\partial}{\partial y} u(t,y)dt$$

exist uniformly for all $y \in [c,d] \subset I$, then $W(y) := \int_a^\infty u(t,y)dt$ exists and is differentiable on (c,d) with derivative

$$W'(y) = \frac{d}{dy} \int_a^\infty u(t,y)dt = \int_a^\infty \frac{\partial}{\partial y} u(t,y)dt.$$

Sketch of the proof Set $U(x,y) := \int_a^x u(t,y)dt$. By Theorem I.25 $\frac{\partial}{\partial y} U(x,y)$ exists and equals $\int_a^x \frac{\partial}{\partial y} u(t,y)dt$. By assumption,

$$U(x,y) \xrightarrow[x \to \infty]{} \int_a^\infty u(t,y)dt \qquad \text{pointwise for all } y \in [c,d],$$

$$\frac{\partial}{\partial y} U(x,y) \xrightarrow[x \to \infty]{} \int_a^\infty \frac{\partial}{\partial y} u(t,y)dt \qquad \text{uniformly for all } y \in [c,d].$$

By a standard theorem on uniform convergence and differentiability, see Rudin [42, Theorem 7.17], we now conclude

$$\frac{d}{dy}\int_a^\infty u(t,y)dt = \int_a^\infty \frac{\partial}{\partial y}u(t,y)dt. \qquad \square$$

Theorem I.32 *Let $u, w \in \mathcal{R}[a,b]$ for all $b \in (a,\infty)$ and assume that $u, w \geqslant 0$ and $\lim_{x\to\infty} u(x)/w(x) = A > 0$ exists. Then $u \in \mathcal{R}[a,\infty)$ if, and only if, $w \in \mathcal{R}[a,\infty)$.*

Proof By assumption we find for every $\epsilon > 0$ some $N_\epsilon \in \mathbb{N}$ such that

$$0 < A - \epsilon \leqslant \frac{u(x)}{w(x)} \leqslant A + \epsilon \quad \forall x \geqslant N_\epsilon\,(>a).$$

Thus $(A-\epsilon)w(x) \leqslant u(x) \leqslant (A+\epsilon)w(x)$ for all $x \geqslant N_\epsilon$. Thus, if $w \in \mathcal{R}[a,\infty)$, we get $(A+\epsilon)w \in \mathcal{R}[a,\infty)$ (see the remark following Lemma I.27) and, by Corollary I.28, $u \in \mathcal{R}[a,\infty)$.

Similarly, if $u \in \mathcal{R}[a,\infty)$, we have $u/(A-\epsilon) \in \mathcal{R}[a,\infty)$ and, again by Theorem I.28, $w \in \mathcal{R}[a,\infty)$. $\square$

We will finally study the interplay of series and improper integrals.

Theorem I.33 *Let $a = b_0 < b_1 < b_2 < \cdots$ be a strictly increasing sequence with $b_k \to \infty$.*

(i) *If $u \in \mathcal{R}[a,\infty)$, then*

$$\sum_{k=1}^\infty \int_{b_{k-1}}^{b_k} u\,dt$$

converges.

(ii) *If $u \geqslant 0$ and $u \in \mathcal{R}[b_{k-1}, b_k]$ for all $k \in \mathbb{N}$, then the convergence of*

$$\sum_{k=1}^\infty \int_{b_{k-1}}^{b_k} u\,dt$$

implies $u \in \mathcal{R}[a,\infty)$.

Proof (i) Since $u \in \mathcal{R}[a,\infty)$,

$$\int_a^\infty u\,dt = \lim_{n\to\infty}\int_a^{b_n} u\,dt = \lim_{n\to\infty}\sum_{k=1}^n \int_{b_{k-1}}^{b_k} u\,dt = \sum_{k=1}^\infty \int_{b_{k-1}}^{b_k} u\,dt.$$

(ii) Define $S := \sum_{k=1}^{\infty} \int_{b_{k-1}}^{b_k} u\,dt$. Since b_k increases to ∞, we find for all $b > a$ some $N \in \mathbb{N}$ such that $b_N > b$. Consequently,

$$\int_a^b u\,dt \leqslant \int_a^{b_N} u\,dt = \sum_{k=1}^{N} \int_{b_{k-1}}^{b_k} u\,dt \leqslant S,$$

which shows that the limit $\lim_{b\to\infty} \int_a^b u\,dt = \sup_{b>0} \int_a^b u\,dt \leqslant S$ exists. □

Theorem I.34 (integral test for series) *Let $u \in C[0,\infty)$, $u \geqslant 0$, be a decreasing function. Then*

$$\int_0^{\infty} u\,dt \quad \textit{and} \quad \sum_{k=0}^{\infty} u(k)$$

either both converge or both diverge.

