

VI. Measure Theory for Euclidean Space, 334-394

DOI: [10.3792/euclid/9781429799997-6](https://doi.org/10.3792/euclid/9781429799997-6)

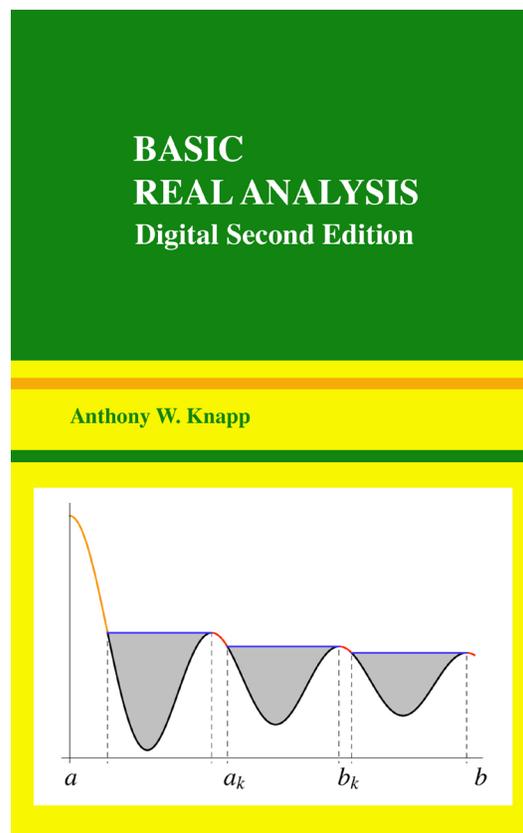
from

Basic Real Analysis
Digital Second Edition

Anthony W. Knapp

Full Book DOI: [10.3792/euclid/9781429799997](https://doi.org/10.3792/euclid/9781429799997)

ISBN: 978-1-4297-9999-7



Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, N.Y. 11733-1729, U.S.A.
Email to: aknapp@math.stonybrook.edu
Homepage: www.math.stonybrook.edu/~aknapp

Title: Basic Real Analysis, with an appendix “Elementary Complex Analysis”

Cover: An instance of the Rising Sun Lemma in Section VII.1.

Mathematics Subject Classification (2010): 28-01, 26-01, 42-01, 54-01, 34-01, 30-01, 32-01.

First Edition, ISBN-13 978-0-8176-3250-2

©2005 Anthony W. Knapp

Published by Birkhäuser Boston

Digital Second Edition, not to be sold, no ISBN

©2016 Anthony W. Knapp

Published by the author

All rights reserved. This file is a digital second edition of the above named book. The text, images, and other data contained in this file, which is in portable document format (PDF), are proprietary to the author, and the author retains all rights, including copyright, in them. The use in this file of trade names, trademarks, service marks, and similar items, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

All rights to print media for the first edition of this book have been licensed to Birkhäuser Boston, c/o Springer Science+Business Media Inc., 233 Spring Street, New York, NY 10013, USA, and this organization and its successor licensees may have certain rights concerning print media for the digital second edition. The author has retained all rights worldwide concerning digital media for both the first edition and the digital second edition.

The file is made available for limited noncommercial use for purposes of education, scholarship, and research, and for these purposes only, or for fair use as understood in the United States copyright law. Users may freely download this file for their own use and may store it, post it online, and transmit it digitally for purposes of education, scholarship, and research. They may not convert it from PDF to any other format (e.g., EPUB), they may not edit it, and they may not do reverse engineering with it. In transmitting the file to others or posting it online, users must charge no fee, nor may they include the file in any collection of files for which a fee is charged. Any exception to these rules requires written permission from the author.

Except as provided by fair use provisions of the United States copyright law, no extracts or quotations from this file may be used that do not consist of whole pages unless permission has been granted by the author (and by Birkhäuser Boston if appropriate).

The permission granted for use of the whole file and the prohibition against charging fees extend to any partial file that contains only whole pages from this file, except that the copyright notice on this page must be included in any partial file that does not consist exclusively of the front cover page. Such a partial file shall not be included in any derivative work unless permission has been granted by the author (and by Birkhäuser Boston if appropriate).

Inquiries concerning print copies of either edition should be directed to Springer Science+Business Media Inc.

CHAPTER VI

Measure Theory for Euclidean Space

Abstract. This chapter mines some of the powerful consequences of the basic measure theory in Chapter V.

Sections 1–3 establish properties of Lebesgue measure and other Borel measures on Euclidean space and on open subsets of Euclidean space. The main general property is the regularity of all such measures—that the measure of any Borel set can be approximated by the measure of compact sets from within and open sets from without. Lebesgue measure in all of Euclidean space has an additional property, translation invariance, which allows for the notion of the convolution of two functions. Convolution gives a kind of moving average of the translates of one function weighted by the other function. Convolution with the dilates of a fixed integrable function provides a handy kind of approximate identity.

Section 4 gives the final form of the comparison of the Riemann and Lebesgue integrals, a preliminary form having been given in Chapter III.

Section 5 gives the final form of the change-of-variables theorem for integration, starting from the preliminary form of the theorem in Chapter III and taking advantage of the ease with which limits can be handled by the Lebesgue integral. Sard’s Theorem allows one to disregard sets of lower dimension in establishing such changes of variables, thereby giving results in their expected form rather than in a form dictated by technicalities.

Section 6 concerns the Hardy–Littlewood Maximal Theorem in N dimensions. In dimension 1, this theorem implies that the derivative of a 1-dimensional Lebesgue integral with respect to Lebesgue measure recovers the integrand almost everywhere. The theorem in the general case implies that certain averages of a function over small sets about a point tend to the function almost everywhere. But the theorem can be regarded as saying also that a particular approximate identity formed by dilations applies to problems of almost-everywhere convergence, as well as to problems of norm convergence and uniform convergence. A corollary of the theorem is that many approximate identities formed by dilations yield almost-everywhere convergence theorems.

Section 7 redevelops the beginnings of the subject of Fourier series using the Lebesgue integral, the theory having been developed with the Riemann integral in Section I.10. With the Lebesgue integral and its accompanying tools, Fourier series are meaningful for more functions than before, Dini’s test applies even to a wider class of Riemann integrable functions than before, and Fejér’s Theorem and Parseval’s Theorem become easier and more general than before. A completely new result with the Lebesgue integral is the Riesz–Fischer Theorem, which characterizes the trigonometric series that are Fourier series of square-integrable functions.

Sections 8–10 deal with Stieltjes measures, which are Borel measures on the line, and their application to Fourier series. Such measures are characterized in terms of a class of monotone functions on the line, and they lead to a handy generalization of the integration-by-parts formula. This formula allows one to bound the size of the Fourier coefficients of functions of bounded variation, which are differences of monotone functions. In combination with earlier results, this bound yields

the Dirichlet–Jordan Theorem, which says that the Fourier series of a function of bounded variation converges pointwise everywhere, the convergence being uniform on any compact set on which the function is continuous. Section 10 is a short section on computation of integrals.

1. Lebesgue Measure and Other Borel Measures

Lebesgue measure on \mathbb{R}^1 was constructed in Section V.1 on the ring of “elementary” sets—the finite disjoint unions of bounded intervals—and extended from there to the σ -algebra of Borel sets by the Extension Theorem (Theorem 5.5), which was proved in Section V.5. Fubini’s Theorem (Theorem 5.47) would have allowed us to build Lebesgue measure in \mathbb{R}^N as an iterated product of 1-dimensional Lebesgue measure, but we postponed the construction in \mathbb{R}^N until the present chapter in order to show that it can be carried out in a fashion independent of how we group 1-dimensional factors.

The Borel sets of \mathbb{R}^1 are, by definition, the sets in the smallest σ -algebra containing the elementary sets, and we saw readily that every set that is open or compact is a Borel set. We write \mathcal{B}_1 for this σ -algebra. In fact, \mathcal{B}_1 may be described as the smallest σ -algebra containing the open sets of \mathbb{R}^1 or as the smallest σ -algebra containing the compact sets. The reason that the open sets generate \mathcal{B}_1 is that every open interval is an open set, and every interval is a countable intersection of open intervals. Similarly the compact sets generate \mathcal{B}_1 because every closed bounded interval is a compact set, and every interval is the countable union of closed bounded intervals.

Now let us turn our attention to \mathbb{R}^N . We have already used the word “rectangle” in two different senses in connection with integration—in Chapter III to mean an N -fold product along coordinate directions of open or closed bounded intervals, and in Chapter V to mean a product of measurable sets. For clarity let us refer to any product of bounded intervals as a **geometric rectangle** and to any product of measurable sets as an **abstract rectangle** or an abstract rectangle in the sense of Fubini’s Theorem. In \mathbb{R}^N , every geometric rectangle under our definition is an abstract rectangle, but not conversely.

Define the **Borel sets** of \mathbb{R}^N to be the members of the smallest σ -algebra \mathcal{B}_N containing all compact sets in \mathbb{R}^N . It is equivalent to let \mathcal{B}_N be the smallest σ -algebra containing all open sets. In fact, every open geometric rectangle is the countable union of compact geometric rectangles, and every open set in turn is the countable union of open geometric rectangles; thus the open sets are in the smallest σ -algebra containing the compact sets. In the reverse direction every closed set is the complement of an open set, and every compact set is closed; thus the compact sets are in the smallest σ -algebra containing the open sets.

Functions on \mathbb{R}^N measurable with respect to \mathcal{B}_N are called **Borel measurable functions** or **Borel functions**. Any continuous real-valued function f on \mathbb{R}^N

is Borel measurable because the inverse image $f^{-1}((c, +\infty])$ of the open set $(c, +\infty]$ has to be open and therefore has to be a Borel set.

Proposition 6.1. If m and n are integers ≥ 1 , then $\mathcal{B}_m \times \mathcal{B}_n = \mathcal{B}_{m+n}$ within the product set $\mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^{m+n}$.

PROOF. If U is open in \mathbb{R}^m and V is open in \mathbb{R}^n , then $U \times V$ is open in \mathbb{R}^{m+n} , and it follows that $\mathcal{B}_m \times \mathcal{B}_n \subseteq \mathcal{B}_{m+n}$. For the reverse inclusion, let W be open in \mathbb{R}^{m+n} . Then W is the countable union of open geometric rectangles, and each of these is of the form $U \times V$ with U open in \mathbb{R}^m and V open in \mathbb{R}^n . Since each such $U \times V$ is in $\mathcal{B}_m \times \mathcal{B}_n$, so is W . Thus we obtain the reverse inclusion $\mathcal{B}_{m+n} \subseteq \mathcal{B}_m \times \mathcal{B}_n$. \square

Lebesgue measure on \mathbb{R}^N will, at least initially, be a measure defined on the σ -algebra \mathcal{B}_N . Proposition 6.1 tells us that the σ -algebra on which the measure is to be defined is independent of the grouping of variables used in Fubini's Theorem. It will be quite believable that different constructions of Lebesgue measure by using different iterated product decompositions of \mathbb{R}^N , such as $(\mathbb{R}^1 \times \mathbb{R}^1) \times \mathbb{R}^1$ and $\mathbb{R}^1 \times (\mathbb{R}^1 \times \mathbb{R}^1)$, will lead to the same measure, but we shall give two abstract characterizations of the result that will ensure uniqueness without any act of faith. These characterizations will take some moments to establish, but we shall obtain useful additional results along the way. The procedure will be to state the constructions of the measure via Fubini's Theorem, then to consider a wider class of measures on \mathcal{B}_N known as the "Borel measures," and finally to establish the two characterizations of Lebesgue measure among all Borel measures on \mathbb{R}^N .

It is customary to write dx in place of $dm(x)$ for Lebesgue measure on \mathbb{R}^1 , and we shall do so except when there is some special need for the symbol m . Then the notation for the measure normally becomes an expression like dx or dy instead of m . To construct Lebesgue measure dx on \mathbb{R}^N , we can proceed inductively, adding one variable at a time. Fubini's Theorem allows us to construct the product of Lebesgue measure on \mathbb{R}^{N-1} and Lebesgue measure on \mathbb{R}^1 , and Proposition 6.1 shows that the result is defined on the Borel sets of \mathbb{R}^N . Let us take this particular construction as an inductive definition of **Lebesgue measure** on \mathbb{R}^N . It is apparent from the construction that the measure of a geometric rectangle is the product of the lengths of the sides.

Alternatively, we could construct Lebesgue measure on \mathbb{R}^N inductively by grouping \mathbb{R}^N as some other $\mathbb{R}^m \times \mathbb{R}^{N-m}$ and using the product measure from versions of the Lebesgue measures on \mathbb{R}^m and \mathbb{R}^{N-m} . Again the result has the property that the measure of a geometric rectangle is the product of the lengths of the sides. It is believable that this condition determines completely the measure on \mathbb{R}^N , and we shall give a proof of this uniqueness shortly.

A **Borel measure** on \mathbb{R}^N is a measure on the σ -algebra \mathcal{B}_N of Borel sets of \mathbb{R}^N that is finite on every compact set. A key property of Borel measures on \mathbb{R}^N is their regularity as expressed in Theorem 6.2 below. The theorem makes use of two simple properties of \mathbb{R}^N :

- (i) there exists a sequence $\{F_n\}_{n=1}^\infty$ of compact sets with union the whole space such that $F_n \subseteq F_{n+1}^\circ$ for all n ,
- (ii) for any compact set K , there exists a decreasing sequence of open sets U_n with compact closure such that $\bigcap_{n=1}^\infty U_n = K$.

For (i), we can take F_n to be the closed ball of radius n centered at the origin. For (ii), we can take $U_n = \{x \mid D(x, K) < 1/n\}$ if $K \neq \emptyset$, and we can take all $U_n = \emptyset$ if $K = \emptyset$.

Theorem 6.2. Every Borel measure μ on \mathbb{R}^N is **regular** in the sense that the value of μ on any Borel set E is given by

$$\mu(E) = \sup_{\substack{K \subseteq E, \\ K \text{ compact}}} \mu(K) = \inf_{\substack{U \supseteq E, \\ U \text{ open}}} \mu(U).$$

REMARK. This conclusion is new for us even for \mathbb{R}^1 . Although regularity of 1-dimensional Lebesgue measure was introduced before Proposition 5.4, it was established only for the elementary sets at that time.

PROOF. We shall begin by showing for each Borel set E and for any $\epsilon > 0$ that

$$\text{there exist closed } C \text{ and open } U \text{ such that} \quad (*) \\ C \subseteq E \subseteq U \text{ and } \mu(U - C) < \epsilon.$$

Let \mathcal{A} be the set of Borel sets E for which $(*)$ holds for all $\epsilon > 0$.

If E is compact, then we can take $C = E$ and $U = U_n$ as in (ii) for a suitable n in order to prove $(*)$; Corollary 5.3 gives us $\lim_n \mu(U_n - C) = 0$, since the compact closure of U_n forces $\mu(U_n)$ to be finite. Therefore \mathcal{A} contains all compact sets.

To see that \mathcal{A} is closed under complements, suppose E is in \mathcal{A} . Let $\epsilon > 0$ be given and choose, by $(*)$ for E , a closed set C and an open set U such that $C \subseteq E \subseteq U$ and $\mu(U - C) < \epsilon$. Taking complements, we have $U^c \subseteq E^c \subseteq C^c$ and $\mu(C^c - U^c) = \mu(U - C) < \epsilon$. Thus E^c is in \mathcal{A} .

Let us see that \mathcal{A} is closed under finite unions. Suppose that E_1 and E_2 are in \mathcal{A} . Let $\epsilon > 0$ be given and choose, by $(*)$ for E_1 and E_2 , two closed sets C_1 and C_2 and two open sets U_1 and U_2 such that $C_1 \subseteq E_1 \subseteq U_1$, $\mu(U_1 - C_1) < \epsilon$, $C_2 \subseteq E_2 \subseteq U_2$, and $\mu(U_2 - C_2) < \epsilon$. Then $C_1 \cup C_2 \subseteq E_1 \cup E_2 \subseteq U_1 \cup U_2$ and $\mu((U_1 \cup U_2) - (C_1 \cup C_2)) \leq \mu(U_1 - C_1) + \mu(U_2 - C_2) < 2\epsilon$. Since ϵ is

arbitrary, $E_1 \cup E_2$ is in \mathcal{A} . Hence \mathcal{A} is closed under finite unions, and \mathcal{A} is an algebra of sets.

The proof that \mathcal{A} is closed under countable unions takes two steps. For the first step we let a sequence of sets E_n in \mathcal{A} be given with union E , and first assume that all E_n lie in one of the sets F_M in (i) above. Let $\epsilon > 0$ be given and choose, by (*) for each E_n , closed sets C_n and open sets U_n such that $C_n \subseteq E_n \subseteq U_n$ and $\mu(U_n - C_n) < \epsilon/2^n$. Possibly by intersecting U_n with F_{M+1}^o , we may assume that all U_n lie in the compact set F_{M+1} . Set $U = \bigcup_{n=1}^{\infty} U_n$ and $C = \bigcup_{n=1}^{\infty} C_n$. Then $C \subseteq E \subseteq U$ with U open but C not necessarily closed. Nevertheless, we have $U - C \subseteq \bigcup_{n=1}^{\infty} (U_n - C_n)$, and Proposition 5.1g gives $\mu(U - C) \leq \sum_{n=1}^{\infty} \mu(U_n - C_n) < \epsilon$. The sets $S_m = U - \bigcup_{n=1}^m C_n$ form a decreasing sequence within F_{M+1} with intersection $U - C$. Since $\mu(F_{M+1})$ is finite, Corollary 5.3 shows that $\mu(S_m)$ decreases to $\mu(U - C)$, which is $< \epsilon$. Thus there is some $m = m_0$ with $\mu(S_{m_0}) < \epsilon$. The set $C' = \bigcup_{n=1}^{m_0} C_n$ is closed, and we have $C' \subseteq E \subseteq U$ and $\mu(U - C') = \mu(S_{m_0}) < \epsilon$. Therefore E is in \mathcal{A} .

For the second step we let the sets E_n be general members of \mathcal{A} . Since \mathcal{A} is an algebra, $E_n \cap (F_{m+1} - F_m)$ is in \mathcal{A} for every n and m . Applying the previous step, we see that $E'_m = E \cap (F_{m+1} - F_m)$ is in \mathcal{A} for every m . The sets E'_m have union E , and E'_m is contained in $F_{m+1} - F_m$. Changing notation, we may assume that the given sets E_n all have $E_n \subseteq F_{n+1} - F_n$. If $\epsilon > 0$ is given, construct U_n open and C_n closed as in the previous paragraph except that U_n is not constrained to lie in a particular F_M . Again let $U = \bigcup_{n=1}^{\infty} U_n$ and $C = \bigcup_{n=1}^{\infty} C_n$, so that $C \subseteq E \subseteq U$ and $\mu(U - C) < \epsilon$. The set U is open, and this time we can prove that the set C is closed. In fact, let $\{x_k\}$ be a sequence in C convergent to some limit point x_0 . The point x_0 is in some F_M since the sets F_M have union the whole space. Since $F_M \subseteq F_{M+1}^o$ and F_{M+1}^o is open, the sequence is eventually in F_{M+1}^o . The inclusion $C_n \subseteq E_n \subseteq F_{n+1} - F_n$ shows that $C_n \cap F_{M+1} = \emptyset$ for $n \geq M+1$. Thus no term of the sequence after some point lies in C_{M+1}, C_{M+2}, \dots , i.e., all the terms of the sequence after some point lie in $\bigcup_{n=1}^M C_n$. This is a closed set, and the limit x_0 must lie in it. Therefore x_0 lies in C , and C is closed. This proves that E is in \mathcal{A} . Hence \mathcal{A} is a σ -algebra and must contain all Borel sets.

From (*) for all Borel sets, it follows that every Borel set E satisfies

$$\mu(E) = \sup_{\substack{C \subseteq E, \\ C \text{ closed}}} \mu(C) = \inf_{\substack{U \supseteq E, \\ U \text{ open}}} \mu(U). \quad (**)$$

Proposition 5.2 shows that the sets F_n of (i) have the property that $\mu(C) = \sup \mu(C \cap F_n)$ for every Borel set C . When C is closed, the sets $C \cap F_n$ are compact, and thus (**) implies the equality asserted in the statement of the theorem. This completes the proof. \square

Recall from Section III.10 that the **support** of a scalar-valued function on a metric space is the closure of the set where it is nonzero. Let $C_{\text{com}}(\mathbb{R}^N)$ be the

space of continuous scalar-valued functions on \mathbb{R}^N of compact support. If there is no special mention of the scalars, the scalars may be either real or complex.

If K is a compact set and the open sets U_n are as in (ii) before Theorem 6.2, Proposition 2.30e gives us continuous functions $f_n : \mathbb{R}^N \rightarrow [0, 1]$ such that f_n is 1 on K and is 0 on U_n^c . The support of the function f_n is then contained in U_n^{cl} , which is compact. By replacing the functions f_n by $g_n = \min\{f_1, \dots, f_n\}$, we may assume that they are pointwise decreasing. Consequently

- (iii) there exists a decreasing sequence of real-valued members of $C_{\text{com}}(\mathbb{R}^N)$ with pointwise limit the indicator function of K .

Corollary 6.3. If μ and ν are Borel measures on \mathbb{R}^N such that $\int_{\mathbb{R}^N} f d\mu = \int_{\mathbb{R}^N} f d\nu$ for all continuous functions on \mathbb{R}^N of compact support, then $\mu = \nu$.

PROOF. Let K be a compact subset of \mathbb{R}^N , and use (iii) to choose a decreasing sequence $\{f_n\}$ of real-valued members of $C_{\text{com}}(\mathbb{R}^N)$ with pointwise limit the indicator function I_K . Since f_1 is integrable, dominated convergence allows us to deduce $\int_{\mathbb{R}^N} I_K d\mu = \int_{\mathbb{R}^N} I_K d\nu$ from the equality $\int_{\mathbb{R}^N} f_n d\mu = \int_{\mathbb{R}^N} f_n d\nu$ for all n . Thus $\mu(K) = \nu(K)$ for every compact set K . Applying Theorem 6.2, we obtain $\mu(E) = \nu(E)$ for every Borel set E . \square

Corollary 6.4. Let $p = 1$ or $p = 2$. If μ is a Borel measure on \mathbb{R}^N , then

- (a) $C_{\text{com}}(\mathbb{R}^N)$ is dense in $L^p(\mathbb{R}^N, \mu)$,
- (b) the smallest closed subspace of $L^p(\mathbb{R}^N, \mu)$ containing all indicator functions of compact sets in \mathbb{R}^N is $L^p(\mathbb{R}^N, \mu)$ itself.

