Introduction to Integration

H. A. PRIESTLEY

CLARENDON PRESS · OXFORD

Introduction to Integration

This book has been printed digitally and produced in a standard specification in order to ensure its continuing availability

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

Reprinted 2004

ISBN 0-19-850123-4

Preface

Integrals play a central role in mathematics, and a university mathematical education would be incomplete without some study of integration. This textbook both provides a rigorous practical guide to the use of integrals and introduces the beautiful and powerful theory due to Lebesgue. It presents 'integration for the working mathematician' and is not directed solely, or even principally, at specialists in mathematical analysis.

Traditionally, a first course on integration is based on the Riemann integral, or something similar. The Lebesgue integral is introduced, quite separately and often very theoretically, in advanced undergraduate or beginning graduate courses aimed at specialists in analysis and probability. The integration of continuous functions on compact intervals can be treated in many different ways, all of which are adequate for discussing elementary applications. However many of the important integrals of applied mathematics are of functions whose domain is $\mathbb{R}$ or $\mathbb{R}^+$: Fourier and Laplace transforms, the distribution functions of the normal and exponential distributions, integrals defining the Gamma function and other special functions, ... Here the Lebesgue approach scores heavily, in both elegance and ease of use, over the extension to unbounded intervals of Riemann integration. I hope that those who would argue in favour of the latter on the grounds that it more accessible will be convinced by this book that their view is mistaken. The approach here is a unified one. The beginners' integral with which the book starts is a simplified Lebesgue-style integral, based on step function approximations, for continuous functions on closed bounded intervals. The subtleties of the general Lebesgue theory—specifically null sets and convergence almost everywhere—do not intrude at this level. The transition to the full-blown Lebesgue integral, defined via monotonic sequences of step functions, can then be made without conceptually starting afresh.

An important part of the philosophy of the book is that it is the **properties** of the integral that are crucial, and that the manner in which it is constructed is secondary. Thus stress is placed on such properties as linearity and positivity and, at a deeper level, the Monotone and Dominated Convergence Theorems. The foundations for the integral—more substantial for the later part of the book than for the more elementary early part—are laid, so that the existence of an integral with the desired properties can be established. However, like the foundations of a building, these underpinnings rarely need to be examined once they are in place. Therefore it would be quite feasible to use the book in a purely dogmatic way with students for whom such an approach is thought appropriate.

The book is a companion volume to my earlier OUP text "Introduction to complex analysis". The prerequisites are similar: knowledge of elementary differential and integral calculus and of the rudiments of mathematical analysis, including simple ε-δ arguments. Chapter 2 summarizes the necessary background. The material is divided into 34 very short chapters. This format is intended to make individual topics easy to locate and easy to digest. On the premise that students of integration ought to be proficient at practical integration, emphasis is put on examples throughout. In addition to sets of exercises at the ends of almost all chapters there are exercises in the body of the text, called 'exercise examples'. Their purpose is to reinforce understanding of the immediately preceding theory.

It is not envisaged that the whole book would be used for a single course, but rather that different parts would be studied at different stages. The book falls roughly into three parts.

- Chapters 3–8 present a beginners' integral, for continuous functions on compact intervals. The theory is developed far enough to link 'integral as area' with 'integral as antiderivative'. It is accompanied by numerous examples and exercises to reinforce and extend students' manipulative skills and by applications (Simpson's rule, for example).
- Chapters 9–26 cover the theory of the Lebesgue integral proper. Here it is in the profusion of concrete examples that this account differs most from existing texts.
- Chapters 27–33 are devoted to illustrations of how the theory is applied in areas with which it is customarily associated: the Lebesgue spaces, Fourier analysis, and orthogonal expansions are briefly discussed. Finally, Chapter 34 makes connections with probability theory.

This is not a book for aficionados of measure theory. The stress throughout is on functions rather than on sets. Lebesgue measure is only introduced late on, and general measures get only a fleeting mention, in the context of probability. Probabilists, too, should not have high expectations that the approach is customized for them: Chapter 34 aims only to provide links to specialized and advanced texts, where topics it merely hints at, or omits altogether, are fully developed.

My reason for writing the book was to share 25 years experience of teaching integration theory. I make no claims for originality in the material. The construction of the Lebesgue integral via monotonic sequences of step functions follows a well-worn route, previously trodden by, among others, T.M. Apostol, B.Sz.-Nagy, and A.J. Weir. Their texts are cited in the bibliography, and the influence of their work on my presentation is considerable. Even more important is my debt to past and present Oxford colleagues. First of all to John Kingman, who in the 1970s persuaded the Oxford Mathematics Faculty to include rudimentary Lebesgue integration in the first year undergraduate mathematics course and who stimulated my interest in the subject. I must also acknowledge a substantial debt to John Roe, who propounded the simplified Lebesgue-style integral presented in the first part of the book. I believe the 'Roe integral' to have considerable pedagogical merits as a stepping stone to the Lebesgue integral, and its use here is a key feature of the approach I have adopted. In addition, I have profited from the mathematical wisdom of other colleagues who have taught

integration theory courses at various times, especially David Edwards, Aubrey Ingleton, and Wilson Sutherland. I am most grateful for the constructive feedback I have had from those upon whom I have inflicted evolutionary drafts of the book. My hearty thanks go in particular to Graham Nelson, for his witty and eclectic comments, and, among my students, Jonathan Bevan, Kathryn Harriman, James Oliver, Tim Ricketts, and Steven Siddals, all of whom have hunted mathematical and typographical errors with zeal.

My thanks are, of course, due to all the staff of Oxford University Press who have been involved with the production of the book. It has been typeset by its author using $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX (supplemented by excellent numbering macros produced by Ralph Freese), a system which was admirably suited to the task.

H.A.P.

March 1997

Contents

Notation

We assume familiarity with the standard notation and vocabulary relating to sets: $\cup$ and $\cap$ (union and intersection), $\in$ (set membership), and so on. We also adopt the symbol $:=$, meaning 'equals by definition' and the Bourbaki dangerous bend sign, ☡ , to warn of a common pitfall. As is now customary, $\square$ signals the end of a proof.

Remarks in the text enclosed in square brackets provide links to results later in the book and comments directed at those with the knowledge to appreciate them. All these remarks can be ignored on a first reading.

The chapters, except the first, are divided into numbered sections. Each section and each exercise is labelled with the chapter number as well as an individual number. Cross-references to sections are purely numeric; those to exercises are prefaced by 'Exercise'. Numbers in square brackets refer to items in the bibliography.

1 Setting the scene

Two millennia ago the Greeks calculated areas by the method of exhaustion. This was the process by which an area was approximated from the inside by regular figures, for example rectangles, until the residual area (or the human calculator?) was exhausted. Figure 1.1 gives an illustration. Nowadays numerical approximations to integrals are obtained by computer, less arduously but in essentially the same way.

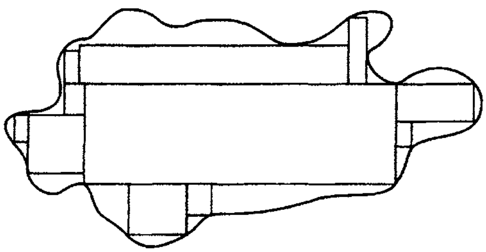


Figure 1.1

In school calculus integrals are of two kinds: indefinite integrals such as $\int \sin x = -\cos x + C$, C being any constant, and definite integrals like

$$\int_0^4 x^{101}\,\mathrm{d}x \quad \text{or} \quad \int_{-\pi}^{\pi} \sin x\,\mathrm{d}x.$$

The statement

$$\int f(x) = F(x) + C$$

serves as shorthand for '$f(x)$ is the derivative of $F(x) + C$', while

$$\int_a^b f(x)\,\mathrm{d}x$$

represents the area under that part of the graph of the function f which lies between the points a and b on the x-axis, as in Fig. 1.2. The link between these apparently unrelated uses of the symbol $\int$ comes through the 'Fundamental Theorem of Calculus':

$$\int_a^b F'(x)\,\mathrm{d}x = [F(x)]_a^b = F(b) - F(a),$$

or, in different notation,

$$\int_a^x F'(t)\,\mathrm{d}t = [F(t)]_a^x = F(x) + C \quad \text{(where } C = -F(a)\text{)}.$$

To call this a theorem is misleading. At this stage it is more an act of faith. Why should area and antidifferentiation be connected in this way? We can check it directly in very simple cases, and approximation methods such as Simpson's rule provide further evidence. However for a function which varies erratically it is not even clear what we should mean by the area under its graph. For example, we may ask what value should be assigned to

$$\int_0^1 \theta(x)\,\mathrm{d}x \quad \text{where} \quad \theta(x) = \begin{cases} 0 & \text{when } x \text{ is rational,} \\ 1 & \text{when } x \text{ is irrational.} \end{cases}$$

[Our answer, later, is 1; see 11.3.]

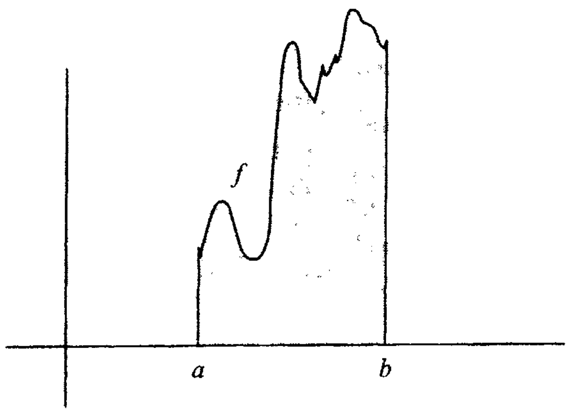


Figure 1.2

We aim to give a precise definition of integral which reconciles the ideas of 'integral as area' and 'integral as antiderivative' and embraces the everyday functions of elementary mathematics and many more exotic functions too. Our strategy is to begin with functions for which the area under the graph is defined in a universally accepted way and to build up in stages to more and more complicated functions, keeping in mind how we expect area to behave. We begin by discussing properties of area rather informally, to motivate the properties we shall demand of integrals. We shall not interrupt our presentation by defining every item of notation we use in this chapter. For anything unfamiliar, consult the notation index, or see Chapter 2, where much standard usage is recalled.

Before embarking we do need to introduce some notation we shall use repeatedly. In the standard way we write

$$[a,b] := \{\, x \in \mathbb{R} \mid a \leqslant x \leqslant b \,\} \quad (a \leqslant b).$$

Given any set S, we define its *characteristic function* (or *indicator function*), χ_S, by

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

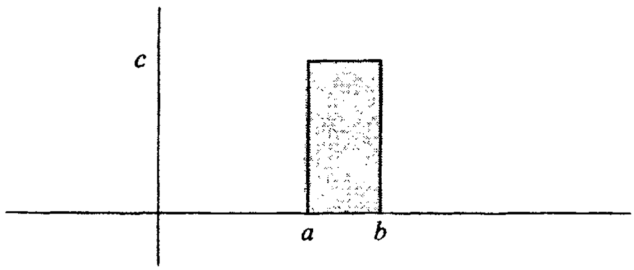
Here χ is the Greek letter chi. Thus $c\chi_{[a,b]}$ denotes the function taking the constant value c between a and b, and is zero elsewhere, while the function θ

above is $\chi_{[0,1]\setminus\mathbb{Q}}$. Our starting premise is simply that the area of a rectangle is its base multiplied by its height. This leads us to define

$$\int_a^b c\,\mathrm{d}x := c(b-a),$$

the rectangular area under the graph of the function $c\chi_{[a,b]}$; see Fig. 1.3. When $c < 0$ the graph lies below the x-axis, and our value for the integral is negative. Thus the areas we deal with are 'signed'. In general we should expect

$$\int_a^b f(x)\,\mathrm{d}x \geqslant 0 \quad \text{when } f(x) \geqslant 0 \text{ for } x \in [a,b] \qquad \text{(positivity)}.$$



$c\chi_{[a,b]}$

Figure 1.3

Our next observation is that if we take two rectangles which do not overlap then their combined area is the sum of their individual areas. So we should expect the integral of the function with the step-like graph shown in Fig. 1.4 to be the sum of the areas of the shaded rectangles. Also, assuming the Fundamental Theorem of Calculus works, we have for example,

$$\begin{aligned}
\int_0^x (\cos t - 3t^2)\,\mathrm{d}t &= \int_0^x \left(\frac{\mathrm{d}}{\mathrm{d}t}\sin t - \frac{\mathrm{d}}{\mathrm{d}t}t^3\right)\mathrm{d}t \\
&= \int_0^x \frac{\mathrm{d}}{\mathrm{d}t}(\sin t - t^3)\,\mathrm{d}t \qquad \text{(because differentiation is linear)} \\
&= \left[\sin t - t^3\right]_0^x \\
&= \sin x - x^3 \\
&= \int_0^x \cos t\,\mathrm{d}t - 3\int_0^x t^2\,\mathrm{d}t.
\end{aligned}$$