Proof Note that by Theorem I.8 $u \in \mathcal{R}[0,b]$ for all $b > 0$, so the improper integral can be defined. Since u is decreasing,

$$u(k+1) \leqslant \int_k^{k+1} u(t)dt \leqslant u(k),$$

see Theorem I.11, and summing these inequalities over $k = 0, 1, \dots, N$ yields

$$\sum_{k=1}^{N+1} u(k) = \sum_{k=0}^{N} u(k+1) \leqslant \int_0^{N+1} u(t)dt \leqslant \sum_{k=0}^{N} u(k).$$

Since u is positive and since the series has only positive terms, it is obvious that $\int_0^{\infty} u\,dt$ converges if, and only if, the series $\sum_{k=0}^{\infty} u(k)$ is finite. □

(B) *Improper Riemann Integrals with Unbounded Integrands*

Definition I.35 If $u \in \mathcal{R}[a,c]$ [resp. $u \in \mathcal{R}[c,b]$] for all $c \in (a,b)$ and if the limit

$$\lim_{c\uparrow b} \int_a^c u\,dt \quad \left[\text{resp. } \lim_{c\downarrow a} \int_c^b u\,dt\right]$$

exists and is finite, we call u *improperly Riemann integrable* and write $u \in \mathcal{R}[a,b)$ [resp. $u \in \mathcal{R}(a,b]$]. The value of the limit is called the (*improper Riemann*) integral and denoted by $\int_a^b u\,dt$.

Notice that the function u in Definition I.35 need not be bounded in (a,b). If it is, the improper integral coincides with the ordinary Riemann integral.

Lemma I.36 *If the function $u \in \mathcal{R}[a,b)$ [or $u \in \mathcal{R}(a,b]$] has an extension to $[a,b]$ which is bounded, then the extension is Riemann integrable over $[a,b]$, and proper and improper Riemann integrals coincide.*

Proof We consider only $[a,b)$, since the other case is similar. Denote, for notational simplicity, the extension of u again by u.

Let $M := \sup u([a,b])$, fix $\epsilon > 0$ and pick $c < b$ with $b - c \leqslant \epsilon/M$. Since $u \in \mathcal{R}[a,c]$, we can find a partition π of $[a,c]$ such that $S^{\pi}[u] - S_{\pi}[u] \leqslant \epsilon$. For the partition $\pi' := \pi \cup \{b\}$ of $[a,b]$ we get

$$S^{\pi'}[u] - S^{\pi}[u] = \sup u([c,b])\frac{\epsilon}{M} \leqslant M\frac{\epsilon}{M} = \epsilon$$

and

$$S_{\pi'}[u] - S_{\pi}[u] = \inf u([c,b])\frac{\epsilon}{M} \leqslant M\frac{\epsilon}{M} = \epsilon,$$

which implies that $S^{\pi'}[u] - S_{\pi'}[u] \leqslant 3\epsilon$ and $u \in \mathcal{R}[a,b]$ by Theorem I.5. The claim now follows from Lemma I.14. □

Many of the results for improper integrals of the form $\int_a^{\infty} u\,dt$, resp. $\int_{-\infty}^{b} u\,dt$, carry over with minor notational changes to the case of half-open bounded intervals. Note, however, that in the convergence theorems some assertions involving uniform convergence are senseless in the presence of unbounded integrands. We leave the details to the reader.

The *typical examples* of improper integrals of this kind are expressions of the type $\int_0^1 t^{\lambda}\,dt$ if $\lambda < 0$. In fact, if $\lambda \neq -1$,

$$\int_0^1 t^{\lambda} = \lim_{\epsilon \to 0} \int_{\epsilon}^1 t^{\lambda}\,dt = \lim_{\epsilon \to 0} \frac{1}{\lambda + 1}(1 - \epsilon^{\lambda+1}) = \begin{cases} \dfrac{1}{\lambda+1} & \text{if } \lambda > -1, \\ \infty & \text{if } \lambda < -1, \end{cases}$$

and a similar calculation confirms that $\int_0^1 t^{-1}\,dt = \infty$. Thus $t^{\lambda} \in \mathcal{R}(0,1]$ if, and only if, $\lambda > -1$.