REMARK. The scalars are assumed to be the same for $C_{\text{com}}(\mathbb{R}^N)$ as for $L^1(\mathbb{R}^N, \mu)$ and $L^2(\mathbb{R}^N, \mu)$; the corollary is valid both for real scalars and for complex scalars.

PROOF. If E is a Borel set of finite μ measure and if ϵ is given, Theorem 6.2 allows us to choose a compact set K with $K \subseteq E$ and $\mu(E - K) < \epsilon$. Then $\int_{\mathbb{R}^N} |I_E - I_K|^p d\mu = \mu(E - K) < \epsilon$, and consequently the closure in $L^p(\mathbb{R}^N, \mu)$ of the set of all indicator functions of compact sets contains all indicator functions of Borel sets of finite μ measure. Proposition 5.56 shows consequently that the smallest closed subspace of $L^p(\mathbb{R}^N, \mu)$ containing all indicator functions of compact sets is $L^p(\mathbb{R}^N, \mu)$ itself. This proves (b).

For (a), let K be compact, and use (iii) to choose a decreasing sequence $\{f_n\}$ of real-valued members of $C_{\text{com}}(\mathbb{R}^N)$ with pointwise limit I_K . Since f_1^p is integrable, dominated convergence yields $\lim_n \int_{\mathbb{R}^N} |f_n - I_K|^p d\mu = 0$. Hence the closure of $C_{\text{com}}(\mathbb{R}^N)$ in $L^p(\mathbb{R}^N, \mu)$ contains all indicator functions of compact sets. By Propositions 5.55c and 5.55d this closure contains the smallest closed subspace of $L^p(\mathbb{R}^N, \mu)$ containing all indicator functions of compact sets. By (b), the latter subspace is $L^p(\mathbb{R}^N, \mu)$ itself. This proves (a). \square

Fix an integer $n \geq 0$, and let (a_1, \dots, a_N) be an N -tuple of integers. The **diadic cube** $Q_n(a_1, \dots, a_N)$ in \mathbb{R}^N of side 2^{-n} is defined to be the geometric rectangle

$$Q_n(a_1, \dots, a_N) = \{(x_1, \dots, x_N) \mid 2^{-n}a_j < x_j \leq 2^{-n}(a_j + 1) \text{ for } 1 \leq j \leq N\}.$$

Let \mathcal{Q}_n be the set of all diadic cubes of side 2^{-n} . The members of \mathcal{Q}_n are disjoint and have union \mathbb{R}^N . Thus we can associate uniquely to each x in \mathbb{R}^N a sequence $\{Q_n\}$ of diadic cubes such that x is in Q_n and Q_n is in \mathcal{Q}_n . Since for each n , the members of \mathcal{Q}_{n+1} are obtained by subdividing each member of \mathcal{Q}_n into 2^N disjoint smaller diadic cubes, the diadic cubes Q_n associated to x must have the property that $Q_n \supseteq Q_{n+1}$ for all $n \geq 0$.

Lemma 6.5. Any open set in \mathbb{R}^N is the countable disjoint union of diadic cubes.

PROOF. Let an open set U be given. We may assume that $U \neq \mathbb{R}^N$, so that $U^c \neq \emptyset$. We describe which diadic cubes to include in a collection \mathcal{A} so that \mathcal{A} has the required properties. If x is in U , then $D(x, U^c) = d$ is positive since U^c is closed and nonempty. Let $\{Q_n\}$ be the sequence of diadic cubes associated to x . The distance between any two points of Q_n is $\leq 2^{-n}\sqrt{N}$, and this is $< d$ if n is sufficiently large. Hence Q_n is contained in U for n sufficiently large. The cube in \mathcal{A} that contains x is to be the Q_n with n as small as possible so that $Q_n \subseteq U$.

The construction has been arranged so that the union of the diadic cubes in \mathcal{A} is exactly U . Suppose that Q and Q' are members of \mathcal{A} obtained from respective points x and x' in U . If $Q \cap Q' \neq \emptyset$, let x'' be in the intersection. Then Q and Q' are two of the diadic cubes in the sequence associated to x'' , and one has to contain the other. Without loss of generality, suppose that $Q \supseteq Q'$. Then x' lies in Q as well as Q' , and we should have selected Q for x' rather than Q' if $Q \neq Q'$. We conclude that $Q = Q'$, and thus the members of \mathcal{A} are disjoint. Each collection \mathcal{Q}_n is countable, and therefore the collection \mathcal{A} is countable. \square

Proposition 6.6. Any Borel measure on \mathbb{R}^N is determined by its values on all the diadic cubes.

REMARK. We shall apply this result in the present section in connection with Lebesgue measure on \mathbb{R}^N and in Section 8 in connection with general Borel measures on \mathbb{R}^1 .

PROOF. The values on the diadic cubes determine the values on all open sets by Lemma 6.5, and the values on all open sets determine the values on all Borel sets by Theorem 6.2. \square

Corollary 6.7. There exists a unique Borel measure on \mathbb{R}^N for which the measure of each geometric rectangle is the product of the lengths of the sides. The measure is the N -fold product of 1-dimensional Lebesgue measure.

REMARKS. The uniqueness is immediate from Proposition 6.6. The first version of Lebesgue measure that we constructed has the property stated in the corollary and therefore proves existence. All the other versions of **Lebesgue measure** we constructed have the same property, and so all such versions are equal. The corollary therefore allows us to use Fubini's Theorem for any decomposition $\mathbb{R}^N = \mathbb{R}^m \times \mathbb{R}^n$ with $m + n = N$. As in the 1-dimensional case, we shall often write dx for Lebesgue measure.

Corollary 6.7 gives one characterization of Lebesgue measure. We shall use Proposition 6.6 to give a second characterization, which will be in terms of translation invariance.

Proposition 6.8. Under a Borel function $F : \mathbb{R}^N \rightarrow \mathbb{R}^{N'}$, $F^{-1}(E)$ is in \mathcal{B}_N whenever E is in $\mathcal{B}_{N'}$. In particular, this conclusion is valid if F is continuous.

PROOF. The set of E 's for which $F^{-1}(E)$ is in \mathcal{B}_N is a σ -algebra, and the result will follow if this set of E 's contains the open geometric rectangles of $\mathbb{R}^{N'}$. If F_j denotes the j^{th} component of F , then $F_j : \mathbb{R}^N \rightarrow \mathbb{R}^1$ is Borel measurable and Proposition 5.6c shows that $F_j^{-1}(U_j)$ is a Borel set in \mathbb{R}^N if U_j is open in \mathbb{R}^1 . Then $F^{-1}(U_1 \times \cdots \times U_{N'}) = \bigcap_{j=1}^{N'} F_j^{-1}(U_j)$ is a Borel set in \mathbb{R}^N . \square

Corollary 6.9. Any homeomorphism of \mathbb{R}^N carries \mathcal{B}_N to \mathcal{B}_N .

Corollary 6.9 is a special case of Proposition 6.8. The particular homeomorphisms of interest at the moment are **translations** and **dilations**. Translation by x_0 is the homeomorphism $\tau_{x_0}(x) = x + x_0$. Its operation on a set E is given by $\tau_{x_0}(E) = \{\tau_{x_0}(x) \mid x \in E\} = \{x + x_0 \mid x \in E\} = E + x_0$, and its operation on a function f on \mathbb{R}^N is given by $\tau_{x_0}(f)(x) = f(\tau_{x_0}^{-1}(x)) = f(x - x_0)$. Its operation on an indicator function I_E is $\tau_{x_0}(I_E)(x) = I_E(x - x_0) = I_{E+x_0}(x) = I_{\tau_{x_0}(E)}(x)$. Because of Corollary 6.9, translations operate on measures, the formula being $\tau_{x_0}(\mu)(E) = \mu(\tau_{x_0}^{-1}(E))$; since homeomorphisms carry compact sets to compact sets, the right side is a Borel measure if μ is a Borel measure. The actions of τ_{x_0} on functions and measures are related by integration. If $f \geq 0$ is a Borel function, then so is $\tau_{x_0}(f)$, and $\int_{\mathbb{R}^N} f d(\tau_{x_0}\mu) = \int_{\mathbb{R}^N} \tau_{x_0}^{-1}(f) d\mu$; this formula is verified by checking it for indicator functions and then passing to simple functions ≥ 0 by linearity and to Borel functions $f \geq 0$ by monotone convergence.

Dilation δ_c by a nonzero real c is given on members of \mathbb{R}^N by $\delta_c(x) = cx$, and the operations on sets, functions, indicator functions, and measures are analogous

to the corresponding operations for translations. Although dilations will play a recurring role in this book, the notation δ_c will be used only in the present section.

Theorem 6.10. Lebesgue measure m on \mathbb{R}^N is translation invariant in the sense that $\tau_{x_0}(m) = m$ for every x_0 in \mathbb{R}^N . In fact, Lebesgue measure is the unique translation-invariant Borel measure on \mathbb{R}^N that assigns measure 1 to the diadic cube $Q_0(0, \dots, 0)$. The effect of dilations on Lebesgue measure is that $\delta_c(m) = |c|^{-N}m$, i.e., $\int_{\mathbb{R}^N} f(cx) dx = |c|^{-N} \int_{\mathbb{R}^N} f(x) dx$ for every nonnegative Borel function f .

REMARKS. From one point of view, translation and dilation are examples of bounded linear operators on each $L^p(\mathbb{R}^N, dx)$, with translation preserving norms and with dilation multiplying norms by a constant depending on p and the particular dilation. From another point of view, translation and dilation are especially simple examples of changes of variables. Operationally the theorem allows us to write $dy = dx$ when $y = \tau_{x_0}(x)$ and $dz = |c|^N dx$ when $z = cx$. These effects of translations and dilations on integration with respect to Lebesgue measure are special cases of the general change-of-variables formula to be proved in Section 5.

PROOF. For any x_0 in \mathbb{R}^N , m and $\tau_{x_0}(m)$ assign the product of the lengths of the sides as measure to any diadic cube. From Proposition 6.6 we conclude that $m = \tau_{x_0}(m)$. The assertion about the effect of dilations on Lebesgue measure is proved similarly.

We still have to prove the uniqueness. Let μ be a translation-invariant Borel measure. The members of \mathcal{Q}_n are translates of one another and hence have equal μ measure. The members of \mathcal{Q}_{n+1} are obtained by partitioning each member of \mathcal{Q}_n into 2^N members of \mathcal{Q}_{n+1} that are translates of one another. Thus the μ measure of any member of \mathcal{Q}_{n+1} is 2^{-N} times the μ measure of any member of \mathcal{Q}_n . Consequently the μ measure of any diadic cube is completely determined by the value of μ on $Q_0(0, \dots, 0)$, which is a member of \mathcal{Q}_0 . The uniqueness then follows by another application of Proposition 6.6. \square

For a continuous function on a closed bounded interval, it was shown at the end of Section V.3 that the Riemann integral equals the Lebesgue integral. The next proposition gives an N -dimensional analog. A general comparison of the Riemann and Lebesgue integrals will be given in Section 4.

Proposition 6.11. For a continuous function on a compact geometric rectangle, the Riemann integral equals the Lebesgue integral.

PROOF. The two are equal in the 1-dimensional case, and the N -dimensional cases of each may be computed by iterated 1-dimensional integrals — as a result of

Corollary 3.33 in the case of the Riemann integral and as a result of the definition of Lebesgue measure as a product and the use of Fubini's Theorem (Theorem 5.47) in the case of the Lebesgue integral. \square

So far, we have worked in this section only with Lebesgue measure on the Borel sets. The **Lebesgue measurable sets** are those sets that occur when Lebesgue measure is completed. The Lebesgue measurable sets of measure 0 are of particular interest. In Section III.8 we defined an ostensibly different notion of measure 0 by saying that a set in \mathbb{R}^N is of measure 0 if for any $\epsilon > 0$, it can be covered by a countable set of open geometric rectangles of total volume less than ϵ , and Theorem 3.29 characterized the Riemann integrable functions on a compact geometric rectangle as those functions whose discontinuities form a set of measure 0 in this sense. Later, Proposition 5.39 showed for \mathbb{R}^1 that a set has measure 0 in this sense if and only if it is Lebesgue measurable of Lebesgue measure 0. This equivalence extends to \mathbb{R}^N , as the next proposition shows.

Proposition 6.12. In \mathbb{R}^N , the Lebesgue measurable sets of measure 0 are exactly the subsets E of \mathbb{R}^N with the following property: for any $\epsilon > 0$, the set E can be covered by countably many geometric rectangles of total volume less than ϵ .

PROOF. Let m be Lebesgue measure on \mathbb{R}^N . If E has the stated property, let E_n be the union of the given countable collection of geometric rectangles of total volume $< 1/n$ used to cover E . Proposition 5.1g shows that $m(E_n) < 1/n$, and hence the Borel set $E' = \bigcap_k E_k$ has $m(E') < 1/n$ for every n . Therefore $m(E') = 0$. Since $E \subseteq E'$, E is Lebesgue measurable and has Lebesgue measure 0.

Conversely if E is Lebesgue measurable of Lebesgue measure 0 and if $\epsilon > 0$ is given, we are to find a union of open geometric rectangles containing E and having total volume $< \epsilon$. Find a set E' in \mathcal{B}_N with $E \subseteq E'$ and $m(E') = 0$. It is enough to handle E' . Writing \mathbb{R}^N as the union of compact geometric cubes C_n of side $2n$ centered at the origin and covering $E' \cap C_n$ up to $\epsilon/2^n$, we see that we may assume that E' is bounded, being contained in some cube C_n .

Within $\mathbb{R}^1 \cap [-n, n]$, we know that the set of finite unions of intervals is an algebra $\mathcal{A}_1^{(n)}$ of sets such that $\mathcal{B}_1^{(n)} = \mathcal{B}_1 \cap [-n, n]$ is the smallest σ -algebra containing $\mathcal{A}_1^{(n)}$. Applying Proposition 5.40 inductively, we see that the set of finite disjoint unions of N -fold products of members of $\mathcal{A}_1^{(n)}$ is an algebra $\mathcal{A}_N^{(n)}$, and then Proposition 6.1 shows that the smallest σ -algebra containing $\mathcal{A}_N^{(n)}$ is $\mathcal{B}_N^{(n)} = \mathcal{B}_N \cap C_n$. Proposition 5.38 shows that the measure m on $\mathcal{B}_N^{(n)}$ is given by m^* , where $m^*(A)$ is the infimum of countable unions of members of $\mathcal{A}_N^{(n)}$ that cover A . Consequently the subset E' of C_n can be covered by countably many

geometric rectangles of total volume $< \epsilon$. Doubling these rectangles about their centers and discarding their edges, we obtain a covering of E' by open rectangles of total volume $< 2^N \epsilon$, and we have the required covering. \square

Borel measurable sets have two distinct advantages over Lebesgue measurable sets. One advantage is that Borel measurable sets are independent of the particular Borel measure in question, whereas the sets in the completion of a σ -algebra relative to a Borel measure very much depend on the particular measure. The other advantage is that Fubini's Theorem applies in a tidy fashion to Borel measurable functions as a consequence of the identity $\mathcal{B}_m \times \mathcal{B}_n = \mathcal{B}_{m+n}$ given in Proposition 6.1. By contrast, there are Lebesgue measurable sets for \mathbb{R}^N that are not in the product of the σ -algebras of Lebesgue measurable sets from \mathbb{R}^m and \mathbb{R}^{N-m} . For example, take a set E in \mathbb{R}^1 that is not Lebesgue measurable; such a set is produced in Problem 1 at the end of the present chapter. Then $E \times \{0\}$ in \mathbb{R}^2 is a subset of the Borel set $\mathbb{R}^1 \times \{0\}$, and hence it is Lebesgue measurable of measure 0. However, $E \times \{0\}$ is not in the product σ -algebra, because a section of a function measurable with respect to the product has to be measurable with respect to the appropriate factor (Lemma 5.46).

On the other hand, Lebesgue measurable functions are sometimes unavoidable. An example occurs with Riemann integrability: In view of Proposition 6.12, Theorem 3.29 says that the Riemann integrable functions on a compact geometric rectangle are exactly the functions whose discontinuities form a Lebesgue measurable set of Lebesgue measure 0, and Problems 31–33 at the end of Chapter V produced such a function in the 1-dimensional case that is not a Borel measurable function.

The upshot is that a little care is needed when using Fubini's Theorem and Lebesgue measurable sets at the same time, and there are times when one wants to do so. The situation is a little messy but not intractable. Problem 12 at the end of Chapter V showed that a Lebesgue measurable function can be adjusted on a set of Lebesgue measure 0 so as to become Borel measurable. Using this fact, one can write down a form of Fubini's Theorem for Lebesgue measurable functions that is usable even if inelegant.

2. Convolution

Convolution is an important operation available for functions on \mathbb{R}^N . On a formal level, the **convolution** $f * g$ of two functions f and g is

$$(f * g)(x) = \int_{\mathbb{R}^N} f(x - y)g(y) dy.$$

One place convolution arises is as a limit of a linear combination of translates: We shall see in Proposition 6.13 that the convolution at x may be written also as $\int_{\mathbb{R}^N} f(y)g(x-y)dy$. If f is fixed and if finite sets of translation operators τ_{y_i} and of weights $f(y_i)$ are given, then the value at x of the linear combination $\sum_i f(y_i)\tau_{y_i}$ applied to g and evaluated at x is $\sum_i f(y_i)g(x-y_i)$. Corollary 6.17 will show a sense in which we can think of $\int_{\mathbb{R}^N} f(y)g(x-y)dy$ as a limit of such expressions.

To make mathematical sense out of $f * g$, let us begin with the case that f and g are nonnegative Borel functions on \mathbb{R}^N . The assertion is that $f * g$ is meaningful as a Borel function ≥ 0 . In fact, $(x, y) \mapsto f(x-y)$ is the composition of the continuous function $F : \mathbb{R}^{2N} \rightarrow \mathbb{R}^N$ given by $F(x, y) = x - y$, followed by the Borel function $f : \mathbb{R}^N \rightarrow [0, +\infty]$. If U is open in $[0, +\infty]$, then $f^{-1}(U)$ is in \mathcal{B}_N , and Proposition 6.8 shows that $(f \circ F)^{-1}(U) = F^{-1}(f^{-1}(U))$ is in \mathcal{B}_{2N} . Then the product $(x, y) \mapsto f(x-y)g(y)$ is a Borel function, and Fubini's Theorem (Theorem 5.47) and Proposition 6.1 combine to show that $x \mapsto (f * g)(x)$ is a Borel function ≥ 0 .

Proposition 6.13. For nonnegative Borel functions on \mathbb{R}^N ,

- (a) $f * g = g * f$,
- (b) $f * (g * h) = (f * g) * h$.

PROOF. We use Theorem 6.10 for both parts and also Fubini's Theorem for (b). For (a), the changes of variables $y \mapsto y + x$ and then $y \mapsto -y$ give $\int_{\mathbb{R}^N} f(x-y)g(y)dy = \int_{\mathbb{R}^N} f(-y)g(y+x)dx = \int_{\mathbb{R}^N} f(y)g(x-y)dy$. For (b), the computation is

$$\begin{aligned} (f * (g * h))(x) &= \int_{\mathbb{R}^N} f(x-y)(g * h)(y)dy \\ &= \int_{\mathbb{R}^N} \left[\int_{\mathbb{R}^N} f(x-y)g(y-z)h(z)dz \right] dy \\ &= \int_{\mathbb{R}^N} \left[\int_{\mathbb{R}^N} f(x-y)g(y-z)h(z)dy \right] dz \\ &= \int_{\mathbb{R}^N} \left[\int_{\mathbb{R}^N} f(x-z-y)g(y)h(z)dy \right] dz \\ &= \int_{\mathbb{R}^N} (f * g)(x-z)h(z)dz = ((f * g) * h)(x), \end{aligned}$$

the change of variables $y \mapsto y + z$ being used for the fourth equality. \square

In order to have a well-defined expression for $f * g$ when f and g are not necessarily ≥ 0 , we need conditions under which the nonnegative case leads to something finite. The conditions we use ensure finiteness of $(|f| * |g|)(x)$ for almost every x . For real-valued f and g , we then define $f * g(x)$ by subtraction at the points where $(|f| * |g|)(x)$ is finite, and we define it to be 0 elsewhere. For complex-valued f and g , we define $(f * g)(x)$ as a linear combination of the

appropriate parts where $(|f| * |g|)(x)$ is finite, and we define it to be 0 elsewhere. When we proceed this way, the commutativity and associativity properties in Proposition 6.13 will be valid even though f and g are not necessarily ≥ 0 .

Proposition 6.14. For nonnegative Borel functions f and g on \mathbb{R}^N , convolution is finite almost everywhere in the following cases, and then the indicated inequalities of norms are satisfied:

- (a) for f in $L^1(\mathbb{R}^N)$ and g in $L^1(\mathbb{R}^N)$, and then $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$;
- (b) for f in $L^1(\mathbb{R}^N)$ and g in $L^2(\mathbb{R}^N)$, and then $\|f * g\|_2 \leq \|f\|_1 \|g\|_2$;
for f in $L^2(\mathbb{R}^N)$ and g in $L^1(\mathbb{R}^N)$, and then $\|f * g\|_2 \leq \|f\|_2 \|g\|_1$;
- (c) for f in $L^1(\mathbb{R}^N)$ and g in $L^\infty(\mathbb{R}^N)$, and then $\|f * g\|_\infty \leq \|f\|_1 \|g\|_\infty$;
for f in $L^\infty(\mathbb{R}^N)$ and g in $L^1(\mathbb{R}^N)$, and then $\|f * g\|_\infty \leq \|f\|_\infty \|g\|_1$;
- (d) for f in $L^2(\mathbb{R}^N)$ and g in $L^2(\mathbb{R}^N)$, and then $\|f * g\|_\infty \leq \|f\|_2 \|g\|_2$.

Consequently $f * g$ is defined in the above situations even if the scalar-valued functions f and g are not necessarily ≥ 0 , and the estimates on the norm of $f * g$ are still valid.

PROOF. For (a) and the first conclusions in (b) and (c), let p be 1, 2, or ∞ as appropriate. By Minkowski's inequality for integrals (Theorem 5.60),

$$\begin{aligned} \|f * g\|_p &= \left\| \int_{\mathbb{R}^N} f(y)g(x-y) dy \right\|_{p,x} \leq \int_{\mathbb{R}^N} \|f(y)g(x-y)\|_{p,x} dy \\ &= \int_{\mathbb{R}^N} |f(y)| \|g(x-y)\|_{p,x} dy = \int_{\mathbb{R}^N} |f(y)| \|g\|_p dy = \|f\|_1 \|g\|_p, \end{aligned}$$

the next-to-last equality following from the translation invariance of dx . The second conclusions in (b) and (c) require only notational changes.