These observations lead us to believe that integration should satisfy:

$$\int_a^b (f(x) + \lambda g(x))\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \lambda\int_a^b g(x)\,\mathrm{d}x \qquad \text{(linearity)}.$$

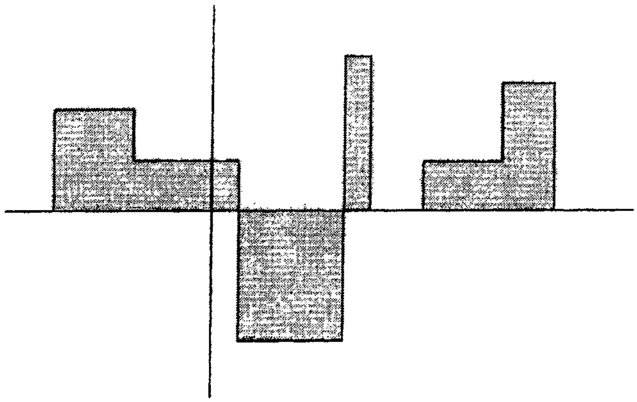


Figure 1.4

One other desirable feature of the integral emerges from considering area. If we take a rectangle and translate it sideways by a distance d, then the area is unchanged. More formally, we shift our origin to d, making the change of variable $y = x + d$. Then $x \in [a, b]$ if and only if $y \in [a + d, b + d]$ and

$$\int_a^b c \,\mathrm{d}x = c(b-a) = c((b+d)-(a+d)) = \int_{a+d}^{b+d} c \,\mathrm{d}y.$$

More generally we should want

$$\int_a^b f(x+d)\,\mathrm{d}x = \int_{a+d}^{b+d} f(y)\,\mathrm{d}y \qquad \text{(translation-invariance)}.$$

So far, we have looked at definite integrals in which the limits of integration are finite. Even in quite elementary applications this is too restrictive. For example, in statistics we meet the normal distribution and the claim that

$$\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x = \sqrt{\pi/2}.$$

Given any real-valued function f defined on a subset S of $\mathbb{R}$ (for example the interval $[a, b]$) we can extend f to a function defined on $\mathbb{R}$ by letting $f(x) = 0$ for $x \notin S$, and the areas under the graph of the original function and its extension will be the same. We therefore take functions $f\colon \mathbb{R} \to \mathbb{R}$ as the norm, restricting to subsets when appropriate. Thus the integral we want to define is of the form

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x,$$

which we shall write more often just as $\int f$. It is to be a real number, representing the area under the graph of f. We cannot expect this area to be finite, or even definable, for all functions $f\colon \mathbb{R} \to \mathbb{R}$. However we do want to define the integral on as large a class of functions as possible, while restricting the class sufficiently to ensure it has nice properties. There is a balancing act here, which explains why there is not a universally accepted way of developing integration theory. In the Lebesgue theory we demand that whenever f can be integrated, then so can its absolute value $|f|$. This means we preclude an area being called finite solely because of cancellation of areas above and below the y-axis (see Fig. 1.5). We should also like integrals to interact well with the limiting processes that abound in analysis. In this respect, Lebesgue integration scores heavily over other developments of integration.

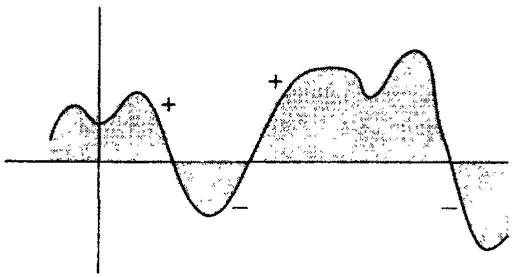


Figure 1.5

In summary, we shall require our real-valued integrable functions to have the Basic Properties listed below. [Together, properties (L) and (P) say that the class of (real-valued) integrable functions forms a real vector space on which $\int$ is a positive linear functional.]

Basic Properties

Building blocks:

(B) For any $a, b \in \mathbb{R}$ with $a \leqslant b$, the characteristic function $\chi_{[a,b]}$ is integrable and

$$\int \chi_{[a,b]} := (b-a).$$

Linearity:

(L) If f and g are integrable, and $\lambda \in \mathbb{R}$, then $f + \lambda g$ is integrable and

$$\int (f + \lambda g) = \int f + \lambda \int g.$$

Positivity:

(P) If f is integrable then

$$f(x) \geqslant 0 \text{ for all } x \Longrightarrow \int f \geqslant 0.$$

Modulus property:

(M) If f is integrable then so is $|f|$ (where $|f|(x) := |f(x)|$ for all x).

Translation-invariance:

(T) Let $d \in \mathbb{R}$. If f is integrable, then so is f_d, where $f_d(x) := f(x+d)$ for all x, and $\int f_d = \int f$.

We devote much of the first part of the book to constructing the class L of functions we shall designate as integrable. We prove, easily, the Fundamental Theorem of Calculus in the familiar form

$$\int_a^b F'(x)\,\mathrm{d}x = F(b) - F(a) \quad \text{if } F' \text{ is continuous on } [a,b].$$

Further, we prove powerful (and beautiful!) theorems about our class of integrable functions which allow us to manipulate integrals cleanly and rigorously. We refer in particular to the convergence theorems in Chapter 14, which are the centrepiece of the Lebesgue theory and its greatest asset.

Exercises

It would be inappropriate to set exercises on abstract vector spaces to students unversed in matrix algebra. In the same way, if your experience of integration so far is limited to elementary calculus, you may wish to defer the exercises below. They are aimed at those who have already studied some form of integration theory, either a Riemann-type integral or a simple integral like that presented in Chapters 3–5.

The exercises are designed to show that many properties of integrals follow directly from the Basic Properties. In doing them you should assume you are given a class of integrable functions which satisfy the Basic Properties, and no assumptions should be made about integration beyond these properties, except where indicated.

1.1 Prove that if $f(x) = 0$ for all except a finite number of values of x then f is integrable and $\int f = 0$.

1.2 Prove that the characteristic function χ_I of **any** bounded interval I with endpoints a, b $(a \leqslant b)$ is integrable, with $\int \chi_I = (b-a)$.

1.3 Prove that, given property (L), property (P) holds if and only if whenever f and g are integrable with $f(x) \geqslant g(x)$ for all x then $\int f \geqslant \int g$.

1.4 Let f be integrable. Prove, by considering the functions $|f| \pm f$, that

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

1.5 Prove that the following functions are **not** integrable:

(i) $\chi_{\mathbb{R}}$ (that is, the function taking the constant value 1),

(ii) the function f where

$$f(x) = \begin{cases} 1 & \text{if } x \geqslant 0, \\ -1 & \text{if } x < 0, \end{cases}$$

(iii) the 'sawtooth' function f which satisfies $f(x) = |x|$ if $-1 \leqslant x \leqslant 1$ and $f(x+2k) = f(x)$ for all x and all integers k.

[Hint: argue by contradiction.]

In the next two exercises you may assume that the series $\sum 1/k$ diverges.

1.6 Let $f(x) = x^{-1}$ for $0 < x \leqslant 1$ and $f(x) = 0$ otherwise. Let

$$\varphi_r(x) = \begin{cases} 1 & \text{if } 0 < x \leqslant 1/r, \\ 0 & \text{otherwise,} \end{cases}$$

for $r = 1, 2, \ldots$. Now define

$$\psi_n := \varphi_1 + \cdots + \varphi_n \qquad (n = 1, 2, \ldots).$$

(a) Sketch on one diagram the graphs of f, ψ_1, ψ_2, ψ_3.

(b) Show that $\psi_n \leqslant f$ for all n.

(c) Calculate $\int \psi_n$ for all n.

(d) Argue by contradiction to show that f is not integrable.

1.7 Let $f(x) = x^{-1} \sin x$ on $[\pi, \infty)$.

(a) Sketch the graphs of f and of $|f|$.

(b) Find a constant $C > 0$ (with C independent of k) such that it is possible to fit, above the x-axis and under the kth hump of the graph of $|f|$, a rectangular box of area C/k $(k = 1, 2, \ldots)$.

(c) Deduce that if $|f|$ were integrable then we would have

$$\int |f| \geqslant C \sum_{k=1}^{n} 1/k \quad \text{for each } n = 1, 2, \ldots.$$

(d) Argue by contradiction to conclude that f is not integrable.

In the last two exercises you will need to assume that if f is integrable then so is $f\chi_{[a,b]}$ for $a < b$; this will be true for the integral we later construct. We write $\int_a^b f$ for $\int f\chi_{[a,b]}$.

1.8 Prove that if f is integrable then, for $a < c < b$,

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

1.9 Let f be a function such that $f(x) = 0$ for $x < 0$, f is integrable, and f is such that $0 \leqslant x \leqslant y$ implies $f(x) \geqslant f(y) \geqslant 0$ (so that f is a non-negative decreasing function on $[0, \infty)$). For $k = 0, 1, \ldots$, let

$$u_k := \int_{2^k}^{2^{k+1}} f.$$

(a) By considering the partial sums, show that the series $\sum u_k$ converges and deduce that

$$\lim_{k \to \infty} u_k = 0.$$

(b) Show that $2^k f(2^{k+1}) \leqslant u_k$ for all k and that $xf(x) \leqslant 2^{k+1} f(2^k)$ if $2^k \leqslant x < 2^{k+1}$.

(c) Deduce that $\lim_{x \to \infty} xf(x) = 0$.

2 Preliminaries

We present below the fundamental notions which we use in a number of contexts, establish notation, and draw attention to features of particular importance for us. The contents of this chapter may be treated as a list of prerequisites for the remainder of the book. The early chapters draw only on those parts of this list which might be expected to be known by a reader familiar with elementary differential calculus but who has only limited experience of ε-δ analysis.

Mathematical analysis, and in particular integration theory, is founded on the real numbers. We need not discuss here how the real numbers are defined. We shall regard them as 'heaven-sent': a set $\mathbb{R}$ equipped with the familiar arithmetic operations $+$ and $\cdot$ of addition and multiplication, and an ordering, $\leqslant$, which allows us to think of $\mathbb{R}$ as a 'number line'. We shall draw, without being explicit, on the familiar properties below. These foundations seldom need detailed examination once in place.

2.1 The ordered field $\mathbb{R}$. We have the following properties.

(a) $(\mathbb{R}, +)$ is an abelian group:

- (A1) $a + (b + c) = (a + b) + c$ for all $a, b, c \in \mathbb{R}$;
- (A2) there exists an element, 0, which acts as additive identity, that is, $a + 0 = 0 + a = a$ for all $a \in \mathbb{R}$;
- (A3) for each $a \in \mathbb{R}$ there exists an element, $-a$ (the additive inverse of a), satisfying $a + (-a) = (-a) + a = 0$;
- (A4) $a + b = b + a$ for all $a, b \in \mathbb{R}$.

(b) $(\mathbb{R} \setminus \{0\}, \cdot)$ is an abelian group:

- (M1) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in \mathbb{R}$;
- (M2) there exists an element, 1, which acts as multiplicative identity, that is, $a \cdot 1 = 1 \cdot a = a$ for all non-zero $a \in \mathbb{R}$;
- (M3) for each non-zero $a \in \mathbb{R}$ there exists an element, $1/a$ (the multiplicative inverse of a) satisfying $a \cdot (1/a) = (1/a) \cdot a = 1$;
- (M4) $a \cdot b = b \cdot a$ for all $a, b \in \mathbb{R}$.

(c) Addition and multiplication are linked by the distributive law:

- (D) $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$ for all $a, b, c \in \mathbb{R}$.

We shall henceforth, as is usual, write ab rather than $a \cdot b$, and omit parentheses round products.