(C) ***Improper Riemann Integrals Where Both Limits Are Critical***

Assume now that the integration interval is (a,b) and that both endpoints a and b, $-\infty \leqslant a < b \leqslant +\infty$, are critical, i.e. that the integrand is unbounded at one or both endpoints and/or that one or both endpoints are infinite.

Let $u \in \mathcal{R}(a,c] \cap \mathcal{R}[c,b)$ for some point $a<c<b$ and suppose that d satisfies $c<d<b$. By the remark following Lemma I.27 and Theorem I.13 we find

$$\begin{aligned}\int_a^c u\,dt + \int_c^b u\,dt &= \lim_{x\downarrow a}\int_x^c u\,dt + \lim_{y\uparrow b}\int_c^y u\,dt \\ &= \lim_{x\downarrow a}\int_x^c u\,dt + \int_c^d u\,dt + \lim_{y\uparrow b}\int_d^y u\,dt \\ &= \lim_{x\downarrow a}\int_x^d u\,dt + \lim_{y\uparrow b}\int_d^y u\,dt \\ &= \int_a^d u\,dt + \int_d^b u\,dt,\end{aligned}$$

which shows that $u \in \mathcal{R}(a,d] \cap \mathcal{R}[d,b)$. Therefore, the following definition makes sense.

Definition I.37 Let $-\infty \leqslant a < b \leqslant +\infty$ and let $(a,b) \subset \mathbb{R}$ be a bounded or unbounded open interval. Then $u:(a,b) \to \mathbb{R}$ is said to be *improperly integrable* if for some (hence, all) $c \in (a,b)$ the function u is improperly integrable both over $(a,c]$ and over $[c,b)$, i.e. we define $\mathcal{R}(a,b) := \mathcal{R}(a,c] \cap \mathcal{R}[c,b)$. The (*improper Riemann*) *integral* is then given by

$$\int_a^b u\,dt := \int_a^c u\,dt + \int_c^b u\,dt = \lim_{x\downarrow a}\int_x^c u\,dt + \lim_{y\uparrow b}\int_c^y u\,dt.$$

The *typical example* of an improper integral of this kind is Euler's gamma function

$$\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}\,dt, \quad x>0,$$

which is treated in Example 12.15 in the framework of Lebesgue theory, but the arguments are essentially similar. The gamma function is only for $0<x<1$ a two-sided improper integral, since for $x \geqslant 1$ it can be interpreted as a one-sided improper integral over $[0,\infty)$, see Lemma I.36.

References

Each reference is followed, in parentheses, by a list of page numbers where it is cited.

[1] Alexits, G.: *Convergence Problems of Orthogonal Series*. Oxford: Pergamon, 1961. (p. 367, 392)

[2] Bauer, H.: Approximation and abstract boundaries. *Am. Math. Monthly* **85** (1978) 632–647. (Also in H. Bauer *Selecta*. Berlin: De Gruyter, 2003, 436–451.) (p. 370)

[3] Bauer, H.: *Probability Theory*. Berlin: De Gruyter, 1996. (p. 400, 401)

[4] Benyamini, Y. and J. Lindenstrauss: *Geometric Nonlinear Functional Analysis*, vol. 1. Providence (RI): American Mathematical Society, 2000. (p. 303)

[5] Besicovitch, A. S.: On the fundamental geometrical properties of linearly measurable plane sets of points. *Math. Annalen* **98** (1928) 422–464. (p. 201)

[6] Boas, R. P.: *A Primer of Real Functions*. Washington (D.C.): Mathematical Association of America, 1960. (p. 123)

[7] Bochner, S.: *Vorlesungen über Fouriersche Integrale*. Leipzig: Akademische Verlagsgesellschaft, 1932. (Reprinted New York: Chelsea, 1948.) (p. 212)

[8] Carathéodory, C.: Über das lineare Maß von Punktmengen – eine Verallgemeinerung des Längenbegriffs. *Nachr. Kgl. Ges. Wiss. Göttingen Math.-Phys. Kl.* (1914) 404–426. (Also in C. Carathéodory *Gesammelte mathematische Schriften*. Munich: C. H. Beck, 1954–57, vol. 4, 249–275. (p. 39, 197)

[9] Ciesielski, K.: How good is Lebesgue measure. *Math. Intelligencer* **11** (1989) 54–58. (p. 45)

[10] Ciesielski, Z.: Hölder condition for realizations of Gaussian processes. *Trans. Am. Math. Soc.* **99** (1961) 403–413. (p. 402)

[11] Diestel, J. and J. J. Uhl Jr.: *Vector Measures*. Providence (RI): American Mathematical Society, 1977. (p. 303)

[12] Dieudonné, J.: Sur un théorème de Jessen. *Fundam. Math.* **37** (1950) 242–248. Also in J. Dieudonné *Choix d'œuvres mathématiques*, vol. 1. Paris: Hermann, 1981, 369–275. (p. 300)

[13] Doob, J. L.: *Stochastic Processes*. New York: Wiley, 1953. (p. 271, 306)

[14] Dudley, R. M.: *Real Analysis and Probability*. Cambridge: Cambridge University Press, 2002 (2nd edn). (First edn: Pacific Grove (CA): Wadsworth & Brooks/Cole, 1989.) (p. 428)

[15] Dunford, N. and J. T. Schwartz: *Linear Operators I*. New York: Wiley–Interscience, 1957. (p. 239, 261)

[16] Evans, L. C. and R. F. Gariepy: *Measure Theory and Fine Properties of Functions*. Boca Raton (FL): CRC Press, 1992. (p. 205, 420)

[17] Fichtenholz, G. M.: *Differential- und Integralrechnung I–III*. Berlin: VEB Deutscher Verlag der Wissenschaften, 1973–74. (p. 176)

[18] Garsia, A. M.: *Topics in Almost Everywhere Convergence*. Chicago (IL): Markham, 1970. (p. 311, 399)

[19] Gradshteyn, I. and I. Ryzhik: *Tables of Integrals, Series, and Products*. San Diego (CA): Academic Press, 1992 (4th corrected and enlarged edn). (p. 374)

[20] Gundy, R. F.: Martingale theory and pointwise convergence of certain orthogonal series. *Trans. Am. Math. Soc.* **124** (1966) 228–248. (p. 399)

[21] Hausdorff, F.: *Grundzüge der Mengenlehre*. Leipzig: Veit & Comp., 1914 (1st edn). (Reprint of the original edn, New York: Chelsea, 1949. Annotated reprint in *Felix Hausdorff – Gesammelte Werke Band III*. Berlin: Springer, 2008). (p. 45)

[22] Hewitt, E. and K. R. Stromberg: *Real and Abstract Analysis*. New York: Springer, 1975. (p. 12, 20, 45)

[23] Hunt, G. A.: *Martingales et Processus de Markov*. Paris: Dunod, 1966. (p. 254)

[24] Kaczmarz, S. and H. Steinhaus: *Theorie der Orthogonalreihen*. New York: Chelsea, 1951 (2nd corrected reprint). [First edn: Warsaw: PWN, 1935.] (p. 367)

[25] Korovkin, P. P.: *Linear Operators and Approximation Theory*. Delhi: Hindustan Publishing Corp., 1960. (p. 370)

[26] Krantz, S. G.: *A Panorama of Harmonic Analysis*. Washington (D.C.): Mathematical Association of America, 1999. (p. 311, 315)

[27] Lévy, P.: *Processus Stochastiques et Mouvement Brownien*. Paris: Gauthier-Villars, 1948. (p. 402)

[28] Lindenstrauss, J. and L. Tzafriri: *Classical Banach Spaces I, II*. Berlin: Springer, 1977–79. (p. 383, 384)

[29] Marcinkiewicz, J. and A. Zygmund: Sur les fonctions indépendantes. *Fundam. Math.* **29** (1937) 309–335. (Also in J. Marcinkiewicz *Collected papers*. Warsaw: PWN, 1964, 233–259.) (p. 271)

[30] Métivier, M.: *Semimartingales. A Course on Stochastic Processes*. Berlin: De Gruyter, 1982. (p. 303)

[31] Munroe, M. E.: *Introduction to Measure and Integration*. Reading (MA): Addison–Wesley, 1953. (p. 196)

[32] Sz.-Nagy, B.: *Introduction to Real Functions and Orthogonal Expansions*. New York: Oxford University Press, 1965. (p. 445)

[33] Olver, F. W. J., Lozier, D. W., Boisvert, R. F. and Clark, C. W.: *NIST Handbook of Mathematical Functions*. Cambridge: Cambridge University Press, 2010. (free online access: http://dlmf.nist.gov/) (p. 212, 367)

[34] Olevskiĭ, A. M.: *Fourier Series with Respect to General Orthogonal Systems*. Berlin: Springer, 1975. (p. 383)

[35] Oxtoby, J. C.: *Measure and Category*, New York: Springer, 1980 (2nd edn). (p. 45)

[36] Paley, R. E. A. C. and N. Wiener: *Fourier Transforms in the Complex Domain*. Providence (RI): American Mathematical Society, 1934. (p. 403)

[37] Pinsky, M. A.: *Introduction to Fourier Analysis and Wavelets*. Pacific Grove (CA): Brooks/Cole, 2002. (Reprinted: Providence (RI): American Mathematical Society, 2009.) (p. 388)

[38] Pratt, J. W.: On interchanging limits and integrals. *Ann. Math. Stat.* **31** (1960) 74–77. (Acknowledgement of priority: *Ann. Math. Stat.* **37** (1966) 1407.) (p. 108)

[39] Riemann, B.: Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe. *Nachr. Kgl. Ges. Wiss. Göttingen* **13** (1867) 227–271. (Also in Bernhard Riemann *Collected papers*, Berlin: Springer, 1990, 259–303.) (p. 433)
[40] Rogers, C. A.: *Hausdorff Measures*. Cambridge: Cambridge University Press, 1970. (p. 202)
[41] Rogers, L. C. G. and D. Williams: *Diffusions, Markov Processes and Martingales* (2 vols). Cambridge: Cambridge University Press, 2000 (2nd edn). (p. 384)
[42] Rudin, W.: *Principles of Mathematical Analysis*. New York: McGraw-Hill, 1976 (3rd edn). (p. 163, 444, 452)
[43] Rudin, W.: *Real and Complex Analysis*. New York: McGraw-Hill, 1987 (3rd edn). (p. 334, 336, 378)
[44] Sellke, T.: Yet another proof of the Radon–Nikodým theorem. *Am. Math. Monthly* **109** (2002) 74–76. (p. 226)
[45] Schilling, R. L. and L. Partzsch: *Brownian Motion. An Introduction to Stochastic Processes*. Berlin: De Gruyter, 2014 (2nd edn). (p. 402)
[46] Schipp, F., Wade, W. R. and P. Simon: *Walsh Series. An Introduction to Dyadic Harmonic Analysis*. Bristol: Adam Hilger, 1990. (p. 392)
[47] Souslin, M. Y.: Sur une définition des ensembles mesurables B sans nombres transfinis, *C. R. Acad. Sci. Paris* **164** (1917) 88–91. (p. 428)
[48] Srivastava, S. M.: *A Course on Borel Sets*. New York: Springer, 1998. (p. 428)
[49] Solovay, R. M.: A model of set theory in which every set of reals is Lebesgue measurable. *Ann. Math.* **92** (1970) 1–56. (p. 423)
[50] Stein, E. M.: *Singular Integrals and Differentiability Properties of Functions*. Princeton (NJ): Princeton University Press, 1970. (p. 313)
[51] Stoer, J. and R. Bulirsch: *Introduction to Numerical Analysis*. New York: Springer, 1980 (3rd edn). (p. 166)
[52] Strichartz, R. S.: *The Way of Analysis*. Sudbury (MA): Jones and Bartlett, 2000 (revised edn). (p. 439)
[53] Stromberg, K.: The Banach–Tarski paradox. *Am. Math. Monthly* **86** (1979) 151–161. (p. 45)
[54] Stroock, D. W.: *A Concise Introduction to the Theory of Integration*. Boston (MA): Birkhäuser, 1999 (3rd edn). (p. 182)
[55] Talvila, E.: Necessary and sufficient conditions for differentiating under the integral sign. *Am. Math. Monthly* **108** (2001) 544–548. (p. 151)
[56] Titkos, T.: A simple proof of the Lebesgue decomposition theorem. *Am. Math. Monthly* **122** (2015) 793–794. (p. 231)
[57] Wagon, S.: *The Banach–Tarski Paradox*. Cambridge: Cambridge University Press, 1985. (p. 45)
[58] Wheeden, R. L. and A. Zygmund: *Measure and Integral. An Introduction to Real Analysis*. New York: Marcel Dekker, 1977. (p. 378)
[59] Yosida, K.: *Functional Analysis*. Berlin: Springer, 1980 (6th edn). (p. 239, 322)
[60] Young, W. H.: On semi-integrals and oscillating successions of functions. *Proc. London Math. Soc.* **9** (1910/11) 286–324. (p.108)

Index

This should be used in conjunction with the bibliography and the list of symbols. Numbers following entries are page numbers which, if accompanied by (Pr *n.m*), refer to Problem *n.m* on that page; a number with a trailing 'n' indicates that a footnote is being referenced. Unless stated otherwise 'integral', 'integrability', etc. always refer to the (abstract) Lebesgue integral. Within the index we use 'L-…' and 'R-…' as a shorthand for '(abstract) Lebesgue-…' and 'Riemann-…'