For (d), we have

$$\begin{aligned} \sup_x |(f * g)(x)| &= \sup_x \left| \int_{\mathbb{R}^N} f(y)g(x-y) dy \right| \\ &\leq \sup_x \|f\|_2 \|g(x-y)\|_{2,y} = \|f\|_2 \|g\|_2, \end{aligned}$$

the inequality following from the Schwarz inequality and the last step following from translation invariance of dy and invariance under $y \mapsto -y$.

Going over these arguments, we see that we may use them even if f and g are not necessarily ≥ 0 . Then the last statement of the proposition follows. \square

Next let us relate the translation operators of Section 1 to convolution. The formula for the effect of a translation operator on a function is $\tau_t(f)(x) = f(x-t)$.

Proposition 6.15. Convolution commutes with translations in the sense that $\tau_t(f * g) = (\tau_t f) * g = f * \tau_t g$.

PROOF. It is enough to treat functions ≥ 0 . Then we have $\tau_t(f * g)(x) = (f * g)(x-t) = \int_{\mathbb{R}^N} f(x-t-y)g(y) dy$, which equals $\int_{\mathbb{R}^N} (\tau_t f)(x-y)g(y) dy = ((\tau_t f) * g)(x)$ on the one hand and, because of translation invariance of Lebesgue measure, equals $\int_{\mathbb{R}^N} f(x-y)g(y-t) dy = (f * \tau_t g)(x)$ on the other hand. \square

Proposition 6.16. If $p = 1$ or $p = 2$, then translation of a function is continuous in the translation parameter in $L^p(\mathbb{R}^N, dx)$. In other words, if f is in L^p relative to Lebesgue measure, then $\lim_{h \rightarrow 0} \|\tau_{t+h}f - \tau_t f\|_p = 0$ for all t .

REMARK. However, continuity fails on L^∞ . In this case, there is a substitute result, and we take that up in a moment.

PROOF. Let f be in L^p . By translation invariance of Lebesgue measure, $\|\tau_{t+h}f - \tau_t f\|_p = \|\tau_h f - f\|_p$. If g is in $C_{\text{com}}(\mathbb{R}^N)$, then $\|\tau_h g - g\|_p^p = \int_{\mathbb{R}^N} |g(x-h) - g(x)|^p dx$, and dominated convergence shows that this tends to 0 as h tends to 0. Let $\epsilon > 0$ and f be given. By Corollary 6.4a, $C_{\text{com}}(\mathbb{R}^N)$ is dense in $L^p(\mathbb{R}^N, dx)$, and thus we can choose g in $C_{\text{com}}(\mathbb{R}^N)$ with $\|f - g\|_p < \epsilon$. Then

$$\begin{aligned} \|\tau_h f - f\|_p &\leq \|\tau_h f - \tau_h g\|_p + \|\tau_h g - g\|_p + \|g - f\|_p \\ &= 2\|f - g\|_p + \|\tau_h g - g\|_p \leq 2\epsilon + \|\tau_h g - g\|_p. \end{aligned}$$

If h is close enough to 0, the term $\|\tau_h g - g\|_p$ is $< \epsilon$, and then $\|\tau_h f - f\|_p < 3\epsilon$. \square

Corollary 6.17. Let $p = 1$ or $p = 2$, and let g_1, \dots, g_r be finitely many functions in $L^p(\mathbb{R}^N)$. If a positive number ϵ and a function f in $L^1(\mathbb{R}^N)$ are given, then there exist finitely many members y_j of \mathbb{R}^N , $1 \leq j \leq n$, and constants c_j such that $\|f * g_k - \sum_{j=1}^n c_j \tau_{y_j} g_k\|_p < \epsilon$ for $1 \leq k \leq r$.

REMARK. In the case $r = 1$, the corollary says that any convolution $f * g$ can be approximated in L^p by a linear combination of translates of g . The result will be used in Chapter VIII with $r > 1$.

PROOF. Let V be the set of functions f in $L^1(\mathbb{R}^N)$ for which this kind of approximation is possible for every $\epsilon > 0$. The main step is to show that V contains the indicator functions of the compact sets in \mathbb{R}^N . Let K be compact, and let I_K be its indicator function. Proposition 6.16 shows that the functions $y \mapsto \tau_y g_k$ are continuous from K into $L^p(\mathbb{R}^N)$ for $1 \leq k \leq r$, and therefore these functions are uniformly continuous. Fix $\epsilon > 0$, and let $\delta > 0$ be such that $\|\tau_y g_k - \tau_{y'} g_k\|_p < \epsilon$ for all k whenever $|y - y'| < \delta$ and y and y' are in K . For each y in K , form the open ball $B(\delta; y)$ in \mathbb{R}^N . These balls cover K , and finitely many suffice; let their centers be y_1, \dots, y_n . Define sets S_1, \dots, S_n inductively as follows: S_j is the subset of K where $|y - y_j| < \delta$ but $|y - y_i| \geq \delta$ for $i < j$. Then $K = \bigcup_{j=1}^n S_j$ disjointly. By the choice of δ , we have $\|\tau_y g_k - \tau_{y_j} g_k\|_p < \epsilon$ for all y in S_j and all k . Using Minkowski's inequality for integrals (Theorem

5.60), and writing m for Lebesgue measure, we have

$$\begin{aligned} \|I_{S_j} * g_k - m(S_j)\tau_{y_j}g_k\|_p &= \left\| \int_{S_j} (g_k(x-y) - g_k(x-y_j)) dy \right\|_p \\ &\leq \int_{S_j} \|g_k(x-y) - g_k(x-y_j)\|_{p,x} dy \\ &\leq \epsilon m(S_j). \end{aligned}$$

Summing over j gives

$$\|I_K * g_k - \sum_{j=1}^n m(S_j)\tau_{y_j}g_k\|_p \leq \epsilon m(K).$$

Since ϵ is arbitrary, I_K lies in V .

If f_1 and f_2 are in V and if g_1, \dots, g_r are given, then we may assume, by taking the union of the sets of members y_j of \mathbb{R}^N and by setting any unnecessary constants c_j equal to 0, that the translates used for f_1 and f_2 with the same $\epsilon > 0$ are the same. Thus we can write $\|f_1 * g_k - \sum_{j=1}^n c_j \tau_{y_j} g_k\|_p < \epsilon/2$ and $\|f_2 * g_k - \sum_{j=1}^n d_j \tau_{y_j} g_k\|_p < \epsilon/2$ for suitable y_j 's and c_j 's, and the triangle inequality gives $\|(f_1 + f_2) * g_k - \sum_{j=1}^n (c_j + d_j)\tau_{y_j} g_k\|_p < \epsilon$. Hence V is closed under addition. Similarly V is closed under scalar multiplication. If $f_l \rightarrow f$ in L^1 with f_l in V and if $\epsilon > 0$ is given, choose l large enough so that $\|f - f_l\|_1 < \epsilon/(2 \max \|g_k\|_p)$. If $\|f_l * g_k - \sum_{j=1}^n c_j^{(l)} \tau_{y_j^{(l)}} g_k\|_p < \epsilon/2$, then the inequality $\|f * g_k - f_l * g_k\|_p \leq \|f - f_l\|_1 \|g_k\|_p$ and the triangle inequality together give $\|f * g_k - \sum_{j=1}^n c_j^{(l)} \tau_{y_j^{(l)}} g_k\|_p < \epsilon$. Hence f is in V , and V is closed. By Corollary 6.4b, $V = L^1(\mathbb{R}^N)$, and the proof is complete. \square

In some cases with $L^\infty(\mathbb{R}^N)$, results have more content when phrased in terms of the supremum norm $\|f\|_{\text{sup}} = \sup_{x \in \mathbb{R}^N} |f(x)|$ defined in Section V.9. For a continuous function f , the two norms agree because the set where $|f(x)| > M$ is open and therefore has positive measure if it is nonempty. For a bounded function f , the condition $\lim_{h \rightarrow 0} \|\tau_h f - f\|_{\text{sup}} = 0$ is equivalent to uniform continuity of f , basically by definition. The functions f in L^∞ for which $\lim_{h \rightarrow 0} \|\tau_h f - f\|_\infty = 0$ are not much more general than the bounded uniformly continuous functions; we shall see shortly that they can be adjusted on a set of measure 0 so as to be bounded and uniformly continuous.

Proposition 6.18. In \mathbb{R}^N with Lebesgue measure, the convolution of L^1 with L^∞ , or of L^∞ with L^1 , or of L^2 with L^2 results in an everywhere-defined bounded uniformly continuous function, not just an L^∞ function. Moreover,

$$\|f * g\|_{\text{sup}} \leq \|f\|_1 \|g\|_\infty, \quad \|f * g\|_{\text{sup}} \leq \|f\|_\infty \|g\|_1, \quad \text{or} \quad \|f * g\|_{\text{sup}} \leq \|f\|_2 \|g\|_2$$

in the various cases.

PROOF. We give the proof when f is in L^1 and g is in L^∞ , the other cases being handled similarly. The bound follows from the computation $\|f * g\|_{\text{sup}} = \sup_x \left| \int_{\mathbb{R}^N} f(x-y)g(y) dy \right| \leq \sup_x \|g\|_\infty \int_{\mathbb{R}^N} |f(x-y)| dy = \|f\|_1 \|g\|_\infty$. For uniform continuity we use Proposition 6.15 and the bound $\|f * g\|_{\text{sup}} \leq \|f\|_1 \|g\|_\infty$ to make the estimate

$$\begin{aligned} \|\tau_h(f * g) - (f * g)\|_{\text{sup}} &= \|(\tau_h f) * g - f * g\|_{\text{sup}} \\ &= \|(\tau_h f - f) * g\|_{\text{sup}} \leq \|\tau_h f - f\|_1 \|g\|_\infty, \end{aligned}$$

and then we apply Proposition 6.16 to see that the right side tends to 0 as h tends to 0. \square

A corollary of Proposition 6.18 gives a first look at how differentiability interacts with convolution.

Corollary 6.19. Suppose that f is a compactly supported function of class C^n on \mathbb{R}^N and that g is in $L^p(\mathbb{R}^N, dx)$ with p equal to 1, 2, or ∞ . Then $f * g$ is of class C^n , and $D(f * g) = (Df) * g$ for any iterated partial derivative of order $\leq n$.

PROOF. First suppose that $n = 1$. Fix j with $1 \leq j \leq N$, and put $D_j = \partial/\partial x_j$. The function $(D_j f) * g$ is continuous by Proposition 6.18. If we can prove that $D_j(f * g)(x)$ exists and equals $((D_j f) * g)(x)$ for each x , then it will follow that $D_j(f * g)$ is continuous. This fact for all j implies that $f * g$ is of class C^1 , by Theorem 3.7, and the result for $n = 1$ will have been proved. The result for higher n can then be obtained by iterating the result for $n = 1$.

Thus we are to prove that $D_j(f * g)(x)$ exists and equals $((D_j f) * g)(x)$ for each x . In the respective cases $p = 1, 2, \infty$, put $p' = \infty, 2, 1$. Let e_j be the j^{th} standard basis vector of \mathbb{R}^N and let h be real with $|h| \leq 1$. Proposition 6.15 gives

$$h^{-1}((f * g)(x + he_j) - (f * g)(x)) = ((h^{-1}(\tau_{-he_j} f - f)) * g)(x). \quad (*)$$

Proposition 3.28a shows that $h^{-1}(\tau_{-he_j} f - f)$ converges uniformly, as $h \rightarrow 0$, to $D_j f$ on any compact set; since the support is compact, $h^{-1}(\tau_{-he_j} f - f)$ converges uniformly to $D_j f$ on \mathbb{R}^N . Hence the convergence occurs in L^∞ , and dominated convergence shows that it occurs in L^1 and L^2 also. Combining Proposition 6.18 and (*), we see that

$$|h^{-1}((f * g)(x + he_j) - (f * g)(x)) - (D_j f * g)(x)| \leq \|h^{-1}(\tau_{-he_j} f - f) - D_j f\|_{p'} \|g\|_p.$$

The right side tends to 0 as $h \rightarrow 0$, and thus indeed $D_j(f * g)(x)$ exists and equals $((D_j f) * g)(x)$. \square

Twice in Chapter I we made use of an “approximate identity” in \mathbb{R}^1 , a system of functions peaking at the origin such that convolution by these functions acts more and more like the identity operator on some class of functions. The first occasion of this kind was in Section I.9 in connection with the Weierstrass Approximation Theorem, where the functions in the system were $\varphi_n(x) = c_n(1-x^2)^n$ on $[-1, 1]$ with the constants c_n chosen to make the total integral be 1. The polynomials φ_n had the properties

- (i) $\varphi_n(x) \geq 0$,
- (ii) $\int_{-1}^1 \varphi_n(x) dx = 1$,
- (iii) for any $\delta > 0$, $\sup_{\delta \leq |x| \leq 1} \varphi_n(x)$ tends to 0 as n tends to infinity,

and the convolutions were with continuous functions f such that $f(0) = f(1) = 0$ and f vanishes outside $[0, 1]$. The second occasion was in Section I.10 in connection with Fejér’s Theorem, where the functions in the system were trigonometric polynomials $K_N(x)$ such that

- (i) $K_N(x) \geq 0$,
- (ii) $\frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(x) dx = 1$,
- (iii) for any $\delta > 0$, $\sup_{\delta \leq |x| \leq \pi} K_N(x)$ tends to 0 as n tends to infinity.

In this case the convolutions were with periodic functions of period 2π over an interval of length 2π , and the integrations involved $\frac{1}{2\pi} dx$ instead of dx .

Now we shall use the dilations of a single function in order to produce a more robust kind of approximate identity, this time on \mathbb{R}^N . One sense in which convolution by this system acts more and more like the identity appears in Theorem 6.20 below, and a sample application appears in Corollary 6.21. The corollary will illustrate how one can use an approximate identity to pass from conclusions about nice functions in some class to conclusions about all functions in the class.

Theorem 6.20. Let φ be in $L^1(\mathbb{R}^N, dx)$, not necessarily ≥ 0 . Define

$$\varphi_\varepsilon(x) = \varepsilon^{-N} \varphi(\varepsilon^{-1}x) \quad \text{for } \varepsilon > 0,$$

and put $c = \int_{\mathbb{R}^N} \varphi(x) dx$. Then the following hold:

- (a) if $p = 1$ or $p = 2$ and if f is in $L^p(\mathbb{R}^N, dx)$, then

$$\lim_{\varepsilon \downarrow 0} \|\varphi_\varepsilon * f - cf\|_p = 0,$$

- (b) the conclusion in (a) is valid for $p = \infty$ if f is in $L^\infty(\mathbb{R}^N, dx)$ and $\lim_{t \rightarrow 0} \|\tau_t f - f\|_\infty = 0$,
- (c) if f is bounded on \mathbb{R}^N and is continuous at x , then $\lim_{\varepsilon \downarrow 0} (\varphi_\varepsilon * f)(x) = cf(x)$,
- (d) the convergence in (c) is uniform for any set E of x ’s such that f is uniformly continuous at the points of E .

PROOF. We prove conclusions (a) and (b) together. Since $\int_{\mathbb{R}^N} \varphi_\varepsilon(y) dy = \varepsilon^{-N} \int_{\mathbb{R}^N} \varphi(\varepsilon^{-1}y) dy = \int_{\mathbb{R}^N} \varphi(y) dy = c$, we have

$$\begin{aligned} |(\varphi_\varepsilon * f)(x) - cf(x)| &= \left| \int_{\mathbb{R}^N} \varphi_\varepsilon(y) f(x-y) dy - cf(x) \right| \\ &= \left| \int_{\mathbb{R}^N} \varphi_\varepsilon(y) [f(x-y) - f(x)] dy \right| \\ &\leq \int_{\mathbb{R}^N} |\varphi_\varepsilon(y)| |f(x-y) - f(x)| dy. \end{aligned} \quad (*)$$

Now we apply Minkowski's inequality for integrals (Theorem 5.60), taking p to be 1, 2, or ∞ , and we obtain

$$\begin{aligned} \|\varphi_\varepsilon * f - cf\|_p &\leq \left\| \int_{\mathbb{R}^N} |\varphi_\varepsilon(y)| |f(x-y) - f(x)| dy \right\|_{p,x} \\ &\leq \int_{\mathbb{R}^N} \|\varphi_\varepsilon(y)| |f(x-y) - f(x)|\|_{p,x} dy \\ &= \int_{\mathbb{R}^N} |\varphi_\varepsilon(y)| \|\tau_y f - f\|_p dy \\ &= \int_{\mathbb{R}^N} |\varphi(y)| \|\tau_{\varepsilon y}(f) - f\|_p dy. \end{aligned}$$

Let ε decrease to 0. For $p = 1$ or $p = 2$, $\|\tau_{\varepsilon y}(f) - f\|_p$ tends to 0 for each y by Proposition 6.16; for $p = \infty$, it tends to 0 for each y by assumption on f . The integrand $|\varphi(y)| \|\tau_{\varepsilon y}(f) - f\|_p$ is dominated pointwise by the integrable function $2|\varphi(y)| \|f\|_p$ independently of ε , and therefore we have dominated convergence along any sequence ε_n tending to 0. Since convergence to a limit within \mathbb{R} occurs as $\varepsilon \downarrow 0$ if and only if convergence to that limit occurs along every sequence decreasing to 0, we conclude that $\lim_{\varepsilon \downarrow 0} \|\varphi_\varepsilon * f - cf\|_p = 0$. This proves (a) and (b).

For (c), inequality (*) and a change of variables by a dilation gives

$$|(\varphi_\varepsilon * f)(x) - cf(x)| \leq \int_{\mathbb{R}^N} |\varphi(y)| |f(x - \varepsilon y) - f(x)| dy.$$

Since f is bounded and φ is integrable, dominated convergence shows that the right side tends to 0 if f is continuous at x . This proves (c).

For (d), let $\eta > 0$ be given, and choose $M > 0$ by the boundedness of f and integrability of φ such that $2(\sup_{t \in \mathbb{R}^N} |f(t)|) \int_{|y| > M} |\varphi(y)| dy \leq \eta$. Then choose $\delta > 0$ by uniform continuity such that $|u| \leq \delta$ implies that $|f(x+u) - f(x)| \leq \eta/\|\varphi\|_1$. Whenever $\varepsilon \leq \delta/M$, the inequality $|y| \leq M$ implies that $|f(x - \varepsilon y) - f(x)| \leq \eta/\|\varphi\|_1$. Then

$$\begin{aligned} \int_{\mathbb{R}^N} |\varphi(y)| |f(x - \varepsilon y) - f(x)| dy &= \left(\int_{|y| > M} + \int_{|y| \leq M} \right) (\text{same}) dy \\ &\leq \eta + \int_{|y| \leq M} |\varphi(y)| (\eta/\|\varphi\|_1) dy \leq 2\eta, \end{aligned}$$

and (d) follows. \square

Corollary 6.21. If f in $L^\infty(\mathbb{R}^N, dx)$ satisfies $\lim_{t \rightarrow 0} \|\tau_t f - f\|_\infty = 0$, then f can be adjusted on a set of measure 0 so as to be uniformly continuous.

PROOF. Let φ be a member of $C_{\text{com}}(\mathbb{R}^N)$ such that $\int_{\mathbb{R}^N} \varphi(x) dx = 1$. Fix a sequence $\{\varepsilon_n\}$ decreasing to 0 in \mathbb{R}^1 . Proposition 6.18 shows that each $\varphi_{\varepsilon_n} * f$ is bounded and uniformly continuous for every n , and Theorem 6.20 shows that $\{\varphi_{\varepsilon_n} * f\}$ is Cauchy in L^∞ . Since the L^∞ and supremum norms coincide for continuous functions, $\{\varphi_{\varepsilon_n} * f\}$ is uniformly Cauchy and must therefore be uniformly convergent. Let g be the limit function, which is necessarily bounded and uniformly continuous. Then $\|f - g\|_\infty \leq \|f - \varphi_{\varepsilon_n} * f\|_\infty + \|\varphi_{\varepsilon_n} * f - g\|_\infty$, and both terms on the right tend to 0 as n tends to infinity. Consequently $\|f - g\|_\infty = 0$, and g is a bounded uniformly continuous function that differs from f only on a set of measure 0. \square

3. Borel Measures on Open Sets

A number of results in Sections 1–2 about Borel measures on \mathbb{R}^N extend to suitably defined Borel measures on arbitrary nonempty open subsets V of \mathbb{R}^N , and we shall collect some of these results here in order to do two things: to prepare for the proof in Section 5 of the change-of-variables formula for the Lebesgue integral in \mathbb{R}^N and to provide motivation for the treatment in Chapter XI of Borel measures on locally compact Hausdorff spaces.

Throughout this section, let V be a nonempty open subset of \mathbb{R}^N . We shall make use of the following lemma that generalizes to V three properties (i–iii) listed for \mathbb{R}^N before Theorem 6.2 and Corollary 6.3. Let $C_{\text{com}}(V)$ be the vector space of scalar-valued continuous functions on V of compact support in V . If nothing is said to the contrary, the scalars may be either real or complex.

Lemma 6.22.

(a) There exists a sequence $\{F_n\}_{n=1}^\infty$ of compact subsets of V with union V such that $F_n \subseteq F_{n+1}^o$ for all n .

(b) For any compact subset K of V , there exists a decreasing sequence of open sets U_n with compact closure in V such that $\bigcap_{n=1}^\infty U_n = K$.

(c) For any compact subset K of V , there exists a decreasing sequence of functions in $C_{\text{com}}(V)$ with values in $[0, 1]$ and with pointwise limit the indicator function of K .

PROOF. In (a), the case $V = \mathbb{R}^N$ was handled by (i) before Theorem 6.2. For $V \neq \mathbb{R}^N$, we can take $F_n = \{x \in V \mid D(x, V^c) \geq 1/n \text{ and } |x| \leq n\}$ as long as n is \geq some suitable n_0 . We complete the definition of the F_n 's by taking $F_1 = \cdots = F_{n_0-1} = \emptyset$.

In (b), the case $V = \mathbb{R}^N$ was handled by (ii) before Theorem 6.2. For $V \neq \mathbb{R}^N$, every x in K has $D(x, V^c) > 0$ since V^c is closed and is disjoint from K . The function $D(\cdot, V^c)$ is continuous and therefore has a positive minimum on K . Choose n_0 such that $D(x, V^c) \geq 1/n_0$ for x in K , i.e., $|x - y| \geq 1/n_0$ for all $x \in K$ and $y \in V^c$. Then $D(y, K) \geq 1/n_0$ if y is not in V . Let $U_n = \{y \in \mathbb{R}^N \mid D(y, K) < 1/n\}$ for $n > n_0$. This is an open set containing K , and its closure in \mathbb{R}^N is contained in the set where $D(y, K) \leq 1/n$, which in turn is contained in V . The set where $D(y, K) \leq 1/n$ is closed and bounded in \mathbb{R}^N and hence is compact. Therefore U_n^{cl} is contained in a compact subset of V . We complete the definition of the U_n 's by letting U_1, \dots, U_{n_0} all equal U_{n_0+1} .

For (c), we argue as with (iii) before Corollary 6.3. Choose open sets U_n as in (b) that decrease and have intersection K , and apply Proposition 2.30e to obtain continuous functions $f_n : \mathbb{R}^N \rightarrow [0, 1]$ such that f_n is 1 on K and is 0 on U_n^c . The support of the function f_n is then contained in U_n^{cl} , which is compact. By replacing the functions f_n by $g_n = \min\{f_1, \dots, f_n\}$, we may assume that they are pointwise decreasing. Then (c) follows. \square

The **Borel sets** in the open set V are the sets in the σ -algebra

$$\mathcal{B}_N(V) = \mathcal{B}_N \cap V = \{E \cap V \mid E \in \mathcal{B}_N\}$$

of subsets of V . We can regard V as a metric space by restricting the distance function on \mathbb{R}^N , and because V is open, the open sets of V are the open sets of \mathbb{R}^N that are subsets of V . We shall prove the following proposition about these Borel sets after first proving a lemma.

Proposition 6.23. The σ -algebra $\mathcal{B}_N(V)$ is the smallest σ -algebra for V containing the open sets of V , and it is the smallest σ -algebra for V containing the compact sets of V .

Lemma 6.24. Let X be a nonempty set, let \mathcal{U} be a family of subsets of X , let \mathcal{B} be the smallest σ -ring of subsets of X containing \mathcal{U} , and let E be a member of \mathcal{B} . Then $\mathcal{B} \cap E$ is the smallest σ -ring containing $\mathcal{U} \cap E$.

PROOF OF LEMMA 6.24. Let \mathcal{A} be the smallest σ -ring containing $\mathcal{U} \cap E$, and let \mathcal{A}' be the smallest σ -ring containing $\mathcal{U} \cap E^c$. Since $\mathcal{B} \cap E$ is a σ -ring of subsets of X containing $\mathcal{U} \cap E$, \mathcal{A} is contained in $\mathcal{B} \cap E$. Similarly $\mathcal{A}' \subseteq \mathcal{B} \cap E^c$. Thus the set of unions $A \cup A'$ with $A \in \mathcal{A}$ and $A' \in \mathcal{A}'$ is contained in \mathcal{B} , contains \mathcal{U} , and is closed under countable unions. To see that it is closed under differences, let $A_1 \cup A'_1$ and $A_2 \cup A'_2$ be such unions. Then $(A_1 \cup A'_1) - (A_2 \cup A'_2) = (A_1 - A_2) \cup (A'_1 - A'_2)$ exhibits the difference of the given sets as such a union. Hence the set of such unions is a σ -ring and must equal \mathcal{B} . \square

PROOF OF PROPOSITION 6.23. The statement about open sets follows from Lemma 6.24 by taking X to be \mathbb{R}^N , \mathcal{U} to be the set of open sets in \mathbb{R}^N , and E to be V . The set $\mathcal{U} \cap E$ is the set of open subsets of V , and the lemma says that the smallest σ -ring containing $\mathcal{U} \cap E$ is $\mathcal{B}_N(V)$. This is a σ -algebra of subsets of V since V itself is in $\mathcal{U} \cap V$.

Let $\{F_n\}$ be the sequence of compact subsets of V produced by Lemma 6.22a. Since $V = \bigcup_{n=1}^{\infty} F_n$, V is a member of the smallest σ -ring of subsets containing the compact subsets of V . If F is a relatively closed subset of V , then each $F \cap F_n$ is compact, and the countable union F is therefore in this σ -algebra. Taking complements, we see that every open subset of V is in the smallest σ -algebra of subsets of V containing the compact sets. Therefore $\mathcal{B}_N(V)$ is contained in this σ -algebra and must equal this σ -algebra. \square

A **Borel function** on V is a scalar-valued function measurable with respect to $\mathcal{B}_N(V)$. A **Borel measure** on V is a measure on $\mathcal{B}_N(V)$ that is finite on every compact set in V .

Theorem 6.25. Every Borel measure μ on the nonempty open subset V of \mathbb{R}^N is **regular** in the sense that the value of μ on any Borel set E in V is given by

$$\mu(E) = \sup_{\substack{K \subseteq E, \\ K \text{ compact in } V}} \mu(K) = \inf_{\substack{U \supseteq E, \\ U \text{ open in } V}} \mu(U).$$

REMARK. If μ is a Borel measure on V and if we define $\nu(E) = \mu(E \cap V)$ for Borel sets E of \mathbb{R}^N , then ν is a measure on the Borel sets of \mathbb{R}^N , but ν need not be finite on compact sets. Thus Theorem 6.25 is not a special case of Theorem 6.2.

PROOF. This is proved from parts (a) and (b) of Lemma 6.22 in exactly the same way that Theorem 6.2 is proved from items (i) and (ii) before the statement of that theorem. \square

Corollary 6.26. If μ and ν are Borel measures on V such that $\int_V f d\mu = \int_V f d\nu$ for all f in $C_{\text{com}}(V)$, then $\mu = \nu$.

PROOF. This is proved from Theorem 6.25 and Lemma 6.22c in the same way that Corollary 6.3 is proved from Theorem 6.2 and item (iii) before the statement of that corollary. \square

Corollary 6.27. Let $p = 1$ or $p = 2$. If μ is a Borel measure on V , then

- (a) $C_{\text{com}}(V)$ is dense in $L^p(V, \mu)$,
- (b) the smallest closed subspace of $L^p(V, \mu)$ containing all indicator functions of compact subsets of V is $L^p(V, \mu)$ itself,

- (c) $C_{\text{com}}(V)$, as a normed linear space under the supremum norm, is separable,
 (d) $L^p(V, \mu)$ is separable.

PROOF. Conclusions (a) and (b) are proved from Lemma 6.22c with the aid of Propositions 5.56 and 5.55d in the same way that Corollary 6.4 is proved from item (iii) before Corollary 6.3 with the aid of those propositions.

For (c), Lemma 6.22a produces a sequence $\{F_n\}_{n=1}^{\infty}$ of compact subsets of V with union V such that $F_n \subseteq F_{n+1}^o$ for all n . Since the open sets F_n^o cover V , they cover any compact subset K of V , and K must be contained in some F_n^o . Let us put that observation aside for a moment. For each n , we can identify the vector subspace of $C_{\text{com}}(V)$ of functions supported in F_n^o with a vector subspace of $C(F_n)$. The latter is separable by Corollary 2.59, and hence the vector subspace of $C_{\text{com}}(V)$ of functions supported in F_n^o is separable. If we form the union on n of these countable dense sets of certain vector subspaces and if we take into account that the functions supported in any compact subset of V have compact support within some F_n^o , we see that $C_{\text{com}}(V)$ is separable.

For (d), we apply (a) and (c) with V replaced by F_n^o and take into account that $\mu(F_n) < \infty$. Let f be arbitrary in $L^p(F_n^o, \mu|_{F_n^o})$. If $\epsilon > 0$ is given, choose g in $C_{\text{com}}(F_n^o)$ with $\|f - g\|_p \leq \epsilon$. Then choose h in the countable dense set D_n of $C_{\text{com}}(F_n^o)$ such that $\|g - h\|_{\text{sup}} \leq \epsilon$. Since $\|f - h\|_p \leq \|f - g\|_p + \|g - h\|_p$ and $\|g - h\|_p^p = \int_{F_n^o} |g(x) - h(x)|^p d\mu(x) \leq \epsilon^p \mu(F_n)$, we obtain $\|f - h\|_p \leq \epsilon + \epsilon \mu(F_n)^{1/p}$. Hence the closure in $L^p(V, \mu)$ of the countable set $D = \bigcup_{n=1}^{\infty} D_n$ contains f . In particular, D^{cl} contains a vector subspace containing all indicator functions of compact subsets of V . By (b), $D^{\text{cl}} = L^p(V, \mu)$. \square

Proposition 6.28. With V still open in \mathbb{R}^N , let V' be an open set in \mathbb{R}^N . Under a continuous function or even a Borel measurable function $F : V \rightarrow V'$, $F^{-1}(E)$ is in $\mathcal{B}_N(V)$ whenever E is in $\mathcal{B}_{N'}(V')$.

PROOF. The set of E 's for which $F^{-1}(E)$ is in $\mathcal{B}_N(V)$ is a σ -algebra, and this σ -algebra contains the open geometric rectangles of \mathbb{R}^N by the same argument as for Proposition 6.8. Thus it contains $\mathcal{B}_{N'}(V')$. \square

Corollary 6.29. If V' is a second nonempty open subset in \mathbb{R}^N besides V , then any homeomorphism of V onto V' carries $\mathcal{B}_N(V)$ to $\mathcal{B}_N(V')$.

If K is a nonempty compact subset of \mathbb{R}^N , it will be convenient to be able to speak of the Borel sets in K , just as we can speak of the Borel sets in an open subset V of \mathbb{R}^N . The theory for K is easier than the theory for V , partly because Borel measures on K can all be obtained by restriction from Borel measures on \mathbb{R}^N .

The **Borel sets** in K are the sets in $\mathcal{B}_N(K) = \mathcal{B}_N \cap K$. Using Lemma 6.24, we readily see that $\mathcal{B}_N(K)$ is the smallest σ -algebra for K containing the compact subsets of K ; the argument is simpler than the corresponding proof for Proposition 6.23 in that it is not necessary to produce some sequence of sets by means of Lemma 6.22.

A **Borel function** on the compact set K is a scalar-valued function measurable with respect to $\mathcal{B}_N(K)$. A **Borel measure** on K is a measure on $\mathcal{B}_N(K)$ that is finite on compact subsets of K . In this situation regularity is a consequence of the regularity of Borel measures on \mathbb{R}^N , and no separate argument is needed. In fact, if μ is a Borel measure on K , we can define $\nu(E) = \mu(E \cap K)$ for each Borel subset E on \mathbb{R}^N , and then the finiteness of $\mu(K)$ implies that ν is a Borel measure on \mathbb{R}^N . Borel measures on \mathbb{R}^N are regular, and therefore we have

$$\nu(E) = \sup_{\substack{K' \subseteq E, \\ K' \text{ compact in } \mathbb{R}^N}} \nu(K') = \inf_{\substack{U \supseteq E, \\ U \text{ open in } \mathbb{R}^N}} \nu(U).$$

Replacing E by $E \cap K$ and substituting from the definition of ν , we obtain the following proposition.

Proposition 6.30. Every Borel measure μ on a compact nonempty set K in \mathbb{R}^N is **regular** in the sense that the value of μ on any Borel set E in K is given by

$$\mu(E) = \sup_{\substack{K' \subseteq E, \\ K' \text{ compact in } K}} \mu(K') = \inf_{\substack{U \supseteq E, \\ U \text{ relatively} \\ \text{open in } K}} \mu(U).$$

4. Comparison of Riemann and Lebesgue Integrals

This section contains the definitive theorem about the relationship between the Riemann integral and the Lebesgue integral in \mathbb{R}^N . The Riemann integral is defined in Section III.7, the Lebesgue integral is defined in Section V.3, and Lebesgue measure in \mathbb{R}^N is defined in Section 1 of the present chapter. In order to have a notational distinction between the Riemann and Lebesgue integrals, we write in this section $\mathcal{R} \int_A f \, dx$ for the Riemann integral of a bounded real-valued function on a compact geometric rectangle A , and we write $\int_A f \, dx$ for the Lebesgue integral.

Theorem 6.31. Suppose that f is a bounded real-valued function on a compact geometric rectangle A in \mathbb{R}^N . Then f is Riemann integrable on A if and only if f is continuous except on a Lebesgue measurable set of Lebesgue measure 0. In this case, f is Lebesgue measurable, and $\mathcal{R} \int_A f \, dx = \int_A f \, dx$.

PROOF. Proposition 6.12 shows that a set in \mathbb{R}^N has “measure 0” in the sense of Chapter III if and only if it is Lebesgue measurable of measure 0, and Theorem 3.29 shows that f is Riemann integrable on A if and only if f is continuous except on a set of measure 0. This proves the first conclusion of the theorem.

For the second conclusion, suppose that $\mathcal{R} \int_A f \, dx$ exists. Lemma 3.23 shows that there exists a sequence of partitions $P^{(k)}$ of A , each refining the previous one, such that the lower Riemann sums $L(P^{(k)}, f)$ increase to $\mathcal{R} \int_A f \, dx$ and the upper Riemann sums $U(P^{(k)}, f)$ decrease to $\mathcal{R} \int_A f \, dx$. For each k , we define Borel functions L_k and U_k on A as follows: If x is an interior point of some component (closed) rectangle S of $P^{(k)}$, we define $L_k(x) = m_S(f)$, where $m_S(f)$ is the infimum of f on S ; otherwise we let $L_k(x) = 0$. If we write $|S|$ for the volume of S , then the Lebesgue integral of L_k over S is given by $\int_S L_k(x) \, dx = m_S(f)|S|$. Consequently

$$\int_A L_k(x) \, dx = \sum_S m_S(f)|S| = L(P^{(k)}, f).$$

We define $U_k(x)$ similarly, using the supremum $M_S(f)$ of f on S instead of the infimum, and then

$$\int_A U_k(x) \, dx = \sum_S M_S(f)|S| = U(P^{(k)}, f).$$

Let E be the subset of points x in A such that x is in the interior of a component rectangle of $P^{(k)}$ for all k . The set $A - E$ is a Borel set of Lebesgue measure 0. Since $P^{(k+1)}$ is a refinement of $P^{(k)}$ for every k , we have $L_k(x) \leq L_{k+1}(x)$ and $U_k(x) \geq U_{k+1}(x)$ for all x in E and all k . Therefore $L(x) = \lim L_k(x)$ and $U(x) = \lim U_k(x)$ exist for x in E . Since $L_k(x) \leq f(x) \leq U_k(x)$ for x in E , we see that

$$L(x) \leq f(x) \leq U(x) \quad \text{for all } x \text{ in } E.$$

Define $L(x) = U(x) = 0$ on $A - E$. Then L and U are Borel functions with $L(x) \leq U(x)$ everywhere on A . On E , we have dominated convergence, and thus

$$\int_E L(x) \, dx = \lim_k \int_E L_k(x) \, dx \quad \text{and} \quad \int_E U(x) \, dx = \lim_k \int_E U_k(x) \, dx.$$

The set $A - E$ has Lebesgue measure 0, and therefore these equations imply that

$$\int_A L(x) \, dx = \lim_k \int_A L_k(x) \, dx \quad \text{and} \quad \int_A U(x) \, dx = \lim_k \int_A U_k(x) \, dx.$$

Consequently

$$\begin{aligned} \int_A L(x) dx &= \lim_k \int_A L_k(x) dx = \lim_k L(P^{(k)}, f) = \mathcal{R} \int_A f dx \\ &= \lim_k U(P^{(k)}, f) = \lim_k \int_A U_k(x) dx = \int_A U(x) dx. \end{aligned}$$

Since $L(x) \leq U(x)$ on A , Corollary 5.23 shows that the set F where $L(x) = U(x)$ is a Borel set such that $A - F$ has Lebesgue measure 0. Since the inequalities $L(x) \leq f(x) \leq U(x)$ are valid for x in E , $f(x)$ equals $L(x)$ at least on $E \cap F$. The set $E \cap F$ is a Borel set, and L is Borel measurable; hence the restriction of f to $E \cap F$ is Borel measurable. The set $A - (E \cap F)$ is a Borel set of measure 0, and the restriction of f to this set is Lebesgue measurable no matter what values f assumes on this set. Thus f is Lebesgue measurable. Then the Lebesgue integral $\int_A f dx$ is defined, and we have

$$\int_A f(x) dx = \int_{E \cap F} f(x) dx = \int_{E \cap F} L(x) dx = \int_A L(x) dx = \mathcal{R} \int_A f dx. \quad \square$$

5. Change of Variables for the Lebesgue Integral

A general-looking change-of-variables formula for the Riemann integral was proved in Section III.10. On closer examination of the theorem, we found that the result did not fully handle even as ostensibly simple a case as the change from Cartesian coordinates in \mathbb{R}^2 to polar coordinates. Lebesgue integration gives us methods that deal with all the unpleasantness that was concealed by the earlier formula.

Theorem 6.32 (change-of-variables formula). Let φ be a one-one function of class C^1 from an open subset U of \mathbb{R}^N onto an open subset $\varphi(U)$ of \mathbb{R}^N such that $\det \varphi'(x)$ is nowhere 0. Then

$$\int_{\varphi(U)} f(y) dy = \int_U f(\varphi(x)) |\det \varphi'(x)| dx$$

for every nonnegative Borel function f defined on $\varphi(U)$.

REMARK. The σ -algebra on $\varphi(U)$ is understood to be $\mathcal{B}_N \cap \varphi(U)$, the set of intersections of Borel sets in \mathbb{R}^N with the open set $\varphi(U)$. If f is extended from $\varphi(U)$ to \mathbb{R}^N by defining it to be 0 off $\varphi(U)$, then measurability of f with respect to this σ -algebra is the same as measurability of the extended function with respect to \mathcal{B}_N .

PROOF. Theorem 3.34 gives us the change-of-variables formula, as an equality of Riemann integrals, for every f in $C_{\text{com}}(\varphi(U))$. In this case the integrands on both sides, when extended to be 0 outside the regions of integration, are continuous on all of \mathbb{R}^N , and the integrations can be viewed as involving continuous functions on compact geometric rectangles. Proposition 6.11 (or Theorem 6.31 if one prefers) allows us to reinterpret the equality as an equality of Lebesgue integrals.

In the extension of this identity to all nonnegative Borel functions, measurability will not be an issue. The function f is to be measurable with respect to $\mathcal{B}_N(\varphi(U))$, and Corollary 6.29 shows that such f 's correspond exactly to functions $f \circ \varphi$ measurable with respect to $\mathcal{B}_N(U)$.

Using Theorem 5.19, define a measure μ on $\mathcal{B}_N(U)$ by

$$\mu(E) = \int_E |\det \varphi'(x)| dx.$$

Corollary 5.28 implies that μ satisfies

$$\int_U g d\mu = \int_U g(x) |\det \varphi'(x)| dx \quad (*)$$

for every nonnegative g on U measurable with respect to $\mathcal{B}_N(U)$. Next define another set function ν on $\mathcal{B}_N(U)$ by

$$\nu(E) = m(\varphi(E)),$$

where m is Lebesgue measure. It is immediate that ν is a measure, and we have $\int_{\varphi(U)} I_E(\varphi^{-1}(y)) dy = \int_{\varphi(U)} I_{\varphi(E)} dy = m(\varphi(E)) = \nu(E) = \int_U I_E d\nu$. Passing to simple functions ≥ 0 and then using monotone convergence, we obtain

$$\int_{\varphi(U)} g \circ \varphi^{-1} dy = \int_U g d\nu \quad (**)$$

for every nonnegative g on U measurable with respect to $\mathcal{B}_N(U)$.

If in (**) and (*) we take $g = f \circ \varphi$ with f in $C_{\text{com}}(\varphi(U))$ and we substitute into the change-of-variables formula as it is given for f in $C_{\text{com}}(\varphi(U))$, we obtain the identity

$$\int_U g d\nu = \int_U g d\mu \quad (\dagger)$$

for all g in $C_{\text{com}}(U)$. From Corollary 6.26 we conclude that $\mu = \nu$. Hence (\dagger) holds for every nonnegative g on U measurable with respect to $\mathcal{B}_N(U)$. We unwind (\dagger) using (**) and (*) with $g = f \circ \varphi$ but now taking f to be any nonnegative function on $\varphi(U)$ measurable with respect to $\mathcal{B}_N(\varphi(U))$, and we obtain the conclusion of the theorem. \square

Let us return to the example of polar coordinates in \mathbb{R}^2 , first considered in Section III.10. The data in the theorem are

$$U = \left\{ \begin{pmatrix} r \\ \theta \end{pmatrix} \mid 0 < r < +\infty \text{ and } 0 < \theta < 2\pi \right\},$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \varphi \begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix},$$

and we have

$$\varphi(U) = \mathbb{R}^2 - \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \mid x \geq 0 \right\}.$$

Since $\det \varphi' \begin{pmatrix} r \\ \theta \end{pmatrix} = r$, Theorem 6.32 gives

$$\int_{\varphi(U)} f(x, y) dx dy = \int_{0 < r < \infty, 0 < \theta < 2\pi} f(r \cos \theta, r \sin \theta) r dr d\theta$$

for every nonnegative Borel function f on $\varphi(U)$. The set of integration $\varphi(U)$ on the left side is not quite the whole plane; it omits the part of the x axis where $x \geq 0$. But this is a harmless defect: this subset of the x axis is contained in the entire x axis, which is an abstract rectangle in the sense of Fubini's Theorem and has measure 0. Thus the formula can be changed to read

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{0 \leq r < \infty, 0 \leq \theta < 2\pi} f(r \cos \theta, r \sin \theta) r dr d\theta$$

for every nonnegative Borel function f on \mathbb{R}^2 . Here is an application of this formula that we shall use in proving the Fourier Inversion Formula in Chapter VIII.

Proposition 6.33. $\int_{-\infty}^{\infty} e^{-\pi x^2} dx = 1.$

REMARK. Since we now know from Theorem 6.31 that there is no discrepancy between the Riemann integral and the Lebesgue integral with respect to Lebesgue measure, there will be no harm in the future in writing limits of integration in the usual way for integrals with respect to 1-dimensional Lebesgue measure.

PROOF. We use polar coordinates and Fubini's Theorem to compute the square of the integral in question:

$$\begin{aligned} \left(\int_{\mathbb{R}} e^{-\pi x^2} dx \right)^2 &= \int_{\mathbb{R}^2} e^{-\pi x^2} e^{-\pi y^2} dx dy = \int_{\mathbb{R}^2} e^{-\pi(x^2+y^2)} dx dy \\ &= \int_0^\infty \int_0^{2\pi} e^{-\pi r^2} r d\theta dr = 2\pi \int_0^\infty r e^{-\pi r^2} dr \\ &= 2\pi \lim_N \int_0^N r e^{-\pi r^2} dr = \lim_N \left[-e^{-\pi r^2} \right]_0^N = 1. \end{aligned}$$

Since the integral in question is certainly > 0 , the proposition follows. \square

Proposition 6.33 is closely related to properties of the “gamma function,” a certain function of a complex variable that reduces essentially to the factorial function on the positive integers. The definition of the gamma function makes use of the expression t^s defined for $0 < t < +\infty$ and s in \mathbb{C} by $t^s = e^{s \log t}$.

Fix $s \in \mathbb{C}$ with $\operatorname{Re} s > 0$. The function $t \mapsto t^{s-1}e^{-t}$ is continuous on $(0, +\infty)$ and hence Borel measurable. Let us see that it is integrable with respect to Lebesgue measure. Since $|t^{s-1}e^{-t}| = t^{\operatorname{Re} s - 1}e^{-t}$, we may assume that s is real (and positive) in showing the integrability. Integrability on $(0, 1]$ is no problem, since we know that $\int_0^1 t^{s-1} dt < \infty$ for $s > 0$. To handle $[1, +\infty)$, let n be an integer $\geq s - 1$. Then $t^{s-1} \leq t^n = 2^n n! \left(\frac{1}{n!} \left(\frac{t}{2}\right)^n\right) \leq 2^n n! \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{t}{2}\right)^k = 2^n n! e^{t/2}$. Hence $t^{s-1}e^{-t} \leq 2^n n! e^{-t/2}$, and the integrability on $[1, +\infty)$ follows.

With this integrability in place, we define the **gamma function** by

$$\Gamma(s) = \int_0^{\infty} t^{s-1} e^{-t} dt \quad \text{for } \operatorname{Re} s > 0.$$

Proposition 6.34. The gamma function has the properties that

- (a) $\Gamma(s + 1) = s\Gamma(s)$ for $\operatorname{Re} s > 0$,
- (b) $\Gamma(1) = 1$ and $\Gamma(n + 1) = n!$ for integers $n > 0$,
- (c) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

PROOF. Part (a) follows from integration by parts, which needs to be done on an interval $[\varepsilon, M]$ and followed by passages to the limit $\varepsilon \rightarrow 0$ and $M \rightarrow \infty$. In (b), the formula $\Gamma(1) = 1$ just amounts to the elementary integral $\int_0^{\infty} e^{-t} dt = 1$, and then the formula $\Gamma(n + 1) = n!$ for integers $n > 0$ follows by iterating (a). For (c), the change of variables $t = \pi x^2$ gives

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} t^{-1/2} e^{-t} dt = \int_0^{\infty} (\pi x^2)^{-1/2} e^{-\pi x^2} 2\pi x dx = 2\sqrt{\pi} \int_0^{\infty} e^{-\pi x^2} dx.$$

Since $\int_0^{\infty} e^{-\pi x^2} dx = \frac{1}{2} \int_{-\infty}^{\infty} e^{-\pi x^2} dx$, Proposition 6.33 allows us to conclude that $2 \int_0^{\infty} e^{-\pi x^2} dx = 1$. Hence $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. \square

It is often true in applications of the change-of-variables formula that the set $\varphi(U)$ does not exhaust the set that one might hope to have as region of integration. For polar coordinates the exceptional set was the part of the x axis with $x \geq 0$, and an easy argument showed that the exceptional set had measure 0. In a more complicated example, that easy argument will not ordinarily apply, but still the exceptional set has a certain “lower-dimensional” quality to it. A general result saying that certain lower-dimensional sets have measure 0 will be given as a corollary of Sard’s Theorem, which we prove now.

Let $\psi : V \rightarrow \mathbb{R}^N$ be a smooth map defined on an open subset V of \mathbb{R}^N . A **critical point** x of ψ is a point where $\psi'(x)$ has rank $< N$. In this case, $\psi(x)$ is called a **critical value**.

Theorem 6.35 (Sard's Theorem). If $\psi : V \rightarrow \mathbb{R}^N$ is a smooth map defined on an open subset V of \mathbb{R}^N , then the set of critical values of ψ is a Borel set of Lebesgue measure 0 in \mathbb{R}^N .

PROOF. The set where $\psi'(x)$ has rank $\leq N - 1$ is relatively closed in V and hence is the union of countably many compact sets. The set of critical values is then the union of the compact images of these sets and consequently is a Borel set. Let us see that this Borel set has Lebesgue measure 0. Since V is the countable union of compact geometric rectangles and since the countable union of sets of measure 0 is of measure 0, it is enough to prove the theorem for the restriction of ψ to a compact geometric rectangle R .

For points $x = (x_1, \dots, x_N)$ and $x' = (x'_1, \dots, x'_N)$ in R , the Mean Value Theorem gives

$$\psi_i(x') - \psi_i(x) = \sum_{j=1}^N \frac{\partial \psi_i}{\partial x_j}(z_i)(x'_j - x_j), \quad (*)$$

where z_i is a point on the line segment from x to x' . Since the $\frac{\partial \psi_i}{\partial x_j}$ are bounded on R , we see as a consequence that

$$|\psi(x') - \psi(x)| \leq a|x' - x| \quad (**)$$

with a independent of x and x' . Let $L_x(x') = (L_{x,1}(x'), \dots, L_{x,N}(x'))$ be the best first-order approximation to ψ about x , namely

$$L_{x,i}(x') = \psi_i(x) + \sum_{j=1}^N \frac{\partial \psi_i}{\partial x_j}(x)(x'_j - x_j).$$

Subtracting this equation from (*), we obtain

$$\psi_i(x') - L_{x,i}(x') = \sum_{j=1}^N \left(\frac{\partial \psi_i}{\partial x_j}(z_i) - \frac{\partial \psi_i}{\partial x_j}(x) \right) (x'_j - x_j).$$

Since $\frac{\partial \psi_i}{\partial x_j}$ is smooth and $|z_i - x| \leq |x' - x|$, we deduce that

$$|\psi(x') - L_x(x')| \leq b|x' - x|^2 \quad (\dagger)$$

with b independent of x and x' .

If x is a critical point, let us bound the image of the set of x' with $|x' - x| \leq c$. The determinant of the linear part of L_x is 0, and hence L_x has image in a hyperplane, not necessarily a coordinate hyperplane. By (\dagger) , $\psi(x')$ has distance

$\leq bc^2$ from this hyperplane. In each of the $N - 1$ perpendicular directions, (**) shows that $\psi(x')$ and $\psi(x)$ are at distance $\leq ac$ from each other. Thus $\psi(x')$ is contained in a box¹ centered at $\psi(x)$ with volume $2^N(ac)^{N-1}(bc^2) = 2^N a^{N-1}bc^{N+1}$.

We subdivide R into M^N smaller compact geometric rectangles whose dimensions are $1/M$ times those of R . If d is the diameter of R and if one of these smaller geometric rectangles R' contains a critical point x , then any point x' in R' has $|x' - x| \leq d/M$. By the result of the previous paragraph, ψ of R' is contained in a box of volume $2^N a^{N-1}b(d/M)^{N+1}$. The union of these boxes, taken over all of the smaller geometric rectangles containing critical points, contains the critical values. Since there are at most M^N of the smaller geometric rectangles, the outer measure m^* of the set of critical values, where m refers to Lebesgue measure, is $\leq 2^N a^{N-1}bd^{N+1}M^{-1}$. This estimate is valid for all M , and hence the set S of critical values has $m^*(S) = 0$. Therefore the Borel set S has Lebesgue measure 0. \square

Corollary 6.36. If $\psi : V \rightarrow \mathbb{R}^N$ is a smooth map defined on an open subset V of \mathbb{R}^M with $M < N$, then the image of ψ is a Borel set of Lebesgue measure 0 in \mathbb{R}^N .

PROOF. Sard's Theorem (Theorem 6.35) applies to the composition of the projection $\mathbb{R}^N \rightarrow \mathbb{R}^M$ followed by ψ . Every point of the domain is a critical point, and hence every point of the image is a critical value. The result follows. \square

We define a **lower-dimensional set** in \mathbb{R}^N to be any set contained in the countable union of smooth images of open sets in Euclidean spaces of dimension $< N$. The following result is immediate from Corollary 6.36.

Corollary 6.37. Any lower-dimensional set in \mathbb{R}^N is Lebesgue measurable of Lebesgue measure 0.

The N -dimensional generalization of polar coordinates in \mathbb{R}^2 is **spherical coordinates** in \mathbb{R}^N . In the notation of Theorem 6.32, we have

$$U = \left\{ \left(\begin{array}{c} r \\ \theta_1 \\ \vdots \\ \theta_{N-1} \end{array} \right) \mid \begin{array}{l} 0 < r < +\infty, \\ 0 < \theta_j < \pi \text{ for } 1 \leq j \leq N - 2, \\ 0 < \theta_{N-1} < 2\pi \end{array} \right\}$$

and
$$\begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} = \varphi \begin{pmatrix} r \\ \theta_1 \\ \vdots \\ \theta_{N-1} \end{pmatrix} = \begin{pmatrix} r \cos \theta_1 \\ r \sin \theta_1 \cos \theta_2 \\ \vdots \\ r \sin \theta_1 \cdots \sin \theta_{N-2} \cos \theta_{N-1} \\ r \sin \theta_1 \cdots \sin \theta_{N-2} \sin \theta_{N-1} \end{pmatrix}.$$

¹This box need not have its faces parallel to the coordinate hyperplanes.

Problem 2 at the end of the chapter asks for three things to be checked:

(i) the determinant factor in the change-of-variables formula is given by

$$|\det \varphi'| = r^{N-1} \sin^{N-2} \theta_1 \sin^{N-3} \theta_2 \cdots \sin \theta_{N-2},$$

(ii) φ is one-one on U ,

(iii) the complement of $\varphi(U)$ in \mathbb{R}^N is a lower-dimensional set.

Then it follows that the change-of-variables formula applies and that the integration over $\varphi(U)$ can be extended over \mathbb{R}^N . We can write the result as

$$\begin{aligned} \int_{\mathbb{R}^N} f(x) dx &= \int_{r=0}^{\infty} \int_{\theta_1=0}^{\pi} \cdots \int_{\theta_{N-2}=0}^{\pi} \int_{\theta_{N-1}=0}^{2\pi} f(r \cos \theta_1, \dots) \\ &\quad \times r^{N-1} \sin^{N-2} \theta_1 \cdots \sin \theta_{N-2} d\theta_{N-1} \cdots d\theta_1 dr. \end{aligned}$$

The expression $\sin^{N-2} \theta_1 \cdots \sin \theta_{N-2} d\theta_{N-1} \cdots d\theta_1$ we abbreviate as $d\omega$. Geometrically it is the contribution to Lebesgue measure on \mathbb{R}^N from the sphere S^{N-1} of radius 1 centered at the origin. In Chapter XI we shall speak of Borel sets in any compact metric space. The sphere S^{N-1} is a compact metric space, and we shall note that $d\omega$ refers to a rotation-invariant Borel measure on S^{N-1} . We write

$$\Omega_{N-1} = \int_{S^{N-1}} d\omega$$

for the “area” of the sphere S^{N-1} . This constant is evaluated in Problem 12 at the end of the present chapter with the aid of Proposition 6.33. In terms of $d\omega$, the change-of-variables formula for spherical coordinates is

$$\int_{\mathbb{R}^N} f(x) dx = \int_{r=0}^{\infty} \int_{\omega \in S^{N-1}} f(r\omega) r^{N-1} d\omega dr.$$

This formula allows us quickly to check the integrability of powers of $|x|$ near the origin and near ∞ . In fact, we have

$$\int_{|x| \leq 1} |x|^q dx = \Omega_{N-1} \int_0^1 r^{q+N-1} dr$$

and

$$\int_{|x| \geq 1} |x|^q dx = \Omega_{N-1} \int_1^{\infty} r^{q+N-1} dr,$$

from which we see that

$$|x|^q \text{ is integrable near } \begin{cases} 0 & \text{for } q > -N, \\ \infty & \text{for } q < -N. \end{cases}$$

6. Hardy-Littlewood Maximal Theorem

This section takes a first look at the theory of almost-everywhere convergence. The theory developed historically out of Lebesgue's work on an extension of the Fundamental Theorem of Calculus to general integrable functions on intervals of the line, work that we address largely in the next chapter. We shall see gradually that the theory applies to a broader range of problems than the ones immediately generalizing Lebesgue's work, and one can make a case that nowadays the theory in this section is of considerably greater significance in real analysis than one might expect from Lebesgue's work on the Fundamental Theorem.

The theory brings together two threads. The first thread is the observation that an effort to differentiate integrals of general integrable functions on an interval of the line can be reinterpreted as a problem of almost-everywhere convergence in connection with an approximate identity of the kind in Theorem 6.20. In explaining this assertion, let us denote Lebesgue measure by m as necessary. To differentiate $F(x) = \int_a^x f(t) dt$, one forms the usual difference quotient $h^{-1}[F(x+h) - F(x)]$, which can be written for $h > 0$ as

$$\frac{1}{m([-h, 0])} \int_{[-h, 0]} f(x-y) dy = \int_{\mathbb{R}^1} f(x-y) m([-h, 0])^{-1} I_{[-h, 0]}(y) dy$$

or as $f * \varphi_h(x)$, where $\varphi(y) = m([-1, 0])^{-1} I_{[-1, 0]}(y)$. Here φ has integral 1, and φ_h is the normalized dilated function defined in Section 2 by $\varphi_h(y) = h^{-1} \varphi(h^{-1}y)$ in the 1-dimensional case. Theorem 6.20 says for $p = 1$ and $p = 2$ that as h decreases to 0, $f * \varphi_h$ converges to f in L^p if f is in L^p . Also, $f * \varphi_h$ converges uniformly to f if f is bounded and uniformly continuous, and $f * \varphi_h(x)$ converges to $f(x)$ at the point x if f is bounded and is continuous at x . The problem about differentiation of integrals asks about convergence almost everywhere.

We shall want to have a theorem in \mathbb{R}^N , and for this purpose an N -dimensional version of $I_{[-1, 0]}$ does not seem attractive for generalizing. Instead, let us generalize from $I_{[-1, 1]}$, taking the N -dimensional problem to involve a ball B of radius 1 centered at the origin; there is some flexibility in choosing the set B , and a cube centered at the origin would work as well. We write rB for the set of dilates of the members of B by the scalar r . Thus we investigate

$$m(rB)^{-1} \int_{rB} f(x-y) dy$$

as r decreases to 0; equivalently we investigate

$$f * \varphi_r(x), \quad \text{where } \varphi(y) = m(B)^{-1} I_B(y).$$

The second thread comes from making a simple observation and then trying to prove the converse in specific settings, as improbable as it sounds. The observation is that if some sequence of nonnegative functions indexed by n is to converge almost everywhere, its supremum on n must be finite almost everywhere. A converse would say that a finite supremum almost everywhere implies convergence almost everywhere. Banach succeeded in proving an abstract such converse in a 1926 paper, making use of the completeness of the space of functions he was studying. In a celebrated 1930 paper, Hardy and Littlewood proved a concrete such converse in connection with differentiation of integrals; they obtained a quantitative estimate about the supremum, and then the almost everywhere convergence followed from that estimate and from the fact that the convergence certainly takes place for nice functions. Here is an N -dimensional version of the basic theorem in that direction.

Theorem 6.38 (Hardy–Littlewood Maximal Theorem). If f is in $L^1(\mathbb{R}^N)$, then

$$m\{x \mid f^*(x) > \xi\} \leq \frac{5^N \|f\|_1}{\xi}$$

for every $\xi > 0$, where

$$f^*(x) = \sup_{0 < r < \infty} m(rB)^{-1} \int_{rB} |f(x-y)| dy.$$

Before examining the statement of the theorem more closely and then proving the theorem, let us see how to derive a corresponding N -dimensional convergence result from it, and let us see how the first part of Lebesgue's version of the Fundamental Theorem of Calculus, the part about differentiation of integrals, follows as well.

Corollary 6.39. If f is integrable on every bounded subset of \mathbb{R}^N , then

$$\lim_{r \downarrow 0} m(rB)^{-1} \int_{rB} f(x-y) dy = f(x) \quad \text{a.e.}$$

PROOF. Since the convergence for a particular x depends on the behavior of the function only near x , we may assume that f is identically 0 off some bounded set. The effect of this assumption for our purposes is that f then has to be in $L^1(\mathbb{R}^N)$. Define

$$T_r(f) = m(rB)^{-1} \int_{rB} f(x-y) dy,$$

bearing in mind that $f^*(x) = \sup_{r > 0} T_r(|f|)(x)$. If g is continuous of compact support, then $\lim_{r \downarrow 0} T_r g(x) = g(x)$ everywhere by Theorem 6.20c. Let $\epsilon > 0$ be

given, and choose by Corollary 6.4 a function g in $C_{\text{com}}(\mathbb{R}^N)$ with $\|f - g\|_1 < \epsilon$. Then

$$\begin{aligned} \limsup_{r \downarrow 0} |T_r f(x) - f(x)| &\leq \limsup_{r \downarrow 0} |T_r(f - g)(x)| + \limsup_{r \downarrow 0} |T_r g(x) - g(x)| + |g(x) - f(x)| \\ &\leq \sup_{r > 0} |T_r(f - g)(x)| + |g(x) - f(x)| \\ &\leq \sup_{r > 0} T_r(|f - g|)(x) + |g(x) - f(x)|. \end{aligned}$$

If the left side is $> \xi$, at least one of the terms on the right side is $> \xi/2$. Hence

$$\begin{aligned} \{x \mid \limsup_{r \downarrow 0} |T_r f(x) - f(x)| > \xi\} \\ \subseteq \{x \mid (f - g)^*(x) > \xi/2\} \cup \{x \mid |f(x) - g(x)| > \xi/2\}. \end{aligned}$$

By Theorem 6.38 and the inequality $\alpha m\{x \mid |F(x)| > \alpha\} \leq \|F\|_1$, the Lebesgue measure of the right side is

$$\leq \frac{2 \cdot 5^N \|f - g\|_1}{\xi} + \frac{2\|f - g\|_1}{\xi} \leq \epsilon \frac{2(5^N + 1)}{\xi}.$$

Since ϵ is arbitrary, $S(\xi) = \{x \mid \limsup_{r \downarrow 0} |T_r f(x) - f(x)| > \xi\}$ has measure 0. Letting ξ tend to 0 through the values $1/n$, we see that $S(0)$ has measure 0, i.e., that $\lim_{r \downarrow 0} T_r f(x) = f(x)$ almost everywhere. \square

Corollary 6.40 (first part of Lebesgue's form of the Fundamental Theorem of Calculus). If f is integrable on every bounded subset of \mathbb{R}^1 , then $\int_a^x f(y) dy$ is differentiable almost everywhere and

$$\frac{d}{dx} \int_a^x f(y) dy = f(x) \quad \text{a.e.}$$

PROOF. For f in $L^1(\mathbb{R}^1)$, let f^* be as in Theorem 6.38, and define

$$f_r^{**}(x) = \sup_{h>0} \frac{1}{h} \int_0^h |f(x+t)| dt \quad \text{and} \quad f_l^{**}(x) = \sup_{h>0} \frac{1}{h} \int_{-h}^0 |f(x+t)| dt.$$

Then

$$f_r^{**}(x) \leq \sup_{h>0} \frac{1}{h} \int_{-h}^h |f(x+t)| dt = 2f^*(x),$$

and similarly $f_l^{**}(x) \leq 2f^*(x)$. From Theorem 6.38 it follows that

$$m\{x \mid f_r^{**}(x) > \xi\} \leq 10\|f\|_1/\xi$$

and

$$m\{x \mid f_l^{**}(x) > \xi\} \leq 10\|f\|_1/\xi.$$

The same argument as for Corollary 6.39 allows us to conclude, for any f integrable on every bounded subset of \mathbb{R}^1 , that $\lim_{h \downarrow 0} \frac{1}{h} \int_0^h f(x+t) dt = f(x)$ a.e. and $\lim_{h \downarrow 0} \frac{1}{h} \int_{-h}^0 f(x+t) dt = f(x)$ a.e. Hence $\frac{d}{dx} \int_a^x f(t) dt = f(x)$ almost everywhere for such f . \square

Let us return to Theorem 6.38. The function $f^*(x)$ is called the **Hardy–Littlewood maximal function** of f . It is measurable because the supremum over rational r gives the same value of $f^*(x)$ for each x . If we let ξ tend to ∞ in the inequality $m\{x \mid f^*(x) > \xi\} \leq 5^N \|f\|_1/\xi$, we see immediately that $f^*(x)$ is finite almost everywhere, i.e., that the supremum in question is actually finite almost everywhere. The inequality is a quantitative version of that qualitative conclusion.

For any situation in which it is desired to prove an almost-everywhere convergence theorem, there is an associated **maximal function** in modern terminology, which can be taken as the supremum of the absolute value of the quantity for which one is trying to prove almost-everywhere convergence. In the above case we used the supremum for $|f|$ instead, which in principle could be larger.

There is no hope that the Hardy–Littlewood maximal function f^* is actually in L^1 if f is not a.e. the 0 function because the occurrence of large values of r in the supremum already rules out L^1 behavior: in fact, $f^*(x)$ is necessarily \geq a positive multiple of $|x|^{-N}$ for large $|x|$, and thus f^* cannot be integrable. On the other hand, f^* is close to integrable: We shall see in Section 10 that the integral of any nonnegative function g can be computed in terms of the function $m\{x \mid g(x) > \xi\}$ of ξ , the formula being $\int_{\mathbb{R}^N} g(x) dx = \int_0^\infty m\{x \mid g(x) > \xi\} d\xi$. Theorem 6.38 shows that the integrand in the case of f^* is \leq a multiple of $1/\xi$, and $1/\xi$ is close to being integrable on $(0, +\infty)$. This is a better qualitative conclusion than merely finiteness almost everywhere, and Theorem 6.38 is a quantitative version of just how close f^* is to being integrable.

The particular property of f^* that is isolated in Theorem 6.38 arises fairly often. If $g \geq 0$ is integrable and S is the set where $g > \xi$, then $g \geq \xi I_S$ everywhere; hence $\|g\|_1 \geq \xi m(S)$ and $m(S) \leq \|g\|_1/\xi$. A function g is said to be in **weak** L^1 if

$$m\{x \mid |g(x)| > \xi\} \leq C/\xi$$

for some constant C and for all $\xi > 0$. Theorem 6.38 says that the nonlinear operator $f \mapsto f^*$ carries L^1 to weak L^1 with C bounded by a multiple of the L^1

norm of f , and an operator of this kind that satisfies also a certain sublinearity property is said to be of **weak type** $(1, 1)$. We return to this matter, with the definition in a clearer context, in Chapter IX.

Now let us prove Theorem 6.38. One modern proof uses the following covering lemma, which takes into account the geometry of \mathbb{R}^N in a surprisingly subtle way. Once the lemma is in hand, the rest is easy.

Lemma 6.41 (Wiener's Covering Lemma). Let $E \subseteq \mathbb{R}^N$ be a Borel set, and suppose that to each x in E there is associated some ball $B(r; x)$ with r perhaps depending on x . If the radii $r = r(x)$ are bounded, then there is a finite or countable *disjoint* collection of these balls, say $B(r_1; x_1), B(r_2; x_2), \dots$, such that either the collection is infinite and $\inf_{1 \leq j < \infty} r_j \neq 0$ or

$$E \subseteq \bigcup_{j=1}^{\infty} B(5r_j; x_j).$$

In either case,

$$m(E) \leq 5^N \sum_{j=1}^{\infty} m(B(r_j; x_j)).$$

REMARK. The shape of the sets of $B(r; x)$ is not very important. What is important is that there be some neighborhood B of the origin that is closed under the operation of multiplying all its members by -1 and by any positive number $r \leq 1$. The other sets are obtained from B by dilation and translation.

PROOF. Let

$$\mathcal{A}_1 = \{\text{all sets } B(r; x) \text{ in question}\}$$

and $R_1 = \sup \{r \mid B(r; x) \text{ is in } \mathcal{A}_1 \text{ for some } x\}$.

By hypothesis, R_1 is finite. Pick some $B(r_1; x_1)$ with $r_1 \geq \frac{1}{2}R_1$, and let

$$\mathcal{A}_2 = \{\text{members of } \mathcal{A}_1 \text{ disjoint from } B(r_1; x_1)\}.$$

If \mathcal{A}_2 is empty, let all further R_j 's be 0 and let all further $B(r_j; x_j)$'s be empty. If \mathcal{A}_2 is nonempty, let

$$R_2 = \sup \{r \mid B(r; x) \text{ is in } \mathcal{A}_2 \text{ for some } x\}.$$

Pick $B(r_2; x_2)$ in \mathcal{A}_2 with $r_2 \geq \frac{1}{2}R_2$. Let

$$\mathcal{A}_3 = \{\text{members of } \mathcal{A}_2 \text{ disjoint from } B(r_2; x_2)\},$$

and proceed inductively to construct $R_3, B(r_3; x_3), \mathcal{A}_4$, etc.

The numbers R_j are monotone decreasing. We may assume that $\lim R_j = 0$, since otherwise $\inf_j r_j \neq 0$ and $\sum m(B(r_j; x_j)) = +\infty$. Let

$$V_j = \text{union of all sets in } \mathcal{A}_j - \mathcal{A}_{j+1} \quad \text{for } j \geq 1$$

and $V_0 = \text{union of all sets in } \mathcal{A}_1$.

Then $V_0 = \bigcup_{j=1}^{\infty} V_j$; in fact, if $B(r; x)$ is in \mathcal{A}_1 , then the equality $\lim R_j = 0$ forces there to be a last index j such that $B(r; x)$ is in \mathcal{A}_j , and this j has the property that $B(r; x)$ is in \mathcal{A}_j and not \mathcal{A}_{j+1} .

Since $E \subseteq \bigcup_{x \in E} B(r; x) = V_0 = \bigcup_{j=1}^{\infty} V_j$, the proof will be complete if we show that

$$V_j \subseteq B(5r_j; x_j). \quad (*)$$

Thus let $B(r; x)$ be in $\mathcal{A}_j - \mathcal{A}_{j+1}$. Then $r \leq R_j$,

$$B(r; x) \cap B(r_j; x_j) \neq \emptyset,$$

and $r_j \geq \frac{1}{2}R_j$. Consequently $r \leq 2r_j$ and

$$B(r; x - x_j) \cap B(r_j; 0) \neq \emptyset.$$

This condition means that there is some p in $B(r_j; 0)$ with $|x - x_j - p| < r$. If q is any member of $B(r; x)$, then

$$|q - x_j| \leq |q - x| + |x - x_j - p| + |p| < r + r + r_j = 2r + r_j.$$

Thus q is in $B(2r + r_j; x_j) \subseteq B(5r_j; x_j)$, and $(*)$ follows. \square

PROOF OF THEOREM 6.38. Let $E = \{x \mid f^*(x) > \xi\}$. If x is in E , then $m(B(r; 0))^{-1} \int_{B(r; x)} |f(y)| dy > \xi$ for some $r > 0$. Associate this r to x in applying Lemma 6.41. Since

$$\xi < m(B(r; 0))^{-1} \int_{B(r; x)} |f(y)| dy \leq r^{-N} m(B(1; 0))^{-1} \|f\|_1,$$

we see that $r^N \leq \xi^{-1} m(B(1; 0))^{-1} \|f\|_1$. Hence the numbers r are bounded. Thus the lemma applies, and we obtain

$$m(E) \leq 5^N \sum_j m(B(r_j; x_j)) \leq 5^N \xi^{-1} \sum_j \int_{B(r_j; x_j)} |f(y)| dy \leq 5^N \xi^{-1} \|f\|_1,$$

the last inequality holding because of the disjointness of the sets $B(r_j; x_j)$. \square

Let us return to the theme of almost-everywhere convergence in connection with approximate identities. Theorem 6.38 has the following consequence of just that kind.

Corollary 6.42. Let $\varphi \geq 0$ be a continuous integrable function on \mathbb{R}^N of the form $\varphi(x) = \varphi_0(|x|)$, where φ_0 is a decreasing C^1 function on $[0, \infty)$, and define $\varphi_\varepsilon(x) = \varepsilon^{-N} \varphi(\varepsilon^{-1}x)$ for $\varepsilon > 0$. Then there is a constant C_φ such that

$$\sup_{\varepsilon > 0} |(\varphi_\varepsilon * f)(x)| \leq C_\varphi f^*(x)$$

for all x in \mathbb{R}^N and for all f in $L^1(\mathbb{R}^N)$. Consequently if $\int_{\mathbb{R}^N} \varphi(x) dx = 1$, then

$$\lim_{\varepsilon \downarrow 0} (\varphi_\varepsilon * f)(x) = f(x)$$

almost everywhere for each f in $L^1(\mathbb{R}^N)$.

PROOF. Put $\psi(r) = -\varphi_0'(r) \geq 0$, so that $\varphi_0(r) - \varphi_0(R) = \int_r^R \psi(s) ds$. The integrability of φ and the fact that φ_0 is decreasing force $\lim_{R \rightarrow \infty} \varphi_0(R) = 0$, and we obtain $\varphi_0(r) = \int_r^\infty \psi(s) ds$ and $\varphi(x) = \int_{|x|}^\infty \psi(r) dr$. Meanwhile, the integrability of φ , together with the formula for integrating in spherical coordinates, shows that $\int_0^\infty \varphi_0(r) r^{N-1} dr = C < +\infty$. Integrating by parts on the interval $[0, M]$ gives

$$C \geq \int_0^M \varphi_0(r) r^{N-1} dr = \frac{1}{N} [\varphi_0(r) r^N]_0^M + \frac{1}{N} \int_0^M \psi(r) r^N dr,$$

and thus

$$\frac{1}{N} \int_0^\infty \psi(r) r^N dr \leq C < +\infty.$$

The form of φ implies that

$$\varphi_\varepsilon(x) = \varepsilon^{-N} \int_{\varepsilon^{-1}|x|}^\infty \psi(r) dr.$$

If, as in the statement of Theorem 6.38, we let B be the ball of radius 1 centered at the origin, we obtain

$$\begin{aligned} |(\varphi_\varepsilon * f)(x)| &\leq \int_{\mathbb{R}^N} \varphi_\varepsilon(y) |f(x-y)| dy \\ &= \int_{y \in \mathbb{R}^N} \varepsilon^{-N} \int_{r=\varepsilon^{-1}|y|}^\infty \psi(r) |f(x-y)| dr dy \\ &= \int_{r=0}^\infty \psi(r) \left[\varepsilon^{-N} \int_{|y| \leq \varepsilon r} |f(x-y)| dy \right] dr \\ &= \int_{r=0}^\infty m(B) \psi(r) r^N \left[m(\varepsilon r B)^{-1} \int_{|y| \leq \varepsilon r} |f(x-y)| dy \right] dr \\ &\leq m(B) \left[\int_{r=0}^\infty \psi(r) r^N dr \right] f^*(x). \end{aligned}$$

The right side is $\leq C_\varphi f^*(x)$ with $C_\varphi = CNm(B)$. Applying Theorem 6.38, we see that the operator $f \mapsto \sup_{\varepsilon > 0} |(\varphi_\varepsilon * f)(x)|$ is of weak type $(1,1)$. Since $\varphi_\varepsilon * f$ converges pointwise (and in fact uniformly) to f when f is in $C_{\text{com}}(\mathbb{R}^N)$, the same argument as for Corollary 6.39 shows that $\lim_{\varepsilon \downarrow 0} (\varphi_\varepsilon * f)(x) = f(x)$ almost everywhere for each f in $L^1(\mathbb{R}^N)$. \square

EXAMPLE. An example of a function φ as in Corollary 6.42 is $P(x) = \frac{1}{1+x^2}$ in \mathbb{R}^1 . We shall see in Chapter VIII that the function $h(x, y) = (P_y * f)(x)$ for this φ is the natural function on the half plane $y > 0$ in \mathbb{R}^2 that is **harmonic**, i.e., has $\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} = 0$, and has boundary value f . Corollary 6.42 says that $h(x, y)$ has $f(x)$ as boundary values almost everywhere if f is in $L^1(\mathbb{R}^1)$.

7. Fourier Series and the Riesz–Fischer Theorem

As mentioned at the beginning of Chapter V, the use of the Riemann integral imposes some limitations on the subject of Fourier series that no longer apply when one uses the Lebesgue integral. In this section we shall redo the elementary theory of Fourier series of Section I.10 with the Lebesgue integral in place, with particular attention to the improved theorems that we obtain. It will be assumed that the reader knows the theory of that section.

The underlying measure space will be $[-\pi, \pi]$ with the σ -algebra of Borel sets and with the measure $\frac{1}{2\pi} dx$, where dx is 1-dimensional Lebesgue measure. The complex-valued functions under consideration will be periodic of period 2π , thus assuming the same value at π as at $-\pi$. The spaces L^1 , L^2 , and L^∞ will refer to this measure space when no other parameters are given. Since the measure of the whole space is finite, these spaces satisfy the inclusions $L^\infty \subseteq L^2 \subseteq L^1$. The functions in L^∞ being essentially bounded, they are certainly integrable and square integrable. The inclusion $L^2 \subseteq L^1$ follows from the Schwarz inequality: $\frac{1}{2\pi} \int_{-\pi}^{\pi} |f| dx \leq \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f|^2 dx\right)^{1/2} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} 1 dx\right)^{1/2}$.

There is another way of viewing this measure space that will be especially helpful in relating convolution to the theory. Namely, a periodic function on the line of period 2π may be viewed as a function on the unit circle of \mathbb{C} with the angle as parameter. In fact, convolution is a construction that combines group theory with measure theory when the measure is invariant under the group, and that is why convolution appears more natural on the circle than on $[-\pi, \pi]$. The limits of integration do not have to be written differently from the way they are written on the line, but we must remember that functions are to be extended periodically when we interpret integrands. The factor $\frac{1}{2\pi}$ in front of the measure means that all convolutions of functions are to contain this factor. Thus the definition of **convolution** for nonnegative f and g is

$$(f * g)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y)g(y) dy.$$

Convolution is commutative and associative on the circle just as in Proposition 6.13, and the various norm estimates of Section 2 are valid in the setting of the

circle. The use of dilations has no analog for the circle, and thus the circle has no approximate identities of the form φ_ε .

If f is in L^1 , the trigonometric series

$$\sum_{n=-\infty}^{\infty} c_n e^{inx} \quad \text{with} \quad c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx$$

is called the **Fourier series** of f . This time we regard the integral as a Lebesgue integral. We write

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx} \quad \text{and} \quad s_N(f; x) = \sum_{n=-N}^N c_n e^{inx}.$$

A Fourier series can be written also with cosines and sines, and the coefficients a_n and b_n are unchanged from Section I.10.

Theorem 6.43. Let f be in L^2 . Among all choices of d_{-N}, \dots, d_N , the expression

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| f(x) - \sum_{n=-N}^N d_n e^{inx} \right|^2 dx$$

is minimized uniquely by choosing d_n , for all n with $|n| \leq N$, to be the Fourier coefficient $c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx$. The minimum value is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx - \sum_{n=-N}^N |c_n|^2.$$

PROOF. The proof is the same as for Theorem 1.53. □

Corollary 6.44 (Bessel's inequality). If f is in L^2 with $f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$, then

$$\sum_{n=-\infty}^{\infty} |c_n|^2 \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx.$$

In particular, $\sum_{n=-\infty}^{\infty} |c_n|^2$ is finite.

PROOF. The proof is the same as for Corollary 1.54. □

Corollary 6.45 (Riemann–Lebesgue Lemma). If f is in L^1 and has Fourier coefficients $\{c_n\}_{n=-\infty}^{\infty}$, then $\lim_{|n| \rightarrow \infty} c_n = 0$.

REMARK. Since L^2 is properly contained in L^1 , this corollary is not a special case of Corollary 6.44, unlike the situation with the Riemann integral.

PROOF. The result is immediate from Corollary 6.44 in the case of L^2 functions and in particular in the case of continuous functions. Write $c_n(h)$ for the n^{th} Fourier coefficient of any function h . Let $\epsilon > 0$ be given. Choose by Corollary 6.27a a continuous g with $\|f - g\|_1 \leq \epsilon/2$. Then choose N such that $|n| \geq N$ implies $|c_n(g)| \leq \epsilon/2$. Then $|c_n(f)| \leq |c_n(f - g)| + |c_n(g)| \leq \epsilon$ since $|c_n(f - g)| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f - g| |e^{-inx}| dx = \|f - g\|_1 \leq \epsilon/2$. \square

Theorem 6.46 (Dini's test). Let f be in L^1 , and fix x in $[-\pi, \pi]$. If there is a constant $\delta > 0$ such that

$$\int_{|t| < \delta} \frac{|f(x+t) - f(x)|}{|t|} dt < \infty,$$

then $\lim_N s_n(f; x) = f(x)$.

REMARK. The condition in the corresponding result for the Riemann integral, namely Theorem 1.57, was that $|f(x+t) - f(x)| \leq M|t|$ for $|t| < \delta$ and some constant M . The condition in the present theorem is satisfied by $f(x) = \sqrt{|x|}$ at $x = 0$, and the condition in the earlier theorem is not.

PROOF. The proof is the same as for Theorem 1.57 except that we need to appeal to the improved version of the Riemann–Lebesgue Lemma in Corollary 6.45. \square

Now we work toward a proof of Parseval's Theorem for all of L^2 . We need to know about Fourier coefficients of convolutions.

Proposition 6.47. If $f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$ and $g(x) \sim \sum_{n=-\infty}^{\infty} d_n e^{inx}$, then $(f * g)(x) \sim \sum_{n=-\infty}^{\infty} c_n d_n e^{inx}$.

PROOF. This is a consequence of Fubini's Theorem and the translation invariance of Lebesgue measure:

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} (f * g)(x) e^{-inx} dx &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y) g(y) e^{-inx} dy \right] dx \\ &= \left(\frac{1}{2\pi} \right)^2 \int_{-\pi}^{\pi} \left[\int_{-\pi}^{\pi} f(x-y) g(y) e^{-inx} dx \right] dy \\ &= \left(\frac{1}{2\pi} \right)^2 \int_{-\pi}^{\pi} \left[\int_{-\pi}^{\pi} f(x) g(y) e^{-in(x+y)} dx \right] dy \\ &= \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx \right) \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} g(y) e^{-iny} dy \right). \quad \square \end{aligned}$$

The proof of the version of Parseval's Theorem for all of L^2 will make use of the Fejér kernel $K_N(t)$ introduced in Section I.10. We do not need to recall the exact formula for K_N , only the fact that it is a trigonometric polynomial of degree N with the following three properties:

- (i) $K_N(x) \geq 0$,
- (ii) $\frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(x) dx = 1$,
- (iii) for any $\delta > 0$, $\sup_{\delta \leq |x| \leq \pi} K_N(x)$ tends to 0 as n tends to infinity.

These three properties identified K_N as an approximate identity in the setting of periodic functions, and Fejér's Theorem in the form of Theorem 1.59 gave the consequence for convergence at points of continuity of f . With the Lebesgue integral, we get also results about norm convergence in L^1 and L^2 .

Theorem 6.48 (Fejér's Theorem). Let f be in L^1 . Then

- (a) $\lim_N \|K_N * f - f\|_1 = 0$ with no additional hypotheses on f ,
- (b) $\lim_N \|K_N * f - f\|_2 = 0$ if f is also in L^2 ,
- (c) $\lim_N (K_N * f)(x_0) = f(x_0)$ if f is bounded on $[-\pi, \pi]$ and is continuous at x_0 ,
- (d) the convergence in (c) is uniform for x_0 in E if f is bounded on $[-\pi, \pi]$ and is uniformly continuous at the points of E ,
- (e) $\lim_N (K_N * f)(x_0) = \frac{1}{2}(f(x_0+) + f(x_0-))$ if f is bounded on $[-\pi, \pi]$ and has right and left limits $f(x_0+)$ and $f(x_0-)$ at x_0 .

PROOF. For (a) and (b), let $p = 1$ or $p = 2$ as appropriate. Then

$$\begin{aligned} \|K_n * f - f\|_p &= \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(t) [f(x-t) - f(x)] dt \right\|_{p,x} && \text{by (ii)} \\ &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(t) \|f(x-t) - f(x)\|_{p,x} dt && \text{by (i) and} \\ &\leq \sup_{|t| \leq \delta} \|f(x-t) - f(x)\|_{p,x} + 2 \left[\sup_{\delta \leq |t| \leq \pi} K_N(t) \right] \|f\|_p. && \text{Theorem 5.60} \end{aligned}$$

Given $\epsilon > 0$, choose δ by Proposition 6.16 to make the first term of the final bound be $< \epsilon/2$, and then choose N_0 by (iii) to make the second term of the final bound be $< \epsilon/2$ for $N \geq N_0$. Then the final bound is $< \epsilon$ for $N \geq N_0$.

Parts (c) and (d) are proved exactly as in Theorem 1.59. For (e), we may assume without loss of generality that $x_0 = 0$ because convolution commutes with translations. If we can prove (e) for a single function g with a jump discontinuity

at $x = 0$ equal to the jump for f , then we can apply (c) to $f - g$ and deduce (e) for f . Let us see that such a function g may be taken as a multiple of

$$h(x) = \begin{cases} \frac{1}{2}(\pi - x) & \text{for } 0 < x \leq \pi \\ \frac{1}{2}(-\pi - x) & \text{for } -\pi \leq x < 0. \end{cases}$$

In fact, a computation at the beginning of Section I.10 shows explicitly that the series $\sum_{n=1}^{\infty} (\sin nx)/n$ converges to $h(x)$ for $x \neq 0$, but we do not need this fact. All that we need is that the series $\sum_{n=1}^{\infty} (\sin nx)/n$ is the Fourier series of h , a fact that we can readily check from the definition. The sum of this series at $x = 0$ is manifestly 0, and this sum matches the average of the jumps $\frac{1}{2}(\frac{\pi}{2} + \frac{-\pi}{2})$. The Cesàro sums of the series $\sum_{n=1}^{\infty} (\sin nx)/n$ must have the same limit 0, according to Theorem 1.47, and (e) is proved. \square

Theorem 6.49 (Parseval's Theorem). If f is a function in L^2 with $f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$, then

$$\lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - s_N(f; x)|^2 dx = 0$$

and

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx = \sum_{n=-\infty}^{\infty} |c_n|^2.$$

PROOF. From the first conclusion of Theorem 6.43, we obtain $0 \leq \|f - s_N\|_2^2 \leq \|f - (K_N * f)\|_2^2$, and we know from Theorem 6.48b that $\|f - (K_N * f)\|_2^2$ tends to 0. This proves the first formula, and the second formula follows by passing to the limit in the second conclusion of Theorem 6.43. \square

Corollary 6.50 (uniqueness theorem). If f is in L^1 and has all Fourier coefficients 0, then f is the 0 element in L^1 .

PROOF. Proposition 6.47 shows that the Fourier coefficients of $K_N * f$ are $c_n(K_N * f) = c_n(K_N)c_n(f)$, and this is 0 for all n . By Proposition 6.18, $K_N * f$ is continuous, and thus $K_N * f = 0$ by Corollary 1.60. Since $K_N * f$ tends to f in L^1 according to Theorem 6.48a, we conclude that f is the 0 element in L^1 . \square

Now we come to the Riesz–Fischer Theorem, which historically was a great triumph for the Lebesgue integral over the Riemann integral. The result uses the completeness of L^2 and has no counterpart with Riemann integration.

Theorem 6.51 (Riesz–Fischer Theorem). If $\{c_n\}$ is a given doubly infinite sequence of complex numbers with $\sum_{n=-\infty}^{\infty} |c_n|^2 < \infty$, then there exists an f in L^2 whose Fourier series is $\sum_{n=-\infty}^{\infty} c_n e^{inx}$.

PROOF. Define $f_N(x) = \sum_{|n| \leq N} c_n e^{inx}$. For $M \geq N$, Parseval's Theorem (Theorem 6.49) gives $\|f_M - f_N\|_2^2 = \sum_{N+1 \leq |n| \leq M} |c_n|^2$, and the right side tends to 0 as M and N tend to infinity because of the convergence of $\sum_{n=-\infty}^{\infty} |c_n|^2$. Thus $\{f_N\}$ is a Cauchy sequence in L^2 . By Theorem 5.59, L^2 is complete as a metric space, and thus $\{f_N\}$ converges in L^2 . Let f be (a function representing) the limit element in L^2 . The inner product in L^2 is a continuous function of the L^2 function in the first variable, and therefore the Fourier coefficients of f satisfy

$$c_n(f) = (f, e^{inx}) = \lim_N (f_N, e^{inx}).$$

As soon as N gets to be $\geq |n|$, (f_N, e^{inx}) equals c_n . Thus $c_n(f) = c_n$ for all n , and f has the required properties. \square

8. Stieltjes Measures on the Line

A **Stieltjes measure**² is a Borel measure on \mathbb{R}^1 . Lebesgue measure dx is an example, as is any measure $f(x) dx$ in which f is nonnegative and Borel measurable and is integrable on every bounded interval. A completely different kind of Stieltjes measure is one that attaches nonnegative weights to countably many points in such a way that the sum of the weights in any bounded interval is finite. In this section we shall see that the Stieltjes measures stand in one-one correspondence with a class of monotone functions on the line that we describe shortly. We shall also obtain an integration-by-parts formula in which a Stieltjes measure plays the role of the derivative of its corresponding monotone function.

If a Stieltjes measure μ is given, we associate to μ the function $F : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ defined by

$$F(x) = \begin{cases} -\mu(x, 0] & \text{if } x \leq 0, \\ \mu(0, x] & \text{if } x \geq 0. \end{cases}$$

The function F is called the **distribution function** of μ . It has the following properties:³

- (i) F is nondecreasing, i.e., is **monotone increasing**.
- (ii) F is **continuous from the right** in the sense that $F(x_0) = \lim_{x \downarrow x_0} F(x)$ for every x_0 in \mathbb{R}^1 , i.e., $\lim_n F(x_n) = F(x_0)$ whenever $\{x_n\}_{n \geq 1}$ is a sequence tending to x_0 such that $x_n > x_0$ for all $n \geq 1$,
- (iii) $F(0) = 0$.

²Many books, this one included, take Stieltjes measures by definition to occur on the line. However, there is a theory, albeit a somewhat unsatisfactory one, of "Stieltjes measures" in higher-dimensional Euclidean space. It is of interest chiefly in probability theory.

³An alternative definition says $F(x)$ equals $-\mu[x, 0)$ and $\mu[0, x)$ in the two cases, and then property (ii) says that F is **continuous from the left**. The choice made here between these alternatives is governed by keeping technicalities to a minimum in Section 10.

Properties (i) and (iii) are immediate from the definition. With (ii), there are two cases according as the limit x_0 is ≤ 0 or > 0 , and both cases are settled by the complete additivity of μ .

The measure μ is completely determined by its distribution function F . In fact, the definition of F forces $\mu((a, b]) = F(b) - F(a)$, and Proposition 6.6 implies that μ is determined as a Borel measure by this formula.

Theorem 6.52. The Stieltjes measures μ stand in one-one correspondence with the functions $F : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ satisfying (i), (ii), and (iii), the correspondence being that F is the distribution function of μ .

PROOF. We have seen that each μ leads to an F and that F uniquely determines μ . We need to see that every F satisfying (i), (ii), and (iii) arises from some μ . If such a function F is given, we define a set function μ on bounded intervals by

$$\begin{aligned}\mu((a, b]) &= F(b) - F(a), \\ \mu((a, b)) &= \lim_n F(b - \frac{1}{n}) - F(a), \\ \mu([a, b]) &= F(b) - \lim_n F(a - \frac{1}{n}), \\ \mu([a, b)) &= \lim_n F(b - \frac{1}{n}) - \lim_n F(a - \frac{1}{n}).\end{aligned}$$

We extend μ to the ring \mathcal{R} of elementary subsets of \mathbb{R}^1 , i.e., the ring of all finite disjoint unions of bounded intervals, by setting μ of a finite disjoint union of bounded intervals equal to the sum of the values of μ on each of the intervals, just as with Lebesgue measure in Example 4 at the end of Section V.1.

To see that μ is unambiguously defined and is additive on \mathcal{R} , we readily reduce matters, just as with Lebesgue measure, to showing that if an interval is decomposed into the union of two smaller intervals, then μ of the union is the sum of μ of the components. Thus let $a \leq b \leq c$, and let an interval I from a to c be the union of an interval from a to b and an interval from b to c . If the interval I from a to c is (a, c) , then the two possible cases are handled by

$$\begin{aligned}\mu((a, b)) + \mu((b, c)) &= \lim_n F(b - \frac{1}{n}) - F(a) + \lim_n F(c - \frac{1}{n}) - \lim_n F(b - \frac{1}{n}) \\ &= \mu((a, c))\end{aligned}$$

and

$$\mu((a, b]) + \mu((b, c)) = F(b) - F(a) + \lim_n F(c - \frac{1}{n}) - F(b) = \mu((a, c)).$$

If the interval I from a to c has one or both endpoints present, then the computation is the same except that $F(a)$ is replaced by $\lim_n F(a - \frac{1}{n})$ if a is in I and

$\lim_n F(c - \frac{1}{n})$ is replaced by $F(c)$ if c is in I . Thus μ is unambiguously defined, and it is nonnegative and additive.

The next step is to prove, just as with Lebesgue measure in Section V.1, that μ is regular on \mathcal{R} in the sense that for each E in \mathcal{R} and $\epsilon > 0$, we can find a compact K in \mathcal{R} and an open U in \mathcal{R} such that $K \subseteq E \subseteq U$, $m(K) \geq m(E) - \epsilon$, and $m(U) \leq m(E) + \epsilon$. As with Lebesgue measure, the proof comes down to the case that E is a single interval, and this time there are four subcases. Choose n large enough so that $\frac{2}{n} < \epsilon$, and then

$$\begin{aligned} \text{for } [a, b), \quad & \text{take } K = [a, b - \frac{1}{n}] \text{ and } U = (a - \frac{1}{n}, b), \\ \text{for } [a, b], \quad & \text{take } K = [a, b] \text{ and } U = (a - \frac{1}{n}, b + \frac{1}{n}), \\ \text{for } (a, b], \quad & \text{take } K = [a + \frac{1}{n}, b] \text{ and } U = (a, b + \frac{1}{n}), \\ \text{for } (a, b), \quad & \text{take } K = [a + \frac{1}{n}, b - \frac{1}{n}] \text{ and } U = (a, b). \end{aligned}$$

An exception occurs in the definition of K if the listed left endpoint of K exceeds the listed right endpoint of K , and then K is defined to be empty. Each of these definitions contains a parameter n ; if we write K_n and U_n for the corresponding sets K and U , then we can check from the definitions and property (ii) of the function F that $\lim_n \mu(K_n) = \mu(E)$ and $\lim_n \mu(U_n) = \mu(E)$. The regularity condition for E follows from these limit relations.

The next step is to prove that μ is completely additive on \mathcal{R} by imitating the proof for Lebesgue measure. In fact, the proof of Proposition 5.4 applies word-for-word except that m has to be changed to μ throughout and the word “proposition” in the last sentence of the proof should be changed to “complete additivity.”

Then μ extends to a measure on the Borel sets by Theorem 5.5. The extended measure is σ -finite on \mathbb{R}^1 because \mathbb{R}^1 is the countable union of bounded intervals and μ is finite on every bounded interval.

Finally we need to show that the distribution function G of μ is equal to F . Our definitions make

$$G(x) = \begin{cases} -\mu((x, 0]) = -(F(0) - F(x)) = F(x) & \text{if } x \leq 0, \\ \mu((0, x]) = F(x) - F(0) = F(x) & \text{if } x \geq 0. \end{cases}$$

Thus $G = F$. □

EXAMPLES.

(1) Let F be any continuous distribution function that has a continuous derivative f except possibly at finitely many points. If x is a point of continuity of f , then the Fundamental Theorem of Calculus (Theorem 1.32) gives

$\frac{d}{dx} \int_0^x f(t) dt = f(x)$. Put

$$G(x) = \begin{cases} -\int_x^0 f(t) dt & \text{if } x \leq 0, \\ \int_0^x f(t) dt & \text{if } x \geq 0. \end{cases}$$

Then G is a continuous distribution function, and the formula for the derivative of the integral shows that $F'(x) = G'(x)$ except at finitely many points. Recursive application of the Mean Value Theorem, starting from $x = 0$, to $F - G$ on intervals having $F' - G' = 0$ in their interiors, shows that $F = G$ everywhere. The Stieltjes measure μ associated to F , by the uniqueness in Theorem 6.52, is given by

$$\mu(E) = \int_E f(t) dt.$$

The special case with $F(x)$ identically equal to x has f identically equal to 1, and the measure is just Lebesgue measure.

(2) The function F with

$$F(x) = \begin{cases} 0 & \text{for } x \geq 0, \\ -1 & \text{for } x < 0, \end{cases}$$

has the three properties of a distribution function, and the associated measure μ is a point mass assigning weight 1 to $x = 0$. The measure μ takes the value 1 on every Borel set containing 0 and takes the value 0 on every Borel set not containing 0. This measure is sometimes called the **delta measure** at 0 or “delta mass” at 0. Whenever a Stieltjes measure ν has $\nu(\{p\}) > 0$ for some p in \mathbb{R}^1 , we say that ν contains a **point mass** at p of weight $\nu(\{p\})$. Then ν is the sum of a point mass at p of weight $\nu(\{p\})$ and a Stieltjes measure containing no point mass at p .

(3) Let $\{x_n\}$ be a sequence in \mathbb{R} . For example, $\{x_n\}$ could be an enumeration of the rationals. Let $\{w_n\}$ be a sequence of positive numbers with $\sum w_n < \infty$, and define

$$F(x) = \begin{cases} \sum_{\{n \mid 0 < x_n \leq x\}} w_n & \text{for } x \geq 0, \\ -\sum_{\{n \mid x < x_n \leq 0\}} w_n & \text{for } x < 0. \end{cases}$$

Then F satisfies (i), (ii), and (iii), and hence F is the distribution function of some Stieltjes measure μ . The measure is given by

$$\mu((a, b]) = \sum_{\{n \mid a < x_n \leq b\}} w_n.$$

It is a countable sum of point masses. The function F , though monotone increasing, is discontinuous at every x_n , and this set is allowed to be dense.

(4) This example will be a nonzero Stieltjes measure that is carried on a Borel set of Lebesgue measure 0 and yet has a continuous distribution function. We start from the standard Cantor set C in $[0, 1]$ described in detail in Section II.9. This set is compact and is obtained as the intersection of a sequence $\{C_n\}$ of sets with each C_n consisting of the finite union of closed bounded intervals. The set C_0 is $[0, 1]$, and C_{n+1} is obtained from C_n by removing the open middle third of each of the

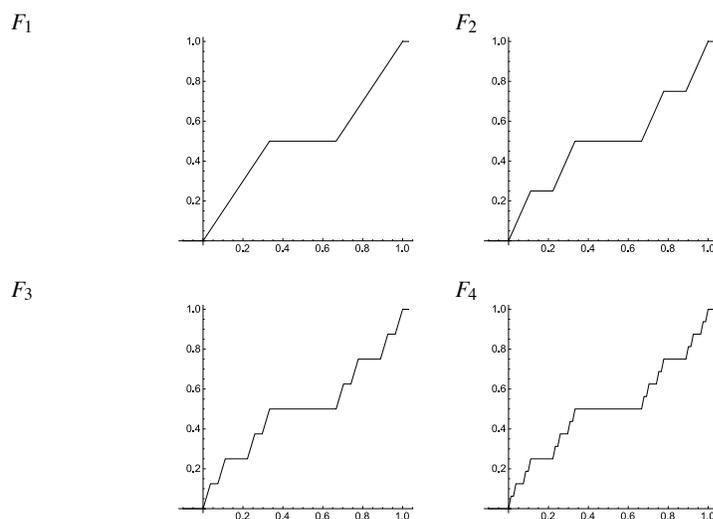


FIGURE 6.1. Construction of a Cantor function F . Graphs of approximations F_1, F_2, F_3, F_4 to F .

constituent closed intervals of C_n . The Lebesgue measure of C_n is $(2/3)^n$, and thus C has Lebesgue measure $\lim_n (2/3)^n = 0$. The measure μ we construct will have $\mu(C) = 1$ and $\mu(C^c) = 0$, yet it will assign 0 measure to every one-point set. The properties that are needed of the corresponding distribution function F so that μ has these properties are that F is continuous, F is 0 for $x \leq 0$, F is 1 for $x \geq 1$, and F is constant on every open interval I of $[0, 1] - C$, i.e., on every open interval of every $[0, 1] - C_n$. This condition will make $\mu(I) = 0$ for all such I . Since the metric space $[0, 1] - C$ has a countable base, it is the union of countably many such open intervals I , and thus $\mu([0, 1] - C) = 0$. Since F is constant for $x \leq 0$ and for $x \geq 1$, μ is 0 on $(-\infty, 0)$ and $(1, +\infty)$ as well, and thus $\mu(C^c) = 0$. To obtain the distribution function F , we construct a sequence of approximating functions F_n and show that the sequence is uniformly Cauchy. The set $C_n^c \cap [0, 1]$ is the union of $2^n - 1$ disjoint open intervals. On the k^{th} such interval we define F_n to be $k2^{-n}$. We let $F_n(x) = 0$ for $x \leq 0$ and $F_n(x) = 1$

for $x \geq 1$. On the complementary closed intervals, define F_n in any fashion that makes F_n monotone increasing and continuous. Graphs of F_1 through F_4 are shown in Figure 6.1 with the interpolation in the graphs done by straight lines. The result is that

$$|F_n(x) - F_{n+k}(x)| \leq 2^{-n}$$

for all x . Hence $\{F_n\}$ is uniformly Cauchy and therefore uniformly convergent. Let F be the limit function. The function F is continuous by Theorem 1.21, and it is monotone increasing, satisfies $F(0) = 0$, and is constant on every open interval contained in C^c . According to Problem 15 at the end of the chapter, it is independent of the method of interpolation used in constructing the F_n 's. The function F is called the **Cantor function** corresponding to the standard Cantor set.

The most general monotone increasing function F on \mathbb{R}^1 is not far from being the distribution function of some Stieltjes measure. In the first place the monotonicity of F implies that F has left and right limits at every point, and consequently its only discontinuities are jumps. There can be only countably many such jumps: in fact, if there were uncountably many jump discontinuities, there would be uncountably many in some bounded interval, and that interval would contain uncountably many of magnitude at least $1/n$ for some integer n ; hence F would have to be unbounded on that bounded interval. Let us define a function F_1 by $F_1(x) = \lim_{t \downarrow x} F(t)$. This is well defined, since F has right limits at every point, and we have $F(x) = F_1(x)$ except on a countable set. If we define $F_2(x) = F_1(x) - F_1(0)$, then F_2 satisfies the three defining properties (i), (ii), (iii) of a distribution function. If μ is the Stieltjes measure corresponding to F_2 under Theorem 6.52, we call μ the **associated Stieltjes measure** for F .

Theorem 6.53 (integration by parts). Let $a < b$, let F be a monotone increasing function on \mathbb{R}^1 that is continuous from the right at a and b , and let μ be the associated Stieltjes measure. If G is a C^1 complex-valued function on $[a, b]$ with derivative g , then

$$\int_a^b F(x)g(x) dx = G(b)F(b) - G(a)F(a) - \int_{(a,b]} G d\mu.$$

PROOF. Without loss of generality, we may assume that G is real-valued. Let F_2 be the distribution function of μ . By construction of F_2 , there is a constant c such that $F - F_2 = c$ except possibly at points of discontinuity of F , and the set S of such points within $[a, b]$ is countable. This exceptional countable set S does not contain a or b , since F and F_2 are continuous from the right at a and b .

We have

$$\begin{aligned}\int_a^b (F - F_2)g \, dx &= \int_S (F - F_2)g \, dx + \int_a^b c g(x) \, dx \\ &= c \int_a^b g(x) \, dx = c(G(b) - G(a))\end{aligned}$$

and also

$$G(b)(F(b) - F_2(b)) - G(a)(F(a) - F_2(a)) = G(b)c - G(a)c = c(G(b) - G(a)).$$

Thus

$$\int_a^b (F - F_2)g \, dx = G(b)(F(b) - F_2(b)) - G(a)(F(a) - F_2(a)).$$

Comparing this formula with the formula in the statement of the theorem, we see that if the theorem holds for F_2 , then it holds for F . Changing notation, we may therefore assume that F is the distribution function of μ .

Let P be a partition $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ of $[a, b]$ with mesh to be specified. For $1 \leq i \leq n$, we use the Mean Value Theorem to choose $t_i \in (x_{i-1}, x_i)$ with $G(x_i) - G(x_{i-1}) = g(t_i)(x_i - x_{i-1})$. We can do so since we have assumed that G is real valued. Then we have

$$F(x_i)g(t_i)(x_i - x_{i-1}) = F(x_i)G(x_i) - F(x_i)G(x_{i-1})$$

and

$$\begin{aligned}\sum_{i=1}^n F(x_i)g(t_i)(x_i - x_{i-1}) &= F(x_n)G(x_n) + \sum_{i=2}^n F(x_{i-1})G(x_{i-1}) - \sum_{i=1}^n F(x_i)G(x_{i-1}) \\ &= F(x_n)G(x_n) - F(x_0)G(x_0) - \sum_{i=1}^n G(x_{i-1})(F(x_i) - F(x_{i-1})) \\ &= F(b)G(b) - F(a)G(a) - \sum_{i=1}^n G(x_{i-1})(F(x_i) - F(x_{i-1})).\end{aligned}$$

We shall show for small enough mesh that $\sum_{i=1}^n F(x_i)g(t_i)(x_i - x_{i-1})$ is close to $\int_a^b F(x)g(x) \, dx$ and that $\sum_{i=1}^n G(x_{i-1})(F(x_i) - F(x_{i-1}))$ is close to $\int_{(a,b]} G \, d\mu$.

Let M be an upper bound for $|g|$ and $|F|$ on $[a, b]$, and let $\epsilon > 0$ be given. Choose a number $\delta > 0$ by uniform continuity of G and g such that $|x - x'| < \delta$ implies $|G(x) - G(x')| < \epsilon/M$ and $|g(x) - g(x')| < \epsilon/(M(b-a))$, as well as another condition to be specified. If the mesh of the partition is $< \delta$, then

$$\begin{aligned}
& \left| \sum_{i=1}^n G(x_{i-1})(F(x_i) - F(x_{i-1})) - \int_{(a,b]} G \, d\mu \right| \\
&= \left| \sum_{i=1}^n \int_{(x_{i-1}, x_i]} [G(x_{i-1}) - G(x)] \, d\mu \right| \\
&\leq \sum_{i=1}^n \int_{(x_{i-1}, x_i]} |G(x_{i-1}) - G(x)| \, d\mu \\
&\leq \sum_{i=1}^n \int_{(x_{i-1}, x_i]} (\epsilon/M) \, d\mu \\
&= (\epsilon/M)(F(b) - F(a)) \\
&\leq 2\epsilon \qquad \text{since } |F| \leq M.
\end{aligned}$$

Also,

$$\begin{aligned}
& \left| \sum_{i=1}^n F(x_i)g(t_i)(x_i - x_{i-1}) - \int_a^b F(x)g(x) \, dx \right| \\
&= \left| \sum_{i=1}^n \int_{(x_{i-1}, x_i]} (F(x_i)(g(t_i) - g(x)) + (F(x_i) - F(x))g(x)) \, dx \right| \\
&\leq \sum_{i=1}^n \int_{(x_{i-1}, x_i]} |F(x_i)| |g(t_i) - g(x)| \, dx + \sum_{i=1}^n \int_{(x_{i-1}, x_i]} |F(x_i) - F(x)| |g(x)| \, dx \\
&\leq \sum_{i=1}^n \int_{(x_{i-1}, x_i]} M(\epsilon/(M(b-a))) \, dx + \sum_{i=1}^n |F(x_i) - F(x_{i-1})| M\delta \\
&= \epsilon + (F(b) - F(a))M\delta \quad \text{by monotonicity of } F \\
&\leq \epsilon + 2M^2\delta.
\end{aligned}$$

Thus if δ satisfies the additional condition that $\delta < \epsilon/(2M^2)$, then the absolute value of the difference of the two sides in the formula of the theorem is $< 2\epsilon + 2\epsilon = 4\epsilon$. This completes the proof. \square

9. Fourier Series and the Dirichlet–Jordan Theorem

A real-valued function f on a bounded interval $[a, b]$ is said to be of **bounded variation** on $[a, b]$ if there is a constant M such that every partition

$$P : a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

has

$$\sup_P \sum_{i=1}^n |f(x_i) - f(x_{i-1})| \leq M.$$

Let us write $\|f\|_{BV}$ for the least M such that this inequality holds. The set of functions f of bounded variation on $[a, b]$ is a pseudo normed linear space in the sense of Section V.9, the pseudonorm being $\|\cdot\|_{BV}$. The only functions f with $\|f\|_{BV} = 0$ are the constants.

Examples of functions of bounded variation are furnished by arbitrary bounded monotone functions and by any function with a continuous derivative. In fact, if f is monotone increasing, then f is of bounded variation with $\|f\|_{BV} = f(b) - f(a)$. If f has f' continuous, then the Mean Value Theorem gives

$$\begin{aligned} \sum_{i=1}^n |f(x_i) - f(x_{i-1})| &= \sum_{i=1}^n |f'(t_i)|(x_i - x_{i-1}) \quad \text{with } x_{i-1} < t_i < x_i \\ &\leq \|f'\|_{\text{sup}} \sum_{i=1}^n (x_i - x_{i-1}) \\ &= \|f'\|_{\text{sup}}(b - a), \end{aligned}$$

and we see that f is of bounded variation with $\|f\|_{BV} \leq \|f'\|_{\text{sup}}(b - a)$.

Let us associate two functions on $[a, b]$ to f if f is of bounded variation. For a real number r , define $r^+ = \max\{r, 0\}$ and $r^- = -\min\{r, 0\}$, so that $r = r^+ - r^-$ and $|r| = r^+ + r^-$. The functions are the **positive** and **negative variations** of f , given by

$$\begin{aligned} V^+(f)(x) &= \sup_{\substack{P \text{ with } x_0=a \\ \text{and } x_n=x}} \sum_{i=1}^n (f(x_i) - f(x_{i-1}))^+, \\ V^-(f)(x) &= \sup_{\substack{P \text{ with } x_0=a \\ \text{and } x_n=x}} \sum_{i=1}^n (f(x_i) - f(x_{i-1}))^-, \end{aligned}$$

the supremum in each case being taken over all partitions of $[a, x]$.

Proposition 6.54. If f is of bounded variation on $[a, b]$, then $V^+(f)$ and $V^-(f)$ are monotone increasing functions such that

$$f(x) = f(a) + V^+(f)(x) - V^-(f)(x)$$

for all x in $[a, b]$. In particular, f is the difference of two monotone increasing functions.

REMARK. Since monotone functions have left and right limits at each point, it follows that every function f of bounded variation has left and right limits at each point. We denote these by $f(x-)$ and $f(x+)$, respectively. The function f is continuous from the left at x if and only if $f(x-) = f(x)$, and it is continuous from the right at x if and only if $f(x) = f(x+)$.

PROOF. It is evident from the definitions that $V^+(f)$ and $V^-(f)$ are monotone increasing. Fix x , and let P be a partition of $[a, x]$. Then

$$\begin{aligned} f(x) - f(a) &= \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \\ &= \sum_{i=1}^n (f(x_i) - f(x_{i-1}))^+ - \sum_{i=1}^n (f(x_i) - f(x_{i-1}))^-. \end{aligned}$$

Hence

$$\begin{aligned} \sum_{i=1}^n (f(x_i) - f(x_{i-1}))^+ &= \sum_{i=1}^n (f(x_i) - f(x_{i-1}))^- + (f(x) - f(a)) \\ &\leq V^-(f)(x) + (f(x) - f(a)), \end{aligned}$$

and

$$V^+(f)(x) \leq V^-(f)(x) + (f(x) - f(a)).$$

Also,

$$\begin{aligned} \sum_{i=1}^n (f(x_i) - f(x_{i-1}))^- &= \sum_{i=1}^n (f(x_i) - f(x_{i-1}))^+ - (f(x) - f(a)) \\ &\leq V^+(f)(x) - (f(x) - f(a)), \end{aligned}$$

and

$$V^-(f)(x) \leq V^+(f)(x) - (f(x) - f(a)).$$

Therefore

$$f(x) - f(a) = V^+(f)(x) - V^-(f)(x),$$

and the proof is complete. \square

Theorem 6.55 (Dirichlet–Jordan Theorem). If f is a function of bounded variation on $[-\pi, \pi]$, then the Fourier series of f converges at each point to $\frac{1}{2}(f(x-) + f(x+))$ and it converges uniformly to $f(x)$ on any compact set on which the periodic extension of f is continuous.

By way of preparation, it will be convenient to extend the definition of Fourier series to allow integrable functions to be replaced by more general Borel measures. If μ is a Borel measure on $[-\pi, \pi]$, we want to be able to regard μ as periodic. One way to proceed would be to insist that μ really be a measure on the circle group, hence be defined on $(-\pi, \pi]$. Alternatively, we could insist that any point

mass contributing to μ at $-\pi$ be matched by an equal point mass for μ at π . A way of avoiding point masses contributing at the endpoints is to change the interval $[-\pi, \pi]$ to a suitable $[c - \pi, c + \pi]$; we can find a number c with no point masses at the ends because only countably many point masses can contribute to μ and still have μ be a finite measure. In any event, we define the **Fourier–Stieltjes series** of μ to be the series

$$\sum_{n=-\infty}^{\infty} c_n e^{inx} \quad \text{with} \quad c_n = \int_{(-\pi, \pi]} e^{-inx} d\mu(x).$$

The usual factor of $\frac{1}{2\pi}$ is dropped because we identify an integrable function $f \geq 0$ with the measure $\frac{1}{2\pi} f dx$ when making the generalization. From the definition of the Fourier–Stieltjes coefficients, we see immediately that $|c_n| \leq \mu((-\pi, \pi])$; hence the coefficients are bounded.

PROOF OF THEOREM 6.55. We take the given function f to be periodic of period 2π . On some closed interval $[a, b]$ containing $[-\pi, \pi]$ in its interior, let us decompose f according to Proposition 6.54 as $f = f(a) + V^+(f) - V^-(f)$. It is then enough to prove the theorem for the monotone increasing functions $f(a) + V^+(f)$ and $V^-(f)$ separately. These functions need to be extended to all of \mathbb{R}^1 , and we may make that extension by taking them to be constant to the left of $[a, b]$ and to the right of $[a, b]$.

Changing notation in the theorem, we may assume from the outset that f is monotone and bounded, though no longer periodic. Neither the Fourier coefficients of f nor the hoped-for values of the sum of the Fourier series are changed if we adjust f on a subset of the countable set where f is discontinuous. Thus we may assume without loss of generality that f is continuous from the right at every point. Let $f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$ be its Fourier series.

Let μ be the Stieltjes measure associated to f . Applying integration by parts (Theorem 6.53) on the interval $[-\pi, \pi]$ with $G(x) = e^{-inx}$ and $g(x) = -ine^{-inx}$, we obtain

$$\int_{-\pi}^{\pi} f(x)(-in)e^{-inx} dx = e^{-in\pi} f(\pi) - e^{-in(-\pi)} f(-\pi) - \int_{(-\pi, \pi]} e^{-inx} d\mu(x).$$

The left side is $-2\pi inc_n$, and the right side is the sum of two bounded terms and the negative of a Fourier–Stieltjes coefficient of μ . These Fourier–Stieltjes coefficients are bounded, and hence $|c_n| \leq C/|n|$ for some constant C .

Let $s_N(x) = \sum_{n=-N}^N c_n e^{inx}$ be the N^{th} partial sum of the Fourier series of f , and let $\sigma_N(x) = \frac{1}{N+1} \sum_{k=0}^N s_k(x)$ be the N^{th} Cesàro sum. We know that $\sigma_N(x) = (K_N * f)(x)$, where K_N is the Fejér kernel. Fejér's Theorem (Theorem 6.48) shows that $\lim_N \sigma_N(x) = \frac{1}{2}(f(x-) + f(x+))$ for all x and

that $\lim_N \sigma_N(x) = f(x)$ uniformly on any compact set of points where f is continuous. The Tauberian theorem stated as Proposition 1.50 allows us to conclude that $s_N(x)$ converges and has the same limit as $\sigma_N(x)$ if it is shown that the sequence $n(c_n e^{inx} + c_{-n} e^{-inx})$ is bounded for $n > 0$. But this boundedness is immediate from the estimate $|c_n| \leq C/|n|$ for the Fourier coefficients of f . \square

10. Distribution Functions

This section concerns the computation of integrals. A measure space (X, \mathcal{A}, ρ) will be fixed throughout. A need to estimate integrals arises in two quite distinct situations, and the emphasis is different for the two situations. One is in connection with problems in Fourier analysis and differential equations, and the underlying measure space typically has X equal to \mathbb{R}^N , \mathcal{A} equal to the σ -algebra of Borel sets, and ρ equal to Lebesgue measure. The other is in connection with probability theory, and the underlying measure space is typically a complicated space with $\rho(X) = 1$. Although the word “distribution” acquires multiple meanings in the process, the theory can begin at the same point in the two cases.

Let $f : X \rightarrow \mathbb{R}$ be a measurable function. We define a measure μ_f on the Borel sets of \mathbb{R} and a function $\lambda_f : (0, +\infty) \rightarrow [0, +\infty]$ by

$$\begin{aligned}\mu_f(E) &= \rho(f^{-1}(E)) = \rho(\{x \in X \mid f(x) \in E\}) \quad \text{for each Borel set } E, \\ \lambda_f(\xi) &= \rho(|f|^{-1}((\xi, +\infty))) = \rho(\{x \in X \mid |f(x)| > \xi\}).\end{aligned}$$

Proposition 6.56. If $f : X \rightarrow \mathbb{R}$ is a measurable function, then

- (a) $\int_X \Phi(f(x)) d\rho(x) = \int_{\mathbb{R}} \Phi(t) d\mu_f(t)$ for every nonnegative Borel measurable function $\Phi : \mathbb{R} \rightarrow \mathbb{R}$,
- (b) $\int_X \Phi(|f(x)|) d\rho(x) = \int_0^\infty \lambda_f(\xi) \varphi(\xi) d\xi$ whenever $\varphi(\xi) d\xi$ is a Stieltjes measure on \mathbb{R}^1 and Φ is its distribution function.

PROOF. In (a), when Φ is an indicator function I_E , the two sides of the identity are $\rho(f^{-1}(E))$ and $\mu_f(E)$, and these are equal by definition of μ_f . We can pass to nonnegative simple functions by linearity and then to general nonnegative Borel measurable functions Φ by monotone convergence.

In (b), when f is a nonnegative simple function s , let $s = \sum_{k=1}^n c_k I_{E_k}$ be the canonical expansion of s as a linear combination of indicator functions, with the c_j 's arranged so that $c_1 > c_2 > \cdots > c_n \geq 0$. Put $c_{n+1} = 0$. Then we have

$$\begin{aligned}\int_0^\infty \lambda_f(\xi) \varphi(\xi) d\xi &= \sum_{k=1}^n \int_{c_{k+1}}^{c_k} \rho(\bigcup_{j=1}^k E_j) \varphi(\xi) d\xi \\ &= \sum_{k=1}^n \rho(\bigcup_{j=1}^k E_j) [\Phi(c_k) - \Phi(c_{k+1})] \\ &= \sum_{k=1}^n \sum_{j=1}^k \rho(E_j) [\Phi(c_k) - \Phi(c_{k+1})]\end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^n \sum_{k=j}^n \rho(E_j) [\Phi(c_k) - \Phi(c_{k+1})] \\
&= \sum_{j=1}^n \rho(E_j) \Phi(c_j) \quad \text{since } \Phi(0) = 0 \\
&= \sum_{j=1}^n \int_{E_j} \Phi(s(x)) d\rho(x) \\
&= \int_X \Phi(s(x)) d\rho(x).
\end{aligned}$$

This proves (b) for nonnegative simple functions f . For a general measurable $|f|$ on X , choose an increasing sequence of nonnegative simple functions s_n with pointwise limit $|f|$. The definition of Φ in terms of φ makes Φ monotone increasing and continuous, and thus $\Phi(s_n(x))$ increases to $\Phi(|f(x)|)$. Also, the set $\{x \in X \mid |f(x)| > \xi\}$, for each fixed ξ , is the increasing union of the sets $\{x \in X \mid s_n(x) > \xi\}$, and thus $\lambda_{s_n}(\xi) = \rho(\{x \in X \mid s_n(x) > \xi\})$ increases to $\lambda_f(\xi) = \rho(\{x \in X \mid |f(x)| > \xi\})$ for each ξ . Hence we can pass to the limit in the identity for each s_n and obtain the identity for $|f|$ by monotone convergence. This proves (b) for a general measurable $|f|$. \square

For applications to Fourier analysis and differential equations, it is (b) that is important, and the function Φ of most interest is $\Phi(t) = t^p$ with $0 < p < +\infty$. The formula in this case is

$$\int_X |f(x)|^p d\rho(x) = p \int_0^\infty \lambda_f(\xi) \xi^{p-1} d\xi.$$

Somewhat unfortunately, the function λ_f is called the **distribution function** of f ; the term does not conflict with the notion of the “distribution function” of a Stieltjes measure as long as one does not make any associations between functions and measures.

A special case of the displayed formula is that X is \mathbb{R}^N , ρ is Lebesgue measure, and p is 1. In this case the formula simplifies to $\int_{\mathbb{R}^N} |f(x)| dx = \int_0^\infty \lambda_f(\xi) d\xi$, a formula that was mentioned after the statement of the Hardy–Littlewood Maximal Theorem (Theorem 6.38).

The displayed formula shows that $\int_X |f|^p d\rho$ can be computed from the function λ_f , and it is apparent that the integral cannot be finite if $\lambda_f(\xi)$ is everywhere \geq some positive multiple of ξ^{-p} . This observation can be improved upon without the aid of Proposition 6.56 in the following way. We have $\int_X |f(x)|^p d\rho(x) \geq \int_{\{x \in X \mid |f(x)| > \xi\}} |f(x)|^p d\rho(x) \geq \xi^p \rho(\{x \in X \mid |f(x)| > \xi\})$. Thus we obtain

$$\rho(\{x \in X \mid |f(x)| > \xi\}) \leq \frac{\int_X |f|^p d\rho}{\xi^p},$$

an inequality that goes under the name **Chebyshev’s inequality**.

11. Problems

- Let S^1 be the unit circle of \mathbb{C} , let T be the subgroup of elements of finite order, and let E be a subset of S^1 that contains exactly one element of each coset in S^1/T . (Such a set E exists by the Axiom of Choice.) Prove that E is not a Lebesgue measurable subset of the circle and therefore that the corresponding subset of $(-\pi, \pi]$ is not Lebesgue measurable on \mathbb{R}^1 .
- Let φ be the mapping given explicitly in Section 5 that allows one to substitute in an expression in Cartesian coordinates and obtain an expression in spherical coordinates. Let U be the domain of φ . Prove that
 - the determinant factor in the change-of-variables formula is given by

$$|\det \varphi'| = r^{N-1} \sin^{N-2} \theta_1 \sin^{N-3} \theta_2 \cdots \sin \theta_{N-2},$$

- φ is one-one on U ,
 - the complement of $\varphi(U)$ in \mathbb{R}^N is a lower-dimensional set.
- Let L be a nonsingular N -by- N real matrix. Prove that

$$\int_{\mathbb{R}^N} f(Lx) dx = |\det L|^{-1} \int_{\mathbb{R}^N} f(x) dx$$

for every nonnegative Borel measurable function f .

- Let M_N denote the N^2 -dimensional Euclidean space of all real N -by- N matrices, and let dx refer to its Lebesgue measure. Prove that

$$\int_{M_N} f(yx) \frac{dx}{|\det x|^N} = \int_{M_N} f(x) \frac{dx}{|\det x|^N}$$

for each nonsingular matrix y and Borel measurable function $f \geq 0$. In the formula, yx is the matrix product of y and x .

- Fix α with $0 < \alpha < 1$. Suppose $f : \mathbb{R} \rightarrow \mathbb{C}$ is periodic of period 2π , is smooth except at multiples of 2π , and satisfies the inequalities $|f(x)| \leq C|x|^\alpha$, $|f'(x)| \leq C|x|^{\alpha-1}$, and $|f''(x)| \leq C|x|^{\alpha-2}$ for $|x| \leq 1$.
 - By breaking the integral at $|x| = 1/|n|$, prove that the Fourier coefficients c_n of f satisfy $|c_n| \leq K/|n|^{1+\alpha}$.
 - How can one conclude from (a) that the Fourier series of f converges uniformly? Why is the limit equal to f ?
 - Prove or disprove: The real and imaginary parts of the function f are of bounded variation on every bounded interval.
- Let μ be a nonzero measure on the σ -algebra of all subsets of \mathbb{R}^1 assigning to each set either measure 0 or measure 1. Prove that μ is a point mass.

7. Determine all Stieltjes measures $\nu \neq 0$ on the line with

$$\int_{\mathbb{R}^1} fg \, d\nu = \left(\int_{\mathbb{R}^1} f \, d\nu \right) \left(\int_{\mathbb{R}^1} g \, d\nu \right)$$

for all continuous nonnegative functions f and g .

Problems 8–10 make use of Fubini's Theorem in unexpected ways.

8. (a) Show that the complement of any Lebesgue measurable set of Lebesgue measure 0 in \mathbb{R}^N is dense.
 (b) Let μ be a Stieltjes measure on the line, and let E be a Borel set in \mathbb{R}^1 with Lebesgue measure 0. Prove that $\mu(E + t) = 0$ for almost every t with respect to Lebesgue measure.
 (c) Suppose that a Stieltjes measure μ on the line satisfies $\lim_{t \rightarrow 0} \mu(E + t) = \mu(E)$ for each bounded Borel set E in \mathbb{R}^1 . Prove that $\mu(E) = 0$ for every Borel set E of Lebesgue measure 0.
9. In potential theory a positive charge on \mathbb{R}^3 is by definition any finite Borel measure μ , and its potential h is the function $h(x) = \int_{\mathbb{R}^3} \frac{d\mu(y)}{|x-y|}$. Prove that the potential is finite almost everywhere with respect to Lebesgue measure.
10. Let $P(x_1, \dots, x_n)$ be a real-valued polynomial on \mathbb{R}^n that is not identically 0. Prove by induction that the set in \mathbb{R}^n where $P = 0$ has Lebesgue measure 0.

Problems 11–14 concern the gamma function and some associated changes of variables.

11. Prove that

$$\int_0^1 t^{x-1} (1-t)^{y-1} dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

by starting from the product of $\Gamma(x+y)$ and the left side, substituting for $\Gamma(x+y)$, making a change of variables, using Fubini's Theorem, and making another change of variables.

12. By evaluating the integral $\int_{\mathbb{R}^N} e^{-\pi|x|^2} dx$ first in Cartesian coordinates by means of Proposition 6.33 and then in spherical coordinates by means of the change-of-variables formula for multiple integrals, obtain an expression for the area $\Omega_{N-1} = \int_{S^{N-1}} d\omega$ of the sphere S^{N-1} . Express the answer in terms of a value of the gamma function.
13. Let I be the “cube” of all $u = (u_1, \dots, u_n)$ in \mathbb{R}^n with $0 < u_i < 1$ for all i , and let S be the “simplex” of all $x = (x_1, \dots, x_n)$ in \mathbb{R}^n with $x_i > 0$ for all i and $\sum_{i=1}^n x_i < 1$. Define $x = \varphi(u)$ by

$$\begin{aligned}x_1 &= u_1, \\x_2 &= (1 - u_1)u_2, \\&\vdots \\x_n &= (1 - u_1) \cdots (1 - u_{n-1})u_n.\end{aligned}$$

- (a) Prove that $\sum_{i=1}^n x_i = 1 - \prod_{i=1}^n (1 - u_i)$.
 (b) Prove that φ maps I one-one onto S , with inverse given by

$$u_i = \frac{x_i}{1 - x_1 - \cdots - x_{i-1}}.$$

- (c) Prove that $|\det \varphi'(u)| = (1 - u_1)^{n-1} (1 - u_2)^{n-2} \cdots (1 - u_{n-1})$ and that
 $|\det(\varphi^{-1})'(x)| = [(1 - x_1)(1 - x_1 - x_2) \cdots (1 - x_1 - \cdots - x_{n-1})]^{-1}$.

14. Using Problems 11 and 13, prove for the simplex S in Problem 13 that

$$\int_S x_1^{a_1-1} x_2^{a_2-1} \cdots x_n^{a_n-1} dx = \frac{\Gamma(a_1)\Gamma(a_2) \cdots \Gamma(a_n)}{\Gamma(a_1 + \cdots + a_n + 1)}$$

when $a_j > 0$ for all j .

Problems 15–17 concern the Cantor function for the standard Cantor set.

15. Prove that the values of the Cantor function F for the standard Cantor set are independent of the method of defining the approximating functions F_n on the complementary closed intervals as long as F_n is monotone increasing and continuous.
 16. Compute $\int_0^1 F(x) dx$ if F is the Cantor function for the standard Cantor set.
 17. The Stieltjes measure μ corresponding to the Cantor function for the standard Cantor set C is called the **Cantor measure**. The set C consists of the members of $[0, 1]$ that can be expanded in the digits 0, 1, 2 of base 3 without using any 1's. Show, for each n -tuple of 0's and 2's, that μ attaches measure 2^{-n} to the subset of C whose base 3 expansion begins with that n -tuple.

Problems 18–20 introduce the **Poisson integral formula** for the unit disk in \mathbb{R}^2 . The Poisson kernel was the subject of Problems 27–29 at the end of Chapter I and is given by

$$P_r(\theta) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta} = \frac{1 - r^2}{1 - 2r \cos \theta + r^2}.$$

Harmonic functions in the unit disk were the subject of Problems 14–15 at the end of Chapter III and also Problems 10–13 at the end of Chapter IV. The present set of problems begins to relate the Poisson kernel to harmonic functions via convolution.

18. If f is in $L^1([-\pi, \pi], \frac{1}{2\pi} d\theta)$, then the **Poisson integral** of f is the function in the unit disk defined in polar coordinates by

$$u(r, \theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\varphi) P_r(\theta - \varphi) d\varphi.$$

If c_n is the n^{th} Fourier coefficient of f , prove that $u(r, \theta) = \sum_{n=-\infty}^{\infty} c_n r^{|n|} e^{in\theta}$, and conclude that u is harmonic in the open unit disk.

19. If p equals 1 or 2 and if f is in $L^p([-\pi, \pi], \frac{1}{2\pi} d\theta)$, prove that the Poisson integral $u(r, \theta)$ of f has the properties that $\|u(r, \cdot)\|_p \leq \|f\|_p$ for $0 \leq r < 1$ and that $u(r, \cdot)$ tends to f in L^p in the sense that $\lim_{r \uparrow 1} \|u(r, \cdot) - f\|_p = 0$.
20. Suppose that f is in $L^\infty([-\pi, \pi], \frac{1}{2\pi} d\theta)$ and that $u(r, \theta)$ is the Poisson integral of f .
- Prove that $\lim_{r \uparrow 1} u(r, \theta) = f(\theta)$ uniformly on any set of θ 's where f is uniformly continuous.
 - For f of class C^2 , prove that the Poisson integral of f is the only harmonic function $u(r, \theta)$ in the disk such that $\lim_{r \uparrow 1} u(r, \theta) = f(\theta)$ uniformly in θ .
 - Prove that $u(r, \cdot)$ tends to f weak-star in L^∞ relative to L^1 in the sense that $\lim_{r \uparrow 1} \int_{-\pi}^{\pi} u(r, \theta) g(\theta) d\theta = \int_{-\pi}^{\pi} f(\theta) g(\theta) d\theta$ for all g in L^1 . (Weak-star convergence was defined in Section V.9.)

Problems 21–25 concern functions of bounded variation. For such a function f , the positive and negative variations of f were defined in Section 9, and their values at x were denoted by $V^+(f)(x)$ and $V^-(f)(x)$.

21. Prove that the product of two functions of bounded variation on $[a, b]$ is of bounded variation.
22. This problem concerns a certain minimality property of the decomposition $f(x) = f(a) + V^+(f)(x) - V^-(f)(x)$ of a function f of bounded variation on $[a, b]$. Prove that if g_1 and g_2 are any two nonnegative monotone increasing functions such that $f(x) = f(a) + g_1(x) - g_2(x)$ for all x , then $V^+(f)(x) \leq g_1(x)$ and $V^-(f)(x) \leq g_2(x)$.
23. Prove that if f is of bounded variation on $[a, b]$ and is continuous at a point x in (a, b) , then both $V^+(f)$ and $V^-(f)$ are continuous at x .
24. If f is of bounded variation on $[a, b]$, define the **total variation** of f as the function given by

$$V(f)(x) = \sup_{\substack{P \text{ with } x_0=a \\ \text{and } x_n=x}} \sum_{i=1}^n |f(x_i) - f(x_{i-1})|,$$

the supremum being taken over all partitions of $[a, x]$. Prove that $V(f)(x) = V^+(f)(x) + V^-(f)(x)$ for all x .

25. Prove that the function f on $[-1, 1]$ given by

$$f(x) = \begin{cases} x \sin(1/x) & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$

is not of bounded variation. Prove or disprove that the function g on $[-1, 1]$ given by

$$g(x) = \begin{cases} x^2 \sin(1/x) & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$

is of bounded variation.

Problems 26–27 use elementary complex analysis as in Appendix B to shed further light on the gamma function as defined in this chapter.

26. (a) Prove that $\Gamma(s)$ is continuous for $\operatorname{Re} s > 0$.
(b) Use Morera's Theorem to prove that $\Gamma(s)$ is analytic for $\operatorname{Re} s > 0$.
27. Prove that the analytic function $\Gamma(s)$, initially defined for $\operatorname{Re} s > 0$, extends to a meromorphic function in \mathbb{C} whose only poles are at the nonpositive integers. Prove moreover that the extended function satisfies $\Gamma(s+1) = s\Gamma(s)$ at all points s other than the nonpositive integers.