(d) Order properties. In the usual way we write interchangeably $a \leqslant b$ and $b \geqslant a$, write $a < b$ to mean $a \leqslant b$ and $a \neq b$, and work with either $<$ or $\leqslant$ depending on the context. Note that $a \leqslant b$ is true when $a < b$ (this causes puzzlement to those who forget that $a \leqslant b$ means $a = b$ **or** $a < b$). The order $\leqslant$ satisfies

(O1) $\leqslant$ is a total order: for all $a, b, c \in \mathbb{R}$, $a \leqslant a$ (reflexivity), $a \leqslant b$ and $b \leqslant a$ imply $a = b$ (antisymmetry), $a \leqslant b$ and $b \leqslant c$ imply $a \leqslant c$ (transitivity), and either $a \leqslant b$ or $b \leqslant a$ (trichotomy—just one of $a < b$, $a = b$, $b < a$ holds);

(O2) for all $a, b, c \in \mathbb{R}$, $a \leqslant b$ implies $a + c \leqslant b + c$;

(O3) for all $a, b, c \in \mathbb{R}$, $0 \leqslant c$ and $a \leqslant b$ imply $a \cdot c \leqslant b \cdot c$.

From the above it is easy to deduce that $a^2 \geqslant 0$ for all $a \in \mathbb{R}$, with equality if and only if $a = 0$; in particular $1 = 1^2 > 0$.

2.2 Concerning infinity. We introduce the usual symbols ∞ and $-\infty$ for positive and negative infinity. We stress that $\pm\infty$ **are not members of** $\mathbb{R}$. The arithmetic of $\mathbb{R}$ does not extend to embrace them: for example we must outlaw bogus statements such as $\infty - \infty = 0$. However, order-wise, it is often convenient to work with $\mathbb{R} \cup \{\pm\infty\}$, thinking of the symbols $-\infty, \infty$ as representing 'knobs' on the ends of the number line, so that $-\infty < a < \infty$ for all $a \in \mathbb{R}$.

2.3 Subsets of $\mathbb{R}$, and $\mathbb{C}$. We consider intervals in $\mathbb{R}$ in detail in Chapter 3 but have prior need of the following notation. We let

$$[a, b] := \{\, x \in \mathbb{R} \mid a \leqslant x \leqslant b \,\} \quad (-\infty < a \leqslant b < \infty),$$
$$(a, b) := \{\, x \in \mathbb{R} \mid a < x < b \,\} \quad (-\infty \leqslant a < b \leqslant \infty).$$

These are called respectively *compact* and *open*. We denote the interval $[0, \infty) := \{\, x \in \mathbb{R} \mid x \geqslant 0 \,\}$ also as $\mathbb{R}^+$ (the non-negative real numbers).

We adopt the standard symbols

$\mathbb{N}$ for the natural numbers, $1, 2, \ldots$,
$\mathbb{Z}$ for the integers, $\ldots, -1, 0, 1, 2, 3, \ldots$,
$\mathbb{Q}$ for the rational numbers, p/q ($p \in \mathbb{Z}$, $q \in \mathbb{N}$),
$\mathbb{C}$ for the complex numbers, $a + \mathrm{i}b$ ($a, b \in \mathbb{R}, \mathrm{i}^2 = -1$).

We may treat $\mathbb{R}$ as the subset $\{\, a + \mathrm{i}0 \mid a \in \mathbb{R} \,\}$ of $\mathbb{C}$, and note that $\mathbb{C}$ with the usual addition and multiplication satisfies the same arithmetic laws ((A1)–(A4), (M1)–(M4), (D)) as hold in $\mathbb{R}$. However we do **not** have an order relation on $\mathbb{C}$ satisfying (O1)–(O3). This is the reason that certain of our major theorems—notably the Monotone Convergence Theorem—apply to functions taking values in $\mathbb{R}$, with no direct extension to complex-valued functions.

It can be shown (see for example [**1**]) that every (non-empty) open interval contains a point of $\mathbb{Q}$ and a point of $\mathbb{R} \setminus \mathbb{Q}$ (in fact infinitely many such points).

The rational numbers, $\mathbb{Q}$, satisfy all the axioms for an ordered field ascribed to $\mathbb{R}$ in 2.1. Before we can distinguish between $\mathbb{Q}$ and the much richer structure $\mathbb{R}$ we need the notions of supremum and infimum.

2.4 Bounds; sup and inf. Let S be a non-empty subset of $\mathbb{R}$. We say $\alpha \in \mathbb{R}$ is an *upper bound* of S if $s \leqslant \alpha$ for all $s \in S$; if S possesses an upper bound we say S is *bounded above*. A real number α is called the *supremum* or *least upper bound* of S (written $\alpha = \sup S$) if

(i) $\alpha \geqslant s$ for all $x \in S$ (α is **an** upper bound), and

(ii) if $\beta \geqslant s$ for all $s \in S$ then $\beta \geqslant \alpha$ (α is the **least** upper bound).

Reversing the inequality signs we get the notions of *lower bound*, *bounded below*, and the *infimum* (*greatest lower bound*), $\inf S$, of a non-empty set S which is bounded below.

2.5 The Completeness Property of $\mathbb{R}$. This states that $\mathbb{R}$ satisfies the two equivalent conditions

(SUP) $\sup S$ exists for any non-empty subset S of $\mathbb{R}$ which is bounded above;

(INF) $\inf S$ exists for any non-empty subset S of $\mathbb{R}$ which is bounded below.

We shall henceforth adopt the convention $\sup S := \infty$ ($\inf S := -\infty$) for any set S which is not bounded above (below). Since ∞ and $-\infty$ are not bona fide members of $\mathbb{R}$ this is no more than a notational convenience.

Intuitively the Completeness Property says that the number line is a continuum, with no 'gaps'. By contrast, $\mathbb{Q}$ is deficient in the sense that certain suprema are missing. The most familiar example is provided by $S := \{ x \in \mathbb{Q} \mid x^2 \leqslant 2 \}$. We would have to have $(\sup S)^2 = 2$ (why?); because $\sqrt{2}$ is irrational this is impossible in $\mathbb{Q}$. In fact, in a constructive approach to the number systems, one way to build $\mathbb{R}$ from $\mathbb{Q}$ is by adding elements to serve as the missing suprema and infima—for a discussion of this process of completion see for example [**6**], and for a further discussion of completeness see Chapter 28.

2.6 Taking stock. In this book we explore just a small part of the fine structure of $\mathbb{R}$. All of these riches emanate from the assumption we have made about $\mathbb{R}$, namely the properties in 2.1 and the Completeness Property.

Having laid out our assumptions we next assemble some items of notation and record some elementary facts. Since our approach to integration relies heavily on the order structure of $\mathbb{R}$, order-theoretic notions will be important. It is an easy exercise to prove from the properties in 2.1 that multiplication by a negative number reverses inequalities: $c < 0$ and $a \leqslant b$ imply $b \cdot c \leqslant a \cdot c$ and in particular $a \leqslant b$ implies $-a \geqslant -b$. This interplay between $\leqslant$, $\geqslant$, and $-$ gives a left-right symmetry to the number line $\mathbb{R}$, resulting in many concepts having order-reversed analogues (upper bound and lower bound, for example), and many properties coming in mutually equivalent pairs (the two clauses of the Completeness Property, for example). Thus the preference we give in our theory to approximation from below, rather than from above, is little more than a matter of choice.

2.7 Max and min. The trichotomy property of $\leqslant$ (see (O1)) implies that we may define, for any $a, b \in \mathbb{R}$,

$$\max\{a,b\} := \begin{cases} a & \text{if } a \geqslant b, \\ b & \text{if } a < b, \end{cases} \qquad \min\{a,b\} := \begin{cases} b & \text{if } a \geqslant b, \\ a & \text{if } a < b. \end{cases}$$

We shall also write $\max\{a,b\}$ as $a \vee b$ and $\min\{a,b\}$ as $a \wedge b$.

2.8 Modulus and related notions. As usual we define the *modulus* or *absolute value* of $a \in \mathbb{R}$ by $|a| := a$ if $a \geqslant 0$ and $-a$ if $a < 0$. For $z = a + ib \in \mathbb{C}$, $|z| := (a^2 + b^2)^{1/2}$. Note that $|z|$ reduces to $|a|$ when z is real (that is, $b = 0$) and that $|z| = 0$ implies $z = 0$—a fact on which the analysis of complex-valued functions relies.

Observe that because $|z|$ is always real we can work with inequalities between moduli of complex numbers, even though, as we have already recalled, $\mathbb{C}$ itself is not ordered. We shall make use of the inequalities:

$$\begin{aligned} &(\Delta) \quad |a+b| \leqslant |a| + |b| \qquad (a, b \in \mathbb{C}), \\ &(\nabla) \quad \frac{1}{|c+d|} \leqslant \frac{1}{|c| - |d|} \qquad (c, d \in \mathbb{C}, |c| > |d|). \end{aligned}$$

The first of these is the well-known *triangle inequality*, the second is obtained from it by taking $a = c + d$ and $b = -c$ in (Δ). Both (Δ) and (∇) are regularly used to estimate integrals.

The modulus on $\mathbb{R}$ is linked to the max and min operations $\vee$ and $\wedge$ by

$$\begin{aligned} |a| &= a \vee (-a), \\ 2(a \vee b) &= a + b + |a - b|, \\ 2(a \wedge b) &= a + b - |a - b|. \end{aligned}$$

To check the last two formulae, consider in turn the cases $a \geqslant b$, $a < b$.

We also define, for any $a \in \mathbb{R}$, the *positive part* and *negative part* of a,

$$a^+ := a \vee 0 \quad \text{and} \quad a^- := (-a) \vee 0.$$

We can recover a and $|a|$ from these: $a = a^+ - a^-$, $|a| = a^+ + a^-$.

We also define, for $a \in \mathbb{C}$,

$$\operatorname{sgn} a = \begin{cases} |a|/a & \text{if } a \neq 0, \\ 0 & \text{if } a = 0. \end{cases}$$

When $a \in \mathbb{R}$, $\operatorname{sgn} a$ tags the sign of a: it is $+1$ when $a > 0$ and -1 when $a < 0$.

2.9 Concerning functions. Given sets A and B, a function $f \colon A \to B$ is simply an assignment to each point $a \in A$ a unique point $f(a)$ in B. Here A is the *domain of* f, denoted $\operatorname{dom} f$, and B the *codomain*. It is useful to write $\operatorname{im} f$ for its image, $\{ f(a) \mid a \in A \}$. We may have $\operatorname{im} f \subsetneqq B$. We also write $a \mapsto f(a)$ (for $a \in \operatorname{dom} f$) to indicate the action of f. Given any two functions f and g we may define the composite, $g \circ f$, by $(g \circ f)(x) := g(f(x))$, so long as $\operatorname{dom} g \subseteq \operatorname{im} f$. If $f\colon A \to B$ and $g\colon B \to A$ are such that $g(f(x)) = x$ for all $x \in A$ and $f(g(y)) = y$ for all $y \in B$ then we say g is the *inverse function* of f and write g as f^{-1}. We adopt the usual definitions of direct and inverse images: for any function $f\colon A \to B$,

$$f(C) := \{\, f(x) \in B \mid x \in C \,\} \quad (C \subseteq A),$$
$$f^{-1}(D) := \{\, x \in A \mid f(x) \in D \,\} \quad (D \subseteq B).$$

Note that $f^{-1}(D)$ is a well-defined set whether or not f has an inverse function.

The functions arising in elementary calculus are normally given by a single formula. However our definition does not require a **single** recipe for $f(x)$ throughout the domain, merely that a value $f(x)$ should be specified for each $x \in \operatorname{dom} f$. Thus the following are examples of functions $f\colon \mathbb{R} \to \mathbb{R}$.

$$f(x) := \mathrm{e}^x \sin(\tan x^3), \qquad f(x) := \begin{cases} \sin x & (x \geqslant 0), \\ \cos x & (x < 0), \end{cases}$$

$$f(x) := \begin{cases} 1/\sin x & \text{if } x \notin \{n\pi \mid n \in \mathbb{Z}\}, \\ 217 & \text{otherwise}, \end{cases} \qquad f(x) := \begin{cases} x & (x \in \mathbb{Q}), \\ -x & (x \notin \mathbb{Q}), \end{cases}$$

$$f(x) := n \quad \text{if } n \leqslant x < n+1 \quad (n \in \mathbb{Z}).$$

Note that the last one is a single function, not a family of functions, one for each n; its graph looks like an infinite staircase.

We take as known the properties of the trigonometric functions, sin, cos, tan, $\ldots$, and of the associated inverse functions. We use the notation $\tan^{-1}$ rather than arctan, and so on. We assume familiarity with logarithms and exponentials too, but also show how the properties of these functions come out of the theory we present; see 5.12, 8.16, and 8.17.

We shall be concerned almost exclusively with functions whose domain is $\mathbb{R}$ or a subset (usually an interval) in $\mathbb{R}$. We sometimes pretend that a function is allowed to have value ∞ or $-\infty$ in a few places, when we really mean that it is not defined there. This convention allows us to write function definitions like $f(x) = 1/x$ without having to explain what to do in the case $x = 0$. Since integration theory comfortably accommodates functions which misbehave or are undefined on 'negligible' sets, this causes no problems. See 11.1 and 12.9. To avoid excessive pedantry we shall allow, for example, $\sin x$ to denote both the sine function (the mapping $x \mapsto sinx$) and its value at the point $x \in \mathbb{R}$. For $\alpha \in \mathbb{R}$ we write α for the function taking value α at every point.

We shall often need to translate and to rescale functions. For this we adopt the following notation: given $f\colon \mathbb{R} \to \mathbb{R}$ and $d \in \mathbb{R}$ we define f_d and f^d by

$$f_d(x) := f(x+d) \quad \text{and} \quad f^d(x) := f(dx).$$

Of course, $f_0 = f$ and f^0 is just the constant function with value $f(0)$.

A function $f\colon S \to \mathbb{R}$ where $S \subseteq \mathbb{R}$ is *increasing* (*decreasing*) if $x \leqslant y$ implies $f(x) \leqslant f(y)$ ($f(x) \geqslant f(y)$), *monotonic* if it is increasing or decreasing, and *strictly increasing* if $x < y$ implies $f(x) < f(y)$.

One final definition: a function $f\colon \mathbb{R} \to \mathbb{R}$ (or $\mathbb{C}$) is said to be of *compact support* if there exists some compact interval $[a,b]$ such that $f = 0$ on $\mathbb{R} \setminus [a,b]$.

2.10 Combining and comparing functions. Let $f, g\colon S \to \mathbb{R}$ be functions, where S is some set. To add f and g we simply add their values at each point:

$$(\forall x \in S)\ (f+g)(x) := f(x) + g(x)$$

(if this formal definition does not look familiar just think what you mean by $x^3 + \cos x$). Note that $f + g$ is a function from S to $\mathbb{R}$, and that in defining it pointwise we make use of the addition on the codomain $\mathbb{R}$; the domain S could be any set whatsoever. In like manner we can define pointwise

$$|f|,\ f^+,\ f^-,\ \operatorname{sgn} f,\ fg,\ f/g,\ f \vee g,\ f \wedge g,$$

and so on. All of these are functions from S to $\mathbb{R}$, with the proviso that for f/g we require $g(x) \neq 0$ for all $x \in S$.

We order real-valued functions pointwise: $f \leqslant g$ means $f(x) \leqslant g(x)$ for all $x \in S$. In particular $g \geqslant 0$ means $g(x) \geqslant 0$ for all $x \in S$.

In applications, to Fourier analysis for example, we often encounter complex-valued functions. Given such a function $f\colon S \to \mathbb{C}$ we can write $f = \operatorname{Re} f + \mathrm{i} \operatorname{Im} f$, where $\operatorname{Re} f$ and $\operatorname{Im} f$ are real-valued functions. The functions $|f|$, fg, and f/g can be constructed when f and g are complex-valued, though f^+, f^-, $f \vee g$, and $f \wedge g$ cannot since they involve order.

2.11 Boundedness. A subset S of $\mathbb{R}$ or $\mathbb{C}$ is *bounded* if it is bounded above and below or, equivalently, if there exists a constant $M \in \mathbb{R}$ such that $|x| \leqslant M$ for all $x \in S$. A function $f\colon S \to \mathbb{R}$ or $\mathbb{C}$ is defined to be bounded if there exists $M \in \mathbb{R}$ such that $|f(x)| \leqslant M$ for all $x \in S$ (that is, $\operatorname{im} f$ is a bounded set). Thus e^{-x} is a bounded function on $\mathbb{R}^+$ (take $M = 1$). The function x^{-1} is unbounded on $S := \{\, x \in \mathbb{R} \mid 0 < x < \infty \,\}$ but bounded on each $S_\delta := \{\, x \in \mathbb{R} \mid \delta < x < \infty \,\}$ $(\delta > 0)$. We stress that for a function f to be bounded we need a universal ceiling M on $|f(x)|$ over the whole of $\operatorname{dom} f$. This is saying more than that the value $f(x)$ is finite at each individual point $x \in \operatorname{dom} f$.

2.12 Limit points. A point c is a *limit point* of a subset S of $\mathbb{R}$ if, given $\delta > 0$, there exists $y \in S$, $y \neq c$, such that $y \in (c - \delta, c + \delta)$. (The requirement '$y \neq c$' serves to prevent every point of S being a limit point of S by default.)

As examples of limit points we draw attention to

(a) the limit points of (a,b) or $[a,b]$ are the points in $[a,b]$ $(-\infty < a < b < \infty)$;

(b) every point of $\mathbb{R}$ is a limit point of $\mathbb{Q}$ and of $\mathbb{R} \setminus \mathbb{Q}$ (for a proof, see [1]).

Since our integrals will be defined by approximation, we shall make extensive use of sequences.

2.13 Sequences of real numbers. Informally, a real sequence is a string of real numbers labelled by the natural numbers: $a_1, a_2, a_3, \ldots$, written more compactly as $\{a_n\}$. Formally, $\{a_n\}$ is shorthand notation for a function $f\colon \mathbb{N} \to \mathbb{R}$ with $f(n) = a_n$. We need the definition of the *limit* of $\{a_n\}$: we say a_n tends to (or converges to) a in $\mathbb{R}$ as $n \to \infty$ if

$$(\forall \varepsilon > 0)(\exists N \in \mathbb{N})\; n \geqslant N \implies |a_n - a| < \varepsilon.$$

In this case we write $a_n \to a$ as $n \to \infty$ or $a = \lim a_n$. We say $a_n \to \infty$ if for any $K \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that $a_n > K$ for all $n \geqslant N$; convergence to $-\infty$ is defined likewise.

Let $a_n \to a$ and $b_n \to b$. If $a_n \leqslant c_n \leqslant b_n$ and $a = b$ then $c_n \to a$ too (**Sandwiching Lemma**). Also $a_n \leqslant b_n$ for all n implies $a \leqslant b$ but in general $a_n < b_n$ for all n does not imply $a < b$ (can you supply a counterexample?). Further,

$$\begin{array}{llll} ca_n \to ca \;(c \in \mathbb{R}), & a_n + b_n \to a + b, & a_n b_n \to ab, & a_n/b_n \to a/b \;\; (b \neq 0), \\ |a_n| \to |a|, & a_n \vee b_n \to a \vee b, & a_n \wedge b_n \to a \wedge b; & \end{array}$$

use the formulae in 2.7 to get the last two.

We note the fact that any convergent sequence is bounded: if $a_n \to a$, then there exists $M \in \mathbb{R}$ such that $|a_n| \leqslant M$ for all n. By choosing $\varepsilon = 1$ in the limit definition and using (Δ) we see we may take $M = \max\{|a_1|, \ldots, |a_{N-1}|, 1 + |a|\}$, where N is such that $|a_n - a| < 1$ for $n \geqslant N$.

2.14 Monotonic sequences. A convergent real sequence must be bounded, but not conversely: look at $\{(-1)^n\}$. Our approach to the Lebesgue integral in Chapter 9 onwards is based on a class of sequences which must converge. These are the bounded sequences which are also monotonic. A real sequence $\{a_n\}$ is *monotonic increasing (decreasing)* if $a_n \leqslant a_{n+1}$ $(a_n \geqslant a_{n+1})$ for all n.

2.15 Monotonic Sequence Theorem. A monotonic real sequence $\{a_n\}$ converges (to a limit $a \in \mathbb{R}$) if and only if it is bounded. Further, whether $\{a_n\}$ is bounded or not, $a_n \to a$, where

$$a = \begin{cases} \sup a_n & \text{if } \{a_n\} \text{ is increasing,} \\ \inf a_n & \text{if } \{a_n\} \text{ is decreasing;} \end{cases}$$

here $\sup a_n$ is interpreted as ∞ if $\{a_n\}$ is not bounded above and $\inf a_n$ as $-\infty$ if $\{a_n\}$ is not bounded below. We write $a_n \nearrow a$ $(a_n \searrow a)$ if $\{a_n\}$ is increasing (decreasing). As is usual we have suppressed set brackets when writing $\sup a_n$ and $\inf a_n$.

2.16 Theorems stemming from the Completeness Property. The Monotonic Sequence Theorem relies on the Completeness Property of $\mathbb{R}$. In its turn it may be used to prove further results which we shall need.

(1) **Subsequence Theorem**. Every bounded sequence $\{a_n\}$ in $\mathbb{R}$ has a subsequence $\{a_{n_k}\}$ $(n_1 < n_2 < \dots)$ which converges.

(2) **Cauchy Convergence Principle**. Let $\{a_n\}$ be a sequence in $\mathbb{R}$. Then $\{a_n\}$ converges if $\{a_n\}$ is a *Cauchy sequence* (that is, given $\varepsilon > 0$ there exists N such that $|a_m - a_n| < \varepsilon$ for all $m, n \geqslant N$).

(3) **Heine–Borel Theorem** (for intervals). Let $[a, b]$ be a compact interval in $\mathbb{R}$. Assume that $\{K_\lambda\}_{\lambda \in \Lambda}$ is a family of open intervals such that $[a, b] \subseteq \bigcup_{\lambda \in \Lambda} K_\lambda$. Then there exists a finite subset Λ_0 of Λ such that $[a, b] \subseteq \bigcup_{\lambda \in \Lambda_0} K_\lambda$.

The converse of (2) is also true, and easy to prove. An alternative route from the Completeness Property to the Cauchy Principle goes via the Bolzano–Weierstrass Theorem—the topological result that every infinite subset of $\mathbb{R}$ has a limit point. The Bolzano–Weierstrass Theorem and the Heine–Borel Theorem are equivalent. The latter enters into integration theory in a technical but crucial way in Chapters 9 and 10. For proofs of the equivalences between the above theorems, see for example **[16]**.

We base our account of integration on monotonic sequences, treating the Cauchy Principle as an ancillary tool. Other approaches to the Lebesgue integral take the Cauchy Principle as fundamental; see Appendix I.

2.17 Complex sequences. Sequences of complex numbers are handled in the same way as real sequences, the only qualification being the usual one that order-theoretic statements must be avoided, unless they involve only associated real numbers (moduli of complex numbers, for example). While the Monotonic Sequence Theorem does not hold for complex sequences, the Subsequence Theorem and the Cauchy Principle do. For a discussion of complex sequences see **[13]**, §§ 1.14–1.16.

2.18 Series. While sequences underpin the theory of integration, series arise principally in applications of it. Therefore we here merely recall a few salient points. For a real or complex sequence $\{a_n\}$, the series $\sum a_n$ is defined to converge if and only if $\{s_n\}$ converges, where $s_n := a_1 + \dots + a_n$ (the nth *partial sum*), and we write $\lim s_n$ as $\sum_{n=1}^{\infty} a_n$. A small irritation arises when we work simultaneously with a series and its sequence of partial sums. Because we must write $s_n = a_1 + \dots + a_n$ as $\sum_{k=1}^{n} a_k$ rather than as $\sum_{n=1}^{n} a_n$ to avoid using n is two different ways at the same time, we have used k, rather than n, to label the individual terms. Sometimes we want k to run from 0 rather than 1, or from some fixed $m \in \mathbb{N}$ (to avoid anomalous initial terms, which don't affect the long-run behaviour of the sequence). This calls for only minor notational adjustments, which we leave the reader to make where necessary.

Since $s_n = a_n - a_{n-1}$ for $n \geqslant 2$ we see that $a_k \to 0$ as $k \to \infty$ whenever $\sum a_k$ converges. For $\{a_k\}$ real, $\{s_k\}$ is monotonic if and only if the terms a_k are of constant sign. We draw attention to the following elementary results.

(1) **Geometric series.** $\sum r^k$ converges if and only if $|r| < 1$.

(2) **Absolute convergence.** $\sum a_k$ converges if $\sum |a_k|$ converges (but not conversely in general).

(3) **Comparison Test.** Suppose $0 \leqslant a_k \leqslant b_k$, where $\sum b_k$ converges. Then $\sum a_k$ converges.

By definition, $\sum a_k$ converges if $\{s_n\}$ converges, where $s_n := a_1 + \cdots + a_n$. In the other direction, it is sometimes convenient to establish convergence of a given sequence $\{s_n\}$ by constructing a series $\sum a_n$ of which $\{s_n\}$ is the sequence of partial sums, and applying one of the standard convergence tests to $\sum a_n$. We shall use this trick several times, but always in the same way. Here it is.

2.19 Telescoping Lemma. Let $\{s_n\}$ be a real or complex sequence. Suppose

$$|s_k - s_{k-1}| \leqslant b_k \quad (k \geqslant 2),$$

where $\sum b_k$ is convergent. Then $\{s_k\}$ converges. This happens in particular if $|s_k - s_{k-1}| \leqslant 2^{-k}C$, for some constant C.

Proof. For any $n \geqslant 2$, we have (by repeated use of (A1)!),

$$s_n = s_1 + (s_2 - s_1) + \cdots + (s_{n-1} - s_{n-2}) + (s_n - s_{n-1}).$$

But $\sum |s_k - s_{k-1}|$ converges by comparison with $\sum b_k$. Hence $\sum (s_k - s_{k-1})$ converges, so $\{s_n\}$ converges. $\square$

2.20 Limits of functions; continuity. Let $f\colon S \to \mathbb{R}$, where $S \subseteq \mathbb{R}$. We presume familiarity with the following ε-δ definitions.

(1) $f(x) \to L$ as $x \to c$ (where $c \in S$ or c is a limit point of S) if

$$(\forall \varepsilon > 0)(\exists \delta > 0)\ x \in S \text{ and } 0 < |x - c| < \delta \implies |f(x) - L| < \varepsilon.$$

This is the definition for $L \in \mathbb{R}$; the adaptation for $f(x) \to \infty$ or $-\infty$ is made in the obvious way, as for sequences.

(2) f is continuous at $c \in S$ if

$$(\forall \varepsilon > 0)(\exists \delta > 0)\ x \in S,\ |x - c| < \delta \implies |f(x) - f(c)| < \varepsilon,$$

and f is continuous (on S) if it is continuous at each $c \in S$.

In the above definitions δ may depend both on a and on ε. In (1), the restriction $0 < |x - c|$, which just says $x \neq c$, means that L is the limiting value of $f(x)$ as x **approaches** c, the value of f **at** c being immaterial, and possibly not even defined.

We use the standard notation $f(c+)$ and $f(c-)$ for the right- and left-hand limits of $f(x)$ at c.

All the definitions above make equally good sense when f is a function taking values in $\mathbb{C}$. It is easy to see that $\lim_{x\to a} f(x)$ exists if and only if $\lim_{x\to a} \operatorname{Re} f(x)$

and $\lim_{x\to a} \operatorname{Im} f(x)$ both exist. Consequently no essentially new features arise when we move from real-valued functions to complex-valued functions. The same proviso applies as for sequences: statements involving order do not lift verbatim to the complex setting.

The next result forms part of the technical apparatus for working with limits. The lemma which follows is in the same spirit.

2.21 Local Boundedness Lemma. Suppose $f(x) \to L$ as $x \to a$. Then there exist M and δ such that $|f(x)| \leqslant M$ for $0 < |x - a| < \delta$.

Proof. From the limit definition with $\varepsilon = 1$ we can choose $\delta > 0$ such that $|f(x) - L| < 1$ for $0 < |x - a| < \delta$. By (Δ), $M := |L| + 1$ will then serve. $\square$

2.22 Lemma. Let $f\colon S \to \mathbb{R}$ be continuous at a point c and let $f(c) > 0$. Then there exists $\delta > 0$ such that $f(x) > f(c)/2$ whenever $x \in S$, $|x - c| < \delta$.

Proof. Take $\varepsilon = f(c)/2$ in the continuity definition. There exists $\delta > 0$ such that for $x \in S$, $|x - c| < \delta$,

$$f(c)/2 > |f(x) - f(c)| \geqslant f(c) - f(x).$$

for $x \in S$, $|x - c| < \delta$; see Fig. 5.2 below. $\square$

We now turn to the global behaviour of a continuous function on a compact interval $[a, b]$ and state two fundamental theorems. We define

$$\mathrm{C}[a, b] := \{\, f\colon [a, b] \to \mathbb{R} \mid f \text{ is continuous on } [a, b] \,\} \quad (-\infty < a \leqslant b < \infty).$$

2.23 Boundedness Theorem. Let $f \in \mathrm{C}[a, b]$. Then f is bounded and f attains its sup and inf, that is, there exist α and β in $[a, b]$ such that

$$f(\alpha) = \sup\{\, f(x) \mid x \in [a, b]\} \quad \text{and} \quad f(\beta) = \inf\{\, f(x) \mid x \in [a, b]\}.$$

2.24 Intermediate Value Theorem. Let $f \in \mathrm{C}[a, b]$. Assume $f(a) < f(b)$. Then for any λ with $f(a) < \lambda < f(b)$ there exists $c \in [a, b]$ such that $f(c) = \lambda$.

Combining the two preceding results, we see that on $[a, b]$ a continuous function f attains every value between $m := \inf f$ and $M := \sup f$. In other words, $f([a, b]) = [m, M]$.

2.25 Differentiable functions. Since differentiation and integration are mutually inverse processes, we cannot study the latter without some technical facility with the former. This means that we presuppose the following.

- Ability to calculate the derivatives of the everyday functions—polynomials and rational functions, exponential and logarithmic functions, trigonometric and hyperbolic functions, and compound functions derived from these, and ability to recognize many of these derivatives, preferably without recourse to the table of derivatives in Appendix II.

- The definition of derivative as a limit, in formal ε-δ form (paralleling 2.20), and consequent results on combining differentiable functions, by pointwise arithmetic operations and by composition (the chain rule, $(g \circ f)'(x) = g'(f(x))f'(x)$).
- Rolle's Theorem and its retinue of consequences, which appear in standard calculus texts but cannot be rigorously proved without calling on the Boundedness Theorem, 2.23. For reference we lay out this array of tools below.

2.26 Theorems of differential calculus. Let $-\infty < a < b < \infty$.

(1) **Rolle's Theorem**. Let $f \in \mathrm{C}[a, b]$. Assume that f is differentiable on (a, b) and such that $f(a) = f(b)$. Then there exists $c \in (a, b)$ such that $f'(c) = 0$.

(2) **Mean Value Theorem** (**MVT**). Let $f \in \mathrm{C}[a, b]$ and assume that f is differentiable on (a, b). Then there exists c with $a < c < b$ such that

$$f(b) - f(a) = (b - a)f'(c).$$

(3) **Corollaries of the MVT**. Let $f\colon (a, b) \to \mathbb{R}$ be differentiable. Then

  (a) $f'(x) \geqslant 0$ for all $x \in (a, b)$ implies f is increasing;

  (b) $f'(x) = 0$ for all $x \in (a, b)$ implies f is a constant function.

(4) **L'Hôpital's rule**. Here let $-\infty \leqslant a < b \leqslant \infty$ and suppose $f, g\colon (a, b) \to \mathbb{R}$ are differentiable and such that $g'(x) \neq 0$ in (a, b). If $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$ then

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$

provided the limit on the right-hand side exists.

2.27 Smoothness. The more derivatives a function possesses, the smoother its graph will look and the better behaved it will be. Let $f\colon \mathbb{R} \to \mathbb{R}$. We say f is k-times continuously differentiable, and write $f \in \mathrm{C}^k(\mathbb{R})$, if the derivatives $f', f'', \ldots, f^{(k)}$ all exist and are continuous. We write $f \in \mathrm{C}^\infty(\mathbb{R})$ if $f \in \mathrm{C}^k(\mathbb{R})$ for all $k \geqslant 1$; this means that f has (necessarily continuous) derivatives of all orders. We define $\mathrm{C}^k[a, b]$ ($k \geqslant 1$) and $\mathrm{C}^\infty[a, b]$ in the same way, but with the proviso that derivatives at the endpoints a and b are taken to be one-sided derivatives.

Exercises

2.1 Check the formulae linking $|\ |$, $\vee$, $\wedge$, ${}^+$, and ${}^-$ from 2.7.

2.2 Show that it is impossible to find a partial order $\leqslant$ on $\mathbb{C}$ extending the usual order on $\mathbb{R}$ and such that for all $w, z \in \mathbb{C}$

  (i) $z = 0$ or $z > 0$ or $z < 0$, and

  (ii) $w, z > 0$ imply $w + z > 0$ and $wz > 0$.

2.3 For each of the following functions f, sketch the graph of f and the graphs of the functions $|f|$, f^+, f^-, $\operatorname{sgn} f$, f_d with $d = -\pi$, and f^d with $d = -2$:

(i) $f(x) = x^2 - 1$, (ii) $f(x) = \sin x$,

(iii) $f(x) = x^3 e^{-x}$, (iv) $f = \chi_{[-1,3]\cup[7,8]}$,

(v) $f = \chi_{[-1,2]} - 2\chi_{[1,3]}$, (vi) $f(x) = \begin{cases} x & (x \in \mathbb{Q}), \\ -x & (x \notin \mathbb{Q}). \end{cases}$

2.4 Sketch the graphs of $f \vee g$ and $f \wedge g$ for the following pairs of functions.
(i) $f(x) = e^x$ and $g(x) = e^{-x}$,
(ii) $f(x) = \sin x$ and $g(x) = \cos x$,
(iii) $f = \chi_{\mathbb{Q}}$ and $g = -\chi_{\mathbb{Q}}$.

2.5 Consider the examples of functions given in 2.9. Which are bounded?

2.6 Let $-\infty < a < b < \infty$ and assume that $f\colon [a,b] \to \mathbb{R}$ is an increasing function. Prove that f is bounded. Does this remain true if $[a,b]$ is replaced by (a,b)?

2.7 Let $C_0(\mathbb{R})$ denote the set of continuous functions $f\colon \mathbb{R} \to \mathbb{C}$ such that $f(x) \to 0$ as $x \to \pm\infty$. Prove that every function in $C_0(\mathbb{R})$ is bounded. [Compare with the result that a convergent sequence is bounded.]

2.8 Show that $f\colon \mathbb{R} \to \mathbb{R}$ is continuous if and only if f^+ and f^- are both continuous. For which functions f is it true that f^+ and f^- are both differentiable?

2.9 Let $\sum a_n$ be an absolutely convergent series of real or complex numbers. Prove that

$$\left|\sum_{n=1}^{\infty} a_n\right| \leqslant \sum_{n=1}^{\infty} |a_n|.$$

2.10 (a) Let $0 < \alpha < 1$ and define $f(x) := \alpha x - x^\alpha$. By applying the MVT to f show that $f(x) \geqslant f(1)$ for all $x \geqslant 0$, with equality if and only if $x = 1$. Hence show that

$$a^\alpha b^{1-\alpha} \leqslant \alpha a + (1-\alpha) b \quad (a, b \geqslant 0).$$

(b) Let p and q be such that $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$, and let A and B be non-negative real numbers. Deduce from (a), with $\alpha = 1/p$, that

$$AB \leqslant \frac{A^p}{p} + \frac{B^q}{q},$$

with equality if and only if $A^p = B^q$. [This inequality is needed in the theory of L^p-spaces; see Chapter 28.]

2.11 On $[0, \pi/2]$ define

$$f(x) = \begin{cases} \operatorname{cosec} x - x^{-1} & (x \neq 0), \\ 0 & (x = 0), \end{cases} \qquad g(x) = \begin{cases} \cot x - x^{-1} & (x \neq 0), \\ 0 & (x = 0). \end{cases}$$

Use L'Hôpital's Rule to prove that f and g belong to $\mathrm{C}[0, \pi/2]$. Prove further that $f \in \mathrm{C}^1[0, \pi/2]$.

3 Intervals and step functions

According to our manifesto in Chapter 1, any finite linear combination of characteristic functions of bounded intervals must be integrable. We now investigate these functions in detail. They have step-like graphs, as in Fig. 1.4, and serve as approximations to more familiar continuous functions, as shown in Fig. 3.1.

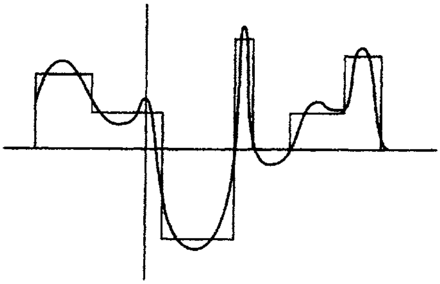


Figure 3.1

3.1 Intervals. A non-empty subset I of $\mathbb{R}$ is an *interval* if whenever $p, q \in I$ with $p \leqslant q$ then $p \leqslant x \leqslant q$ implies $x \in I$. (The exclusion of the empty set does us no harm, and saves us having to worry about this as a special case in statements and proofs.) An interval I takes one of the following forms:

$$\begin{array}{rlll} (a,b) :=& \{\, x \in \mathbb{R} \mid a < x < b \,\} & (-\infty \leqslant a < b \leqslant \infty) & (\text{an } \textit{open interval}), \\ [a,b] :=& \{\, x \in \mathbb{R} \mid a \leqslant x \leqslant b \,\} & (-\infty < a \leqslant b < \infty) & (\text{a } \textit{closed interval}), \\ (a,b] :=& \{\, x \in \mathbb{R} \mid a < x \leqslant b \,\} & (-\infty \leqslant a < b < \infty), & \\ [a,b) :=& \{\, x \in \mathbb{R} \mid a \leqslant x < b \,\} & (-\infty < a < b \leqslant \infty). & \end{array}$$

In each case the points a and b are the *endpoints* of I. The endpoints are allowed to coincide only for a closed interval; $[a, a]$ contains the single point a. The interval I is *bounded* if and only if $a \neq -\infty$ and $b \neq \infty$. Note that neither ∞ nor $-\infty$, which are not members of $\mathbb{R}$, belongs to an interval of any type. Bounded closed intervals are called *compact*.

We say an interval J is a *subinterval* of an interval I if $J \subseteq I$. Let I_1 and I_2 be intervals. The intersection $I_1 \cap I_2$ is an interval if and only if it is non-empty. The union $I_1 \cup I_2$ is an interval if and only if $I_1 \cap I_2 \neq \varnothing$. For example, $S := [-2, 3] \cup (6, 8]$ is not an interval ($3 \in S$, $7 \in S$, but $3 \leqslant 4 \leqslant 7$ and $4 \notin S$). On the other hand,

$$(1,4) \cap [2,7] = [2,4) \quad \text{and} \quad (1,4) \cup [2,7] = (1,7],$$

and these are intervals.

Intervals I and J are said to be *disjoint* if $I \cap J = \varnothing$, and a union of a finite or infinite collection $\{I_r\}$ of intervals is a *disjoint union* if $I_r \cap I_s = \varnothing$ for $r \neq s$. We write $\dot\cup$ in place of $\cup$ when we want to stress that we are dealing with a disjoint union. Note that, for example, the intervals $[-3,2)$ and $[2,8]$ are disjoint but that $[-3,2]$ and $[2,8]$ do not qualify as disjoint because of the shared endpoint 2. This can be a minor irritation; see 9.2.

3.2 The length of a bounded interval. Given a bounded interval I with endpoints a and b ($a \leqslant b$) we define the *length* of I to be

$$\ell(I) := b - a.$$

This is always non-negative, and is zero if and only if $I = [a,b]$ with $a = b$.

Observe that the length of I is the same whichever type of interval I is, that is, whether or not it contains one or both of its endpoints. For this reason, it is convenient to adopt the notation $\langle a, b\rangle$ for a given interval of any type whose endpoints are a and b.

Clearly $\ell(J) \leqslant \ell(I)$ whenever J is a subinterval of I. Suppose now that I is the union of two subintervals I_1 and I_2. We should expect that $\ell(I) = \ell(I_1) + \ell(I_2)$ if I_1 and I_2 are disjoint, but that $\ell(I_1) + \ell(I_2)$ will generally give an over-estimate of $\ell(I)$ if I_1 and I_2 overlap. For example,

$$(-1,4] = (-1,1] \,\dot\cup\, (1,4] = (-1,1] \cup [1,4] = (-1,3) \cup (1,4],$$

$\ell((-1,4]) = 5$, and

$$\begin{aligned} \ell((-1,1]) + \ell((1,4]) &= 2+3 = 5,\\ \ell((-1,1]) + \ell([1,4]) &= 2+3 = 5,\\ \ell((-1,3)) + \ell((1,4]) &= 4+3 > 5. \end{aligned}$$

Lemma 3.3 gives the generalities which underlie this behaviour. Note that we have defined the length of a set S **only** when S is a single bounded interval. We consider more general sets later; compare 3.3 with 4.2 and 22.5.

3.3 Lemma. Let $I = \langle a, b\rangle$ be a bounded interval ($a < b$).

(a) Assume that I is the union of subintervals $I_1 = \langle a, c\rangle$ and $I_2 = \langle c, b\rangle$. Then $\ell(I) = \ell(I_1) + \ell(I_2)$.

(b) Assume that I is the disjoint union of subintervals $I_1, \ldots, I_n$ ($n \geqslant 2$). Then $\ell(I) = \ell(I_1) + \cdots + \ell(I_n)$.

(c) Assume that I is the union of non-disjoint subintervals I_1 and I_2. Then

$$\ell(I_1 \cup I_2) = \ell(I_1) + \ell(I_2) - \ell(I_1 \cap I_2) \leqslant \ell(I_1) + \ell(I_2).$$

Proof. By definition,

$$\ell(\langle a, b\rangle) = b - a = (c-a) + (b-c) = \ell(\langle a, c\rangle) + \ell(\langle c, b\rangle).$$

This proves (a). In case the common endpoint c belongs to exactly one of I_1 and I_2 we have the case $n = 2$ in (b). The general case of (b) is then easily proved by induction.

For (c) we have already covered in (a) the case when $I_1 \cap I_2$ is a single point, and the required result is obvious if $I_1 \subseteq I_2$ or $I_2 \subseteq I_1$. So suppose, without loss of generality, that $I_1 = \langle a, b \rangle$ and $I_2 = \langle \alpha, \beta \rangle$, where $-\infty < a \leqslant \alpha < b \leqslant \beta < \infty$. Then $I_1 \cup I_2$ takes the form $\langle a, \beta \rangle$ and $I_1 \cap I_2$ the form $\langle \alpha, b \rangle$. Hence

$$\ell(I_1 \cup I_2) = (\beta - a) = (\beta - \alpha) + (b - a) - (b - \alpha) = \ell(I_1) + \ell(I_2) - \ell(I_1 \cap I_2),$$

as required. $\square$

3.4 Step functions. In a sketch graph of $\chi_{(a,b]}$ an open circle indicates that the value at a is 0, while a solid circle indicates that the value at b is 1; see Fig. 3.2. We shall adopt the same conventions for characteristic functions of other intervals of fixed type; for $\chi_{\langle a,b \rangle}$ we shall draw the graph with no circles.

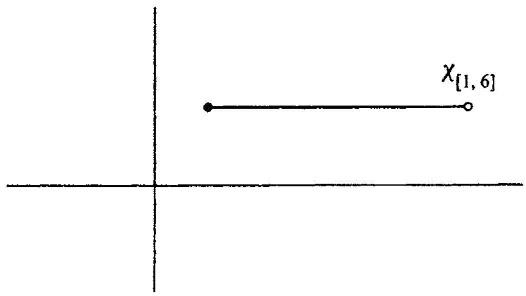


Figure 3.2

Figure 3.3 depicts the graph of φ, where

$$\varphi := 4\chi_{(-2,1)} + \chi_{(1,2)} - 3\chi_{[3,4]} + \chi_{(4,27/5]}.$$

The area under the graph is a disjoint union of rectangles, of total (signed) area

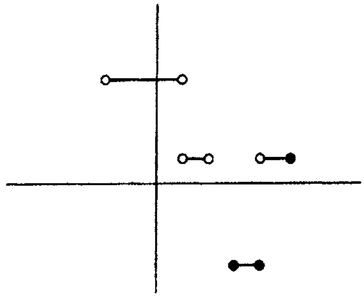
$$(4 \times 3) + (1 \times 1) - (3 \times 1) + (1 \times \tfrac{7}{5}) = 4\ell((-2,1)) + \ell((1,2)) - 3\ell([3,4]) + \ell((4, \tfrac{27}{5}]).$$

A function $\varphi \colon \mathbb{R} \to \mathbb{R}$ is a *step function* if φ can be represented as

$$\varphi = \sum_{r=1}^{n} c_r \chi_{I_r},$$

where $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{R}$, and $I_1, \ldots, I_n$ are bounded intervals. We are going to define the integral of φ to be

$$\int \varphi := \sum_{r=1}^{n} c_r \ell(I_r).$$



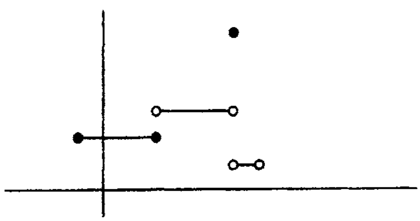
$4\chi_{(-2,1)} + \chi_{(1,2)} - 3\chi_{[3,4]} + \chi_{(4,27/5]}$

Figure 3.3

We do not insist that the intervals $I_1, \ldots, I_n$ are disjoint, as they were in our illustrative example. Disjointness is not necessary to get a step-like graph. For example, Fig. 3.4 shows the graph of

$$\psi := 2\chi_{[-1,5]} + \chi_{(2,5]} + \chi_{[5,6)}.$$

Note that $\chi_{\langle a,b\rangle}$ is a step function whenever $\langle a,b\rangle$ is bounded. In particular $\chi_{[a,b]}$ is a step function for any compact interval $[a,b]$, and our definition above implies that its integral is $(b-a)$, in accordance with the Building Blocks Property (B) in Chapter 1.



$2\chi_{[-1,5]} + \chi_{(2,5]} - \chi_{[5,6)}$

Figure 3.4

The reason for certain features of the definition of a step function emerges from the definition of the integral we have proposed for $\sum_{r=1}^{n} c_r \chi_{I_r}$. Each term in the sum $\sum_{r=1}^{n} c_r \ell(I_r)$ is **finite**, because the intervals are assumed to be bounded, and we have, by assumption, **finitely many terms.** Thus $\int \varphi$ is a real number. There is however one possible problem. The same step function may be represented in different ways: for example, the function ψ above is equal at each point to θ, where

$$\theta := 2\chi_{(-1,2]} - \chi_{[3,4)} + 3\chi_{(2,4)} + \chi_{[3,6)} + \chi_{\{5\}};$$

sketch the graph of θ to convince yourself. The formula for the integral gives

$$\int\psi = (2 \times (4 - (-1))) + (1 \times (5 - 2)) + (1 \times (6 - 5)) = 14,$$

$$\int\theta = (2 \times (2 - (-1))) - (1 \times (4 - 3)) + (3 \times (4 - 2)) + (1 \times (6 - 3)) + (1 \times (5 - 5)) = 14,$$

so—fortunately—we have the same value for either representation. Later, in 4.1, we validate the definition of the integral of a step function by proving in general that different representations give the same integral.

Manipulating step functions can be a messy business, though the underlying ideas are easy. We shall introduce some technical tricks which streamline our arguments later on. You may like to do Exercise example 3.6 in parallel with reading the next section.

3.5 Properties of step functions. Given a step function

$$\varphi = \sum_{r=1}^{n} c_r \chi_{I_r}, \quad \text{where } I_r = \langle a_r, b_r \rangle \ (r = 1, \ldots, n),$$

let

$$\alpha_0 < \alpha_1 < \cdots < \alpha_m$$

be the points of the set $\{a_1, \ldots, a_n, b_1, \ldots, b_n\}$, which contains the endpoints of all the intervals $I_1, \ldots, I_n$. Then

(S1) φ is constant on each interval (α_{i-1}, α_i) $(i = 1, \ldots, m)$, and

(S2) $\varphi = 0$ outside the bounded interval $[\alpha_0, \alpha_m]$.

Further,

(S3) φ takes only a finite number of values (and so is bounded), and

(S4) φ has at most a finite number of discontinuities.

To confirm (S1) note that for each i and each r, either (α_{i-1}, α_i) is contained wholly in I_r or fails to intersect it. Thus for all $x \in (\alpha_{i-1}, \alpha_i)$ the value of $\varphi(x)$ is the sum of those numbers c_r for which the corresponding interval I_r contains (α_{i-1}, α_i). As an example, take φ to be the step function θ defined in the preceding section. We have $\alpha_0 = -1$, $\alpha_1 = 2$, $\alpha_2 = 3$, $\alpha_3 = 4$, $\alpha_4 = 5$.

Properties (S1) and (S2) characterize step functions: we may alternatively define a step function to be a function φ such that (S1) and (S2) hold, for some α_i $(i = 0, \ldots, m)$ with $-\infty < \alpha_0 < \cdots < \alpha_m < \infty$. Any such set of points is called a *partition* for φ, which is said to be a step function on $[\alpha_0, \alpha_m]$. Of course, the distinct endpoints of the intervals $I_1, \ldots, I_n$ arranged in increasing order always give a partition for $\sum_{r=1}^{n} c_r \chi_{I_r}$ (the *standard partition*).

Property (S3) is immediate: the values of $\sum_{r=1}^{n} c_r \chi_{I_r}$ are $0, c_1, \ldots, c_n$, if the intervals I_r are disjoint, and finite sums of these values in general. For (S4) note that φ is continuous except possibly at the points of its standard partition. Remember that our definition of integral implies that $\int \chi_{\langle a,b \rangle}$ is the same whatever type of interval we are considering. In particular the values of a step function φ at its discontinuity points can be changed without affecting $\int \varphi$.

Properties (S2)–(S4) are useful for deciding that a given function cannot be (represented as) a step function. By (S3), a continuous function on $\mathbb{R}$ is a step function only if it is identically zero (invoke the Intermediate Value Theorem). Also, for example, none of the following is a step function:

(i) $\chi_{[0,\infty)}$ ((S2) cannot hold),

(ii) $\chi_{\mathbb{Q}\cap[0,1]}$ (discontinuous at every point of $[0,1]$),

(iii) $\displaystyle\sum_{r=1}^{\infty} r\chi_{(1/(r+1),1/r]}$ (both (S3) and (S4) fail).

Sketch the graphs to check the assertions.

We can now present our technical devices.

Disjunction. Given a step function $\varphi = \sum_{r=1}^{n} c_r\chi_{I_r}$, let $\alpha_0 < \alpha_1 < \cdots < \alpha_m$ be the standard partition. Let $K_1, \ldots, K_{2m+1}$ be the family of intervals $(\alpha_0,\alpha_1), \ldots, (\alpha_{m-1},\alpha_m)$, $[\alpha_0,\alpha_0]$, $[\alpha_1,\alpha_1]$, $\ldots$, $[\alpha_m,\alpha_m]$. Then we can re-express φ as $\varphi = \sum_{i=1}^{2m+1} b_i\chi_{K_i}$. Here some of the numbers b_i may be zero. The important point is that the intervals $K_1, \ldots, K_{2m+1}$ are **disjoint**, whereas the original intervals $I_1, \ldots, I_r$ need not be.

Refinement. Assume that $-\infty < \alpha_0 < \cdots < \alpha_m < \infty$ is a partition for a step function φ, so φ is constant on the open intervals between the points. If we add extra division points this remains true. Hence $-\infty < \beta_0 < \cdots < \beta_p < \infty$ is also a partition for φ whenever $\{\alpha_0, \ldots, \alpha_m\} \subseteq \{\beta_0, \ldots, \beta_p\}$; we call it a *refinement* of the original partition. The disjunction process is an instance of refinement.

Refinement is the key to the formal proof of the consistency of the integral definition. It is also invaluable when we wish to consider sums and other combinations of step functions, as follows.

- Given two different representations, $\sum_{r=1}^{n} c_r\chi_{I_r}$ and $\sum_{s=1}^{p} d_s\chi_{J_s}$, for a step function φ, we can use a common refinement of the standard partitions to re-express φ as $\sum_{t=1}^{q} e_t\chi_{K_t}$, where each K_t lies wholly inside some I_r and wholly inside some J_s (and the intervals $K_1, \ldots, K_q$ may be taken to be disjoint, by disjunction).
- Given any two step functions φ and ψ, we may find bounded intervals $K_1, \ldots, K_q$ (which may be taken to be disjoint) such that

$$\varphi = \sum_{t=1}^{q} e_t\chi_{K_t} \quad \text{and} \quad \psi = \sum_{t=1}^{q} f_t\chi_{K_t},$$

for some $e_1, \ldots, e_q, f_1, \ldots, f_q \in \mathbb{R}$.

3.6 Exercise example. Work through as many of the following exercises as you need to convince yourself that the assertions in 3.5 are true.

Let φ, ψ, and θ be defined by

$$\varphi := 2\chi_{[-2,2]} - \chi_{(0,4)} + \chi_{(4,5)}, \qquad \psi := -\chi_{(0,1]} - \chi_{[1,5)},$$
$$\theta := 3\chi_{\{0\}} + \chi_{[0,3]} + 4\chi_{(1,4)} - \chi_{(1,6]}.$$

For each of φ, ψ, and θ

(a) sketch the graph;
(b) identify its points of discontinuity;
(c) write down the standard partition;
(d) express the function as a linear combination of characteristic functions of disjoint bounded intervals.

Find a common partition for

(i) φ and ψ, (ii) φ and θ, (iii) ψ and θ.

In each case express the pair of step functions in terms of this partition.

In what follows, remember what was said in 2.9 and 2.10 about functions, and in particular about the way functions are combined, pointwise. We let $\mathrm{L}^{\mathrm{step}}$ denote the set of all step functions.

3.7 Combinations of step functions. Let $\varphi, \psi \in \mathrm{L}^{\mathrm{step}}$ and $\lambda \in \mathbb{R}$. Then

(a) $\varphi + \lambda\psi$ is a step function (and $\mathrm{L}^{\mathrm{step}}$ is a vector space);
(b) $\varphi\psi$ is a step function;
(c) $\varphi \vee \psi$, $\varphi \wedge \psi$, and $|\varphi|$ are step functions.

Proof. By refinement, we may represent φ and ψ as

$$\varphi = \sum_{t=1}^{q} e_t \chi_{K_t} \quad \text{and} \quad \psi = \sum_{t=1}^{q} f_t \chi_{K_t},$$

using the same disjoint intervals for both. Then

$$\varphi + \lambda\psi = \sum_{t=1}^{q} e_t \chi_{K_t} + \lambda \sum_{t=1}^{q} f_t \chi_{K_t} = \sum_{t=1}^{q} (e_t + \lambda f_t) \chi_{K_t},$$

which belongs to $\mathrm{L}^{\mathrm{step}}$. We prove (b) likewise. For (c), note that

$$\varphi \vee \psi = \sum_{t=1}^{q} (e_t \vee f_t) \chi_{K_t} \quad \text{and} \quad \varphi \wedge \psi = \sum_{t=1}^{q} (e_t \wedge f_t) \chi_{K_t},$$

so that both $\varphi \vee \psi$ and $\varphi \wedge \psi$ belong to $\mathrm{L}^{\mathrm{step}}$. Similarly, $|\varphi| \in \mathrm{L}^{\mathrm{step}}$ because

$$|\varphi| = \sum_{r=1}^{n} |c_r| \chi_{I_r}. \qquad \square$$

3.8 Exercise example 3.6 continued. Express as a step function

(i) $|\varphi|$, (ii) ψ^3, (iii) $\varphi + \psi$, (iv) $\psi - \theta$, (v) $\varphi \vee \theta$.

Exercises

3.1 Let I and J be bounded intervals. Express χ_S as a step function when S is

$$\text{(i) } I \cap J, \quad \text{(ii) } I \cup J, \quad \text{(iii) } I \setminus J, \quad \text{(iv) } I \triangle J.$$

[For sets S and T, the *symmetric difference* $S \triangle T$ of S and T is $(S \cup T) \setminus (S \cap T)$.]

3.2 State in which of the following cases f is a step function:

$$\text{(i) } f = \chi_{[1,\infty)} + \chi_{(-\infty,-1)}, \qquad \text{(ii) } f = \chi_{[0,\infty)} - \chi_{(1,\infty)},$$

$$\text{(iii) } f(x) = \begin{cases} 1/x & \text{if } x \in \mathbb{N}, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{(iv) } f(x) = \begin{cases} x & \text{if } 1/x \in \mathbb{N}, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{(v) } f = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \chi_{(0,1/k)}, \qquad \text{(vi) } f(x) = \operatorname{sgn}(\cos x).$$

For those which are not, give a reason or reasons.

3.3 Let $[x]$ denote the largest integer n such that $n \leqslant x$. Sketch the graph of $[x]$. Which of the following are step functions?

$$\text{(i) } [x]\chi_{[-100,100]}(x), \qquad \text{(ii) } [x], \qquad \text{(iii) } [\sin x], \qquad \text{(iv) } [x^{-1}]\chi_{(0,\infty)}(x).$$

3.4 Let φ be a step function such that $\varphi \geqslant 0$. Prove that $\sqrt{\varphi}$ is a step function.

3.5 Let $\varphi \in \mathrm{L}^{\mathrm{step}}$. Prove that φ is continuous at x if and only if there exists an open interval $I(x)$ containing x such that φ is constant on $I(x)$.

3.6 Let $\varphi \in \mathrm{L}^{\mathrm{step}}$ and let $\varepsilon > 0$. Show that there exists a continuous function $g \colon \mathbb{R} \to \mathbb{R}$ such that $g = \varphi$ except on a finite union of intervals, $J_1, \ldots, J_n$, with $\sum_{r=1}^{n} \ell(J_r) < \varepsilon$. How small can $\sup |\varphi - g|$ be made?

3.7 Suppose that $F \colon \mathbb{R} \to \mathbb{R}$ is a differentiable function. Prove that if $F' \in \mathrm{L}^{\mathrm{step}}$ then F is constant.

3.8 Let $\varphi, \psi \in \mathrm{L}^{\mathrm{step}}$.
  (a) Prove that the composite function $\varphi \circ \psi$ belongs to $\mathrm{L}^{\mathrm{step}}$.
  (b) Let $f(x) = \sin x$.
    (i) Under what conditions on φ is $\varphi \circ f$ a step function?
    (ii) Under what conditions on φ is $f \circ \varphi$ a step function?

3.9 Let $\varphi : \mathbb{R} \to \mathbb{R}$ and define

$$S^c := \{ x \in \mathbb{R} \mid \varphi(x) \geqslant c \} \quad (c \in \mathbb{R}).$$

  (a) Prove that if $\varphi \in \mathrm{L}^{\mathrm{step}}$ then S^c is a finite union of intervals for each $c \in \mathbb{R}$.
  (b) Assume that S^c is a finite union of bounded intervals for each $c > 0$. Prove that $\varphi \in \mathrm{L}^{\mathrm{step}}$.

4 Integrals of step functions

As proposed in Chapter 3 we define the integral of a step function by

$$\int \varphi := \sum_{r=1}^{n} c_r \ell(I_r) \quad \text{when} \quad \varphi = \sum_{r=1}^{n} c_r \chi_{I_r}.$$

We have already pointed out that this gives a finite value for the integral. Before proceeding we must pay our debts and check consistency of the definition. In part, what this consistency result amounts to geometrically is simply this: if we have a bounded region of the plane with its boundary made up of a finite number of horizontal and vertical lines and we calculate its area by breaking it into rectangles, then we get the same answer irrespective of the way we divide it up. Trusting readers may skip the formal proof given below.

4.1 Consistency lemma for the $\mathrm{L}^{\mathrm{step}}$ integral. Assume that

$$\sum_{r=1}^{p} c_r \chi_{I_r} \quad \text{and} \quad \sum_{s=1}^{p'} c'_s \chi_{I'_s}$$

are two representations of the same step function, φ. Then

$$\sum_{r=1}^{p} c_r \ell(I_r) = \sum_{s=1}^{p'} c'_s \ell(I'_s).$$

Proof. We prove the result by reduction to easier special cases.

Stage 1. We may assume without loss of generality that all $c_r \neq 0$ and all $c'_s \neq 0$—terms involving zero constants make no contribution to the integral.

Stage 2. We shall prove that $\sum_{r=1}^{p} c_r \ell(I_r) = \sum_{t=1}^{q} a_t \ell(K_t)$, where $\sum_{t=1}^{q} a_t \chi_{K_t}$ is another representation of φ, of our choosing, in which the intervals $\{K_t\}$ are disjoint..

By refinement we can find a representation $\sum_{t=1}^{q} a_t \chi_{K_t}$ of φ such that

(i) $K_1, \ldots, K_q$ are disjoint bounded intervals (non-empty by definition),

(ii) $a_t \neq 0$ for all t,

(iii) for each t and r, either $K_t \subseteq I_r$ or $K_t \cap I_r = \varnothing$, and

(iv) $\bigcup_t K_t = \bigcup_r I_r$.

Then

$$I_r = \bigcup \{ K_t \mid t \in T_r \} \quad \text{where } T_r := \{ t \mid K_t \subseteq I_r \}.$$

Now let $R_t := \{ r \mid K_t \subseteq I_r \}$; of course $r \in R_t$ if and only if $t \in T_r$. Then

$$a_t = \sum_{r \in R_t} c_r,$$

since each side equals $\varphi(x)$ where $x \in K_t$. Therefore, using arithmetic properties of $\mathbb{R}$ from 2.1,

$$\begin{aligned}
\sum_{t=1}^{q} a_t \ell(K_t) &= \sum_{t=1}^{q} \Big(\sum_{r \in R_t} c_r \Big) \ell(K_t) \\
&= \sum_{t=1}^{q} \Big(\sum_{r \in R_t} c_r \ell(K_t) \Big) \qquad \text{(by (D))} \\
&= \sum_{r=1}^{p} \Big(\sum_{t \in T_r} c_r \ell(K_t) \Big) \qquad \text{(by (A1) and (A4))} \\
&= \sum_{r=1}^{p} c_r \Big(\sum_{t \in T_r} \ell(K_t) \Big).
\end{aligned}$$

It remains to prove that $\sum_{t \in T_r} \ell(K_t) = \ell(I_r)$ for each r. Since I_r is the disjoint union of the intervals K_t for $t \in T_r$, this is true by 3.3.

Stage 3. Stages 1 and 2 show that we may assume that $\{I_r\}$ and $\{I'_s\}$ each consist of pairwise disjoint intervals and each $c_r \neq 0$ and each $c'_s \neq 0$. This ensures that $\bigcup_r I_r = \bigcup_s I'_s$; each union is $\{ x \mid \varphi(x) \neq 0 \}$. In this situation we can choose by common refinement a representation for φ as in Stage 2, so that the argument given there applies simultaneously to the representations $\sum_r c_r \chi_{I_r}$ and $\sum_s c'_s \chi_{I'_s}$, resulting in the same value for the integral in each case. □

An an immediate application we add a postscript to the discussion of length we gave in Chapter 3.

4.2 Unions of intervals. Suppose that S is the union of bounded disjoint intervals $K_1, \ldots, K_q$. Then $\chi_S = \sum_{t=1}^{q} \chi_{K_t}$. To check this, observe that at a point x the function on the right-hand side takes value 0 if x belongs to no K_t, and value 1 if x belongs to some (necessarily unique) K_t. We define the length of S by

$$\ell(S) := \int \chi_S = \sum_{t=1}^{q} \ell(K_t).$$

The Consistency Lemma guarantees that different decompositions of S as the disjoint union of finitely many bounded intervals give the same value for $\ell(S)$. Any finite union of bounded intervals can be expressed as the disjoint union of finitely many intervals (see the disjunction process described in 3.5). Therefore we have succeeded in extending our notion of length: we can now unambiguously assign a length to any finite union of bounded intervals, in a way which is in line with intuition.

We can quickly derive properties of the integral on $\mathrm{L}^{\mathrm{step}}$, benefiting greatly from the technical devices in 3.5.

4.3 Linearity property (L) for $\mathrm{L}^{\mathrm{step}}$. Let $\varphi, \psi \in \mathrm{L}^{\mathrm{step}}$ and $\lambda \in \mathbb{R}$. Then $\varphi + \lambda\psi \in \mathrm{L}^{\mathrm{step}}$ and

$$\int(\varphi + \lambda\psi) = \int\varphi + \lambda\int\psi.$$

Proof. We proved the first assertion in 3.7. To check the second, we represent φ and ψ using the same intervals for both:

$$\varphi = \sum_{t=1}^{q} e_t\chi_{K_t} \quad \text{and} \quad \psi = \sum_{t=1}^{q} f_t\chi_{K_t}.$$

Then

$$\int(\varphi + \lambda\psi) = \sum_{t=1}^{q}(d_t + \lambda e_t)\ell(K_t) = \sum_{t=1}^{q} e_t\ell(K_t) + \lambda\sum_{t=1}^{q} f_t\ell(K_t) = \int\varphi + \lambda\int\psi,$$

as required. □

[Note that we set up the definition of the integral to make it linear on characteristic functions of bounded intervals. This is analogous to defining a linear transformation T on a finite-dimensional vector space V by setting $T(\sum_{r=1}^{n}\lambda_r e_r) := \sum_{r=1}^{n}\lambda_r T(e_r)$, where $e_1, \ldots, e_n$ is a basis for V.]

4.4 Positivity property (P) for $\mathrm{L}^{\mathrm{step}}$. Let $\varphi, \psi \in \mathrm{L}^{\mathrm{step}}$ with $\varphi \geqslant \psi$. Then $\int\varphi \geqslant \int\psi$. In particular, if $\varphi \in \mathrm{L}^{\mathrm{step}}$ is such that $\varphi \geqslant 0$ then $\int\varphi \geqslant 0$.
[A stronger statement is made in 4.6 below.]

Proof. By 4.3 it suffices to prove the final statement. Let φ be a step function. Without loss of generality we may represent φ as $\sum_{r=1}^{n} c_r\chi_{I_r}$ where the intervals $I_1, \ldots, I_n$ are disjoint. Then $\varphi \geqslant 0$ means exactly that $c_r \geqslant 0$ for each r. Hence

$$\int\varphi := \sum_{r=1}^{n} c_r\ell(I_r) \geqslant 0. \qquad \square$$

4.5 Order properties for $\mathrm{L}^{\mathrm{step}}$. Let φ be a step function.

(a) **Modulus property (M)**: $|\varphi| \in \mathrm{L}^{\mathrm{step}}$ and $\left|\int\varphi\right| \leqslant \int|\varphi|$.

(b) φ may be written as the difference $\varphi = \varphi^+ - \varphi^-$ of non-negative step functions $\varphi^+ := \varphi \vee 0$, $\varphi^- := (-\varphi) \vee 0$ (and $|\varphi| = \varphi^+ + \varphi^-$).

Proof. We have already proved that $\varphi^+, \varphi^-, |\varphi| \in \mathrm{L}^{\mathrm{step}}$ (3.7). The remaining assertions in (b) hold for any real-valued function; see Exercise 2.3.

To prove the inequality in (a), we must show that $\int\varphi \leqslant \int|\varphi|$ and $-\int\varphi \leqslant \int|\varphi|$. Note that $|\varphi| \pm \varphi \geqslant 0$, so that by properties (L) and (P),

$$0 \leqslant \int(|\varphi| \pm \varphi) = \int|\varphi| \pm \int\varphi. \qquad \square$$

4.6 Zero integrals. Our definition of integral implies that $\int \chi_{\langle a,b\rangle}$ is the same whatever type of interval we are considering. In particular the values of a step function φ at its discontinuity points can be changed without affecting $\int \varphi$. A step function whose integral is zero need not itself be zero. For example $\int(\chi_{(0,1)} - \chi_{(-1,0)}) = 1 - 1 = 0$. Here the zero value arises by cancellation. Let's preclude this by considering a non-negative step function φ. Represent φ as $\sum_{r=1}^{n} c_r \chi_{I_r}$, with disjoint intervals I_r, so that $c_r \geqslant 0$ for all r. Then

$$\int \varphi = 0 \implies \sum_{r=1}^{n} c_r \ell(I_r) = 0.$$

Since every term in the sum is non-negative, each term must be zero. So, for each r, either $c_r = 0$ or $\ell(I_r) = 0$. The latter occurs if and only if I_r is a 1-point closed interval, $[a_r, a_r]$. We conclude that $\int \varphi = 0$ for a non-negative step function φ if and only if $\varphi = 0$ except possibly at a finite number of points. In the same way, we can obtain the conclusions of 4.4 under weaker hypotheses:

$$\varphi \geqslant \psi \text{ except at finitely many points} \implies \int \varphi \geqslant \int \psi.$$

4.7 Translation-invariance property (T) for $\mathrm{L}^{\mathrm{step}}$. Let $\varphi \in \mathrm{L}^{\mathrm{step}}$. For $d \in \mathbb{R}$, let φ_d be defined by $\varphi_d(x) := \varphi(x+d)$. Then $\varphi_d \in \mathrm{L}^{\mathrm{step}}$ and $\int \varphi_d = \int \varphi$.

Proof. We get the graph of φ_d by sliding the graph of φ a distance $|d|$ along the x-axis, left if $d \geqslant 0$ and right if $d < 0$; it is clear that doing this will not change the area under the graph. Here is the formal proof. Write φ as $\sum_{r=1}^{n} c_r \chi_{I_r}$. Then

$$\varphi_d(x) := \varphi(x+d) = \sum_{r=1}^{n} c_r \chi_{I_r}(x+d).$$

For any interval I,

$$\chi_I(x+d) = 1 \iff x + d \in I \iff x \in I - d,$$

where $I - d := \{\, y - d \mid y \in I \,\}$. If $I = \langle a, b\rangle$ we have $I - d = \langle a - d, b - d\rangle$ —the interval I shifted by d. Therefore

$$\varphi_d = \sum_{r=1}^{n} c_r \chi_{I_r - d},$$

which is a step function. Finally, $\ell(\langle a-d, b-d\rangle) = \ell(\langle a, b\rangle)$, so that

$$\int \varphi_d := \sum_{r=1}^{n} c_r \ell(I_r - d) = \sum_{r=1}^{n} c_r \ell(I_r) = \int \varphi. \qquad \square$$

Exercises

4.1 Prove that if φ, ψ are step functions then

$$\int (\varphi \vee \psi) \geqslant \max\{\int \varphi, \int \psi\}.$$

When does equality occur?

4.2 Let $\varphi \in \mathrm{L}^{\mathrm{step}}$ and recall that φ^c is defined by $\varphi^c(x) := \varphi(cx)$. Prove that $\varphi^c \in \mathrm{L}^{\mathrm{step}}$ and that $|c| \int \varphi^c = \int \varphi$ $(c \neq 0)$.

4.3 We have proved that the Basic Properties stated in Chapter 1 are true for the integral we have defined on $\mathrm{L}^{\mathrm{step}}$. Suppose we define the class $\mathcal{L}$ of integrable functions to be the minimal set of functions for which the Basic Properties hold. Show that $\mathcal{L} = \mathrm{L}^{\mathrm{step}}$.

4.4 Let φ be a step function on $[0, c]$ $(c > 0)$ such that $\varphi(0) = 0$, $\varphi(c) = d$, and φ is increasing on $[0, c]$. Show that there is a step function ψ on $[0, d]$ such that

$$cd = \int \varphi + \int \psi.$$

[Hint: draw a picture.]

5 Continuous functions on compact intervals

The last chapter concentrated on 'integral as area', for step functions. We wish to reconcile this with 'integral as antiderivative'. The Fundamental Theorem of Calculus, as stated in Chapter 1, concerned continuous functions on compact intervals. We therefore investigate how we can extend our integral to such functions, using step function approximations.

5.1 Definition of the integral on $\mathrm{C}[a,b]$**.** We fix a compact interval $[a,b]$ and take $f \in \mathrm{C}[a,b]$, the continuous functions from $[a,b]$ to $\mathbb{R}$. We define $\mathrm{L}^{\mathrm{step}}[a,b]$ to be the set of step functions which are zero outside $[a,b]$.

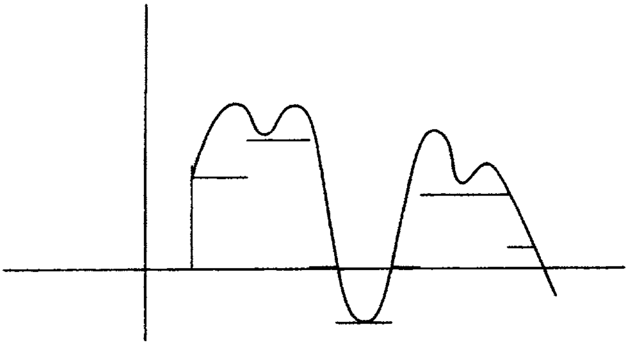


Figure 5.1

Assume for the moment that we have found a way of defining A, the area under the graph of f. Then we should want

$$A \geqslant \int \varphi \quad \text{whenever } \varphi \in \mathrm{L}^{\mathrm{step}}[a,b] \text{ and } \varphi \leqslant f \text{ on } [a,b];$$

see Fig. 5.1. This suggests that the supremum of $\int \varphi$, where φ is as above, is as good an approximation to the shaded area as we can get by approximating by step functions from below. So this is how we are going to **define** $\int_a^b f$:

$$\int_a^b f := \sup_{\varphi \in S_f} \int \varphi \quad \text{where } S_f := \{\, \varphi \in \mathrm{L}^{\mathrm{step}}[a,b] \mid \varphi \leqslant f \,\}.$$

If f is the constant function $c\chi_{[a,b]}$, then f itself is in S_f; indeed, f is the largest member of S_f. In this case the supremum of $\int \varphi$ for $\varphi \in S_f$ is just the $\mathrm{L}^{\mathrm{step}}$ integral of f, namely $c(b-a)$, the expected 'