Anthony W. Knapp

Basic Real Analysis

Along with a companion volume
Advanced Real Analysis

Selected Pages from Chapter VI: pp. 296–334

Birkhäuser
Boston • Basel • Berlin

Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, NY 11733-1729
U.S.A.
e-mail to: aknapp@math.sunysb.edu
http://www.math.sunysb.edu/~ aknapp/books/basic.html

Cover design by Mary Burgess.

Mathematics Subject Classicification (2000): 28-01, 26-01, 42-01, 54-01, 34-01

Library of Congress Cataloging-in-Publication Data
Knapp, Anthony W.
Basic real analysis: along with a companion volume Advanced real analysis / Anthony W. Knapp
p. cm. – (Cornerstones)
Includes bibliographical references and index.
ISBN 0-8176-3250-6 (alk. paper)
1. Mathematical analysis. I. Title. II. Cornerstones (Birkhäuser)

QA300.K56 2005
515–dc22 2005048070

ISBN-10 0-8176-3250-6 eISBN 0-8176-4441-5 Printed on acid-free paper.
ISBN-13 978-0-8176-3250-2

Advanced Real Analysis ISBN 0-8176-4382-6
Basic Real Analysis and *Advanced Real Analysis* (Set) ISBN 0-8176-4407-5

Printed in the United States of America. (MP)

9 8 7 6 5 4 3 2 1 SPIN 10934074

Birkhäuser is a part of *Springer Science+Business Media*

www.birkhauser.com

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 is $\mathbb{R}^m$ 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 $\mathcal{B}_{m+n} \subseteq \mathcal{B}_m \times \mathcal{B}_n$.

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 $F_n \subseteq F_{n+1}^o$ 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 $K \neq \varnothing$, and we can take all $U_n = \varnothing$ if $K = \varnothing$.

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

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

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 $\mu(U_1)$ 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^o_{M+1}, 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 $F_{m+1} - F_n$. 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 of C. The point x_0 is in some F_M since the sets F_M have union the whole space. Since $F_M \subseteq F^o_{M+1}$ and F^o_{M+1} is open, the sequence is eventually in F^0_{M+1}. The inclusion $C_n \subseteq E_n \subseteq F_{n+1} - F_n$ shows that $C_n \cap F_{M+1} = \varnothing$ 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

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

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.

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 $f_n : \mathbb{R}^N \to [0, 1]$ 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 $g_n = \min\{f_1, \ldots, f_n\}$, we may assume that they are pointwise decreasing. Consequently

(iii) there exists a decreasing sequence of real-valued members of $C_{\mathrm{com}}(\mathbb{R}^N)$ 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 $C_{\mathrm{com}}(\mathbb{R}^N)$ 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.

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

(a) $C_{\mathrm{com}}(\mathbb{R}^N)$ 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 $C_{\mathrm{com}}(\mathbb{R}^N)$ 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 $L^p(\mathbb{R}^N)$ 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 $L^p(\mathbb{R}^N)$ containing all indicator functions of compact sets is $L^p(\mathbb{R}^N)$ 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 $C_{\mathrm{com}}(\mathbb{R}^N)$ 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 $C_{\mathrm{com}}(\mathbb{R}^N)$ in $L^p(\mathbb{R}^N)$ contains all indicator functions of compact sets. By Proposition 5.55d this closure contains the smallest closed subspace of $L^p(\mathbb{R}^N)$ containing all indicator functions of compact sets. Conclusion (b) shows that the latter subspace is $L^p(\mathbb{R}^N)$ itself. This proves (a).

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

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

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 $U^c \neq \varnothing$. 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 $Q \cap Q' \neq \varnothing$, 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.

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.

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 $F : \mathbb{R}^N \to \mathbb{R}^{N'}$, $F^{-1}(E)$ 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 $F_j : \mathbb{R}^N \to \mathbb{R}^1$ 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 $\mathbb{R}^N$.

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.

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.

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.

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 $F:\mathbb{R}^{2N}\to\mathbb{R}^N$ given by $F(x,y)=x-y$, followed by the Borel function $f:\mathbb{R}^N\to[0,+\infty]$. 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

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

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

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),

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

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

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

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.

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.

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 $\lim_{h\to 0} \|\tau_{t+h} f - \tau_t f\|_p = 0$ 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

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

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

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 $\mathbb{R}^N$, $1 \le j \le n$, and constants c_j such that $\left\| f * g_k - \sum_{j=1}^n c_j \tau_{y_j} g_k \right\|_p < \epsilon$ for $1 \le k \le r$.

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 $1 \le k \le r$, 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 $|y - y_i| \ge \delta$ 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

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

Summing over j gives

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

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 $\big\| f_1 * g_k - \sum_{j=1}^n c_j \tau_{y_j} g_k \big\|_p < \epsilon/2$ and $\big\| f_2 * g_k - \sum_{j=1}^n d_j \tau_{y_j} g_k \big\|_p < \epsilon/2$ for suitable y_j's and c_j's, and the triangle inequality gives $\big\|(f_1 + f_2) * g_k - \sum_{j=1}^n (c_j + d_j)\tau_{y_j} g_k \big\|_p < \epsilon$. Hence V is closed under addition. Similarly V is closed under scalar multiplication. If $f_l \to f$ in L^1 with f_l in V and if $\epsilon > 0$ is given, choose l large enough so that $\|f - f_l\|_1 < \epsilon\big/(2\max \|g_k\|_p)$. If $\big\| f_l * g_k - \sum_{j=1}^n c_j^{(l)} \tau_{y_j^{(l)}} g_k \big\|_p < \epsilon/2$, then the inequality $\|f * g_k - f_l * g_k\|_p \le \|f - f_l\|_1 \|g_k\|_p$ and the triangle inequality together give $\big\| f * g_k - \sum_{j=1}^n c_j^{(l)} \tau_{y_j^{(l)}} g_k \big\|_p < \epsilon$. Hence f is in V, and V is closed. By Corollary 6.4b, $V = L^1(\mathbb{R}^N)$, and the proof is complete.

In some cases with $L^\infty(\mathbb{R}^N)$, results have more content when phrased in terms of the supremum norm $\|f\|_{\sup} = \sup_{x\in\mathbb{R}^N} |f(x)|$ 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 $\lim_{h\to 0} \|\tau_h f - f\|_{\sup} = 0$ is equivalent to uniform continuity of f, basically by definition. The functions f in L^∞ for which $\lim_{h\to 0} \|\tau_h f - f\|_\infty = 0$ 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,

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

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 $\|f * g\|_{\sup} = \sup_x \big| \int_{\mathbb{R}^N} f(x-y) g(y)\, dy \big| \le \sup_x \|g\|_\infty \int_{\mathbb{R}^N} |f(x-y)|\, dy = \|f\|_1 \|g\|_\infty$.

For uniform continuity we use Proposition 6.15 and the bound $\|f * g\|_{\sup} \le \|f\|_1 \|g\|_\infty$ to make the estimate

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

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

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 $\le n$.

PROOF. First suppose that $n = 1$. Fix j with $1 \le j \le N$, 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 $|h| \le 1$. Proposition 6.15 gives

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

Proposition 3.28a shows that $h^{-1}(\tau_{-he_j} f - f)$ converges uniformly, as $h \to 0$, 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

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

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

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

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

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 $\lim_{t \to 0} \|\tau_t f - f\|_\infty = 0$,

(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

$$\begin{aligned}|(\varphi_\varepsilon * f)(x) - cf(x)| &= \Big|\int_{\mathbb{R}^N} \varphi_\varepsilon(y) f(x-y)\,dy - cf(x)\Big|\\ &= \Big|\int_{\mathbb{R}^N} \varphi_\varepsilon(y)[f(x-y) - f(x)]\,dy\Big|\\ &\le \int_{\mathbb{R}^N} |\varphi_\varepsilon(y)|\,|f(x-y) - f(x)|\,dy. \qquad (*)\end{aligned}$$

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

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

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

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

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 $2(\sup_{t\in\mathbb{R}^N} |f(t)|)\int_{|y|>M} |\varphi(y)|\,dy \le \eta$. Then choose $\delta > 0$ by uniform continuity such that $|u| \le \delta$ implies that $|f(x+u) - f(x)| \le \eta/\|\varphi\|_1$. Whenever $\varepsilon \le \delta/M$, the inequality $|y| \le M$ implies that $|f(x-\varepsilon y) - f(x)| \le \eta\big/\|\varphi\|_1$. Then

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

and (d) follows.

Corollary 6.21. If f in $L^\infty(\mathbb{R}^N, dx)$ satisfies $\lim_{t\to 0} \|\tau_t f - f\|_\infty = 0$, 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 $\|f-g\|_\infty \le \|f-\varphi_{\varepsilon_n} * f\|_\infty + \|\varphi_{\varepsilon_n} * f - g\|_\infty$, 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.

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 $V \ne \mathbb{R}^N$, we can take $F_n = \{x \in V \mid D(x, V^c) \ge 1/n \text{ and } |x| \le n\}$ as long as n is $\ge$ some suitable n_0. We complete the definition of the F_n's by taking $F_1 = \cdots = F_{n_0-1} = F_{n_0}$.

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 $U_n = \big\{y \in \mathbb{R}^N \bigm| D(y, K) < 1/n\big\}$ 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 $f_n : \mathbb{R}^N \to [0, 1]$ 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.

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 $\mathcal{B}$.

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.

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.

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.

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 $L^p(F_n^o, \mu\big|_{F_n^o})$. If $\epsilon > 0$ is given, choose g in $C_{\text{com}}(F_n^o)$ with $\|f-g\|_p \le \epsilon$. Then choose h in the countable dense set D_n of $C_{\text{com}}(F_n^o)$ such that $\|g-h\|_{\sup} \le \epsilon$. Since $\|f-h\|_p \le \|f-g\|_p + \|g-h\|_p$ and $\|g-h\|_p^p = \int_{F_n^o} |g(x) = h(x)|^p\,d\mu(x) \le \epsilon^p \mu(F_n)$, we obtain $\|f-h\|_p \le \epsilon + \epsilon\mu(F_n)^{1/p}$. 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} is a vector subspace containing all indicator functions of compact subsets of V. By (b), $D^{\text{cl}} = L^p(V,\mu)$.

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 $F : V \to V'$, $F^{-1}(E)$ 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 $\mathcal{B}_{N'}(V')$.

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

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

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 $L_k(x) \le L_{k+1}(x)$ and $U_k(x) \ge U_{k+1}(x)$ 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 $L_k(x) \le f(x) \le U_k(x)$ for x in E, we see that

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

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

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

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

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

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 $L(x) \le U(x)$ 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 $L(x) \le f(x) \le U(x)$ 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

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

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

$$\int_U g\, d\mu = \int_U g(x)\, |\det \varphi'(x)|\, dx \tag{$*$}$$

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

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

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

$$\int_U g\, d\nu = \int_U g\, d\mu \tag{$\dagger$}$$

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.

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

$$U = \left\{ \left(\begin{smallmatrix} r \\ \theta \end{smallmatrix}\right) \;\middle|\; 0 < r < +\infty \text{ and } 0 < \theta < 2\pi \right\},$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \varphi \left(\begin{smallmatrix} r \\ \theta \end{smallmatrix}\right) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix},$$

and we have

$$\varphi(U) = \mathbb{R}^2 - \left\{ \left(\begin{smallmatrix} x \\ 0 \end{smallmatrix}\right) \;\middle|\; x \geq 0 \right\}.$$

Since $\det \varphi' \left(\begin{smallmatrix} r \\ \theta \end{smallmatrix}\right) = r$, 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. $\displaystyle\int_{-\infty}^{\infty} e^{-\pi x^2}\, dx = 1.$

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:

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

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

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 $t^{s-1} \leq t^n = 2^n n!\big(\frac{1}{n!}\big(\frac{t}{2}\big)^n\big) \leq 2^n n! \sum_{k=0}^{\infty} \frac{1}{k!}\big(\frac{t}{2}\big)^k = 2^n n! e^{t/2}$. 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

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

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) $\Gamma\big(\frac{1}{2}\big) = \sqrt{\pi}$.

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 $\varepsilon \to 0$ and $M \to \infty$. 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

$$\Gamma\big(\tfrac{1}{2}\big) = \int_0^{\infty} t^{-1/2}e^{-t}\,dt = \int_0^{\infty}(\pi x^2)^{-1/2}e^{-\pi x^2}2\pi x\,dx = 2\sqrt{\pi}\int_0^{\infty} e^{-\pi x^2}\,dx.$$

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 $\Gamma\big(\frac{1}{2}\big) = \sqrt{\pi}$.

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 $\psi : V \to \mathbb{R}^N$ 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 $\psi : V \to \mathbb{R}^N$ 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

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

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

$$|\psi(x') - \psi(x)| \leq a|x' - x| \tag{$**$}$$

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

$$|\psi(x') - L_x(x')| \leq b|x' - x|^2 \tag{$\dagger$}$$

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

$\le bc^2$ from this hyperplane. In each of the $N-1$ perpendicular directions, $(**)$ shows that $\psi(x')$ and $\psi(x)$ are at distance $\le ac$ from each other. Thus $\psi(x')$ is contained in a box[1] 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 $|x'-x| \le d/M$. 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 $\le 2^N a^{N-1} bd^{N+1} M^{-1}$. 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.

Corollary 6.36. If $\psi : V \to \mathbb{R}^N$ 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 $\mathbb{R}^N \to \mathbb{R}^M$ 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.

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

$$U = \left\{ \begin{pmatrix} r \\ \theta_1 \\ \vdots \\ \theta_{N-1} \end{pmatrix} \;\middle|\; \begin{array}{l} 0 < r < +\infty, \\ 0 < \theta_j < \pi \text{ for } 1 \le j \le N-2, \\ 0 < \theta_{N-1} < 2\pi \end{array} \right\}$$

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

[1] This 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

$$\int_{\mathbb{R}^N} f(x)\,dx = \int_{r=0}^{\infty}\int_{\theta_1=0}^{\pi}\cdots\int_{\theta_{N-2}=0}^{\pi}\int_{\theta_{N-1}=0}^{2\pi} f(r\cos\theta_1,\dots)$$

$$\times\, r^{N-1}\sin^{N-2}\theta_1\cdots\sin\theta_{N-2}\,d\theta_{N-1}\cdots d\theta_1\,dr.$$

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

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

and

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

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

$$m\big\{x \,\big|\, f^*(x) > \xi\big\} \le \frac{5^N \|f\|_1}{\xi}$$

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

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

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

$$\begin{aligned}
&\big\{x \bigm| \limsup |T_r f(x) - f(x)| > \xi\big\} \\
&\qquad\qquad \subseteq \big\{x \bigm| (f-g)^*(x) > \xi/2\big\} \cup \big\{x \bigm| |f(x) - g(x)| > \xi/2\big\}.
\end{aligned}$$

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

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

Since ϵ is arbitrary, $S(\xi) = \big\{x \bigm| \limsup |T_r f(x) - f(x)| > \xi\big\}$ 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.

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

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

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

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

and

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

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.

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 $m\{x \mid f^*(x) > \xi\} \le 5^N\|f\|_1/\xi$, 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 $\ge$ 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 $\le$ 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 $g \ge 0$ is integrable and S is the set where $g > \xi$, then $g \ge \xi I_S$ everywhere; hence $\|g\|_1 \ge \xi\, m(S)$ and $m(S) \le \|g\|_1/\xi$. A function g is said to be in **weak** L^1 if

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

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 $B(r_1; x_1), B(r_2; x_2), \ldots$, such that either the collection is infinite and $\inf_{1 \le j < \infty} r_j \ne 0$ or

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

In either case,

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

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 $r \le 1$. 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 $r_1 \ge \frac{1}{2} R_1$, 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 $r_2 \ge \frac{1}{2} R_2$. 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

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

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

$$V_j \subseteq B(5r_j; x_j). \tag{$*$}$$

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

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

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

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

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.

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 $B(r_j; x_j)$.

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 $\lim_{R\to\infty} \varphi_0(R) = 0$, 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

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

and thus

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

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

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

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 $L^1(\mathbb{R}^N)$.

EXAMPLE. An example of a function φ as in Corollary 6.42 is $P(x) = \dfrac{1}{1+x^2}$ 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 $\dfrac{\partial^2 h}{\partial x^2} + \dfrac{\partial^2 h}{\partial y^2} = 0$, 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 with 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: $\frac{1}{2\pi}\int_{-\pi}^{\pi} |f|\,1\,dx \le \big(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f|^2\,dx\big)^{1/2}\big(\frac{1}{2\pi}\int_{-\pi}^{\pi} 1\,dx\big)^{1/2}$.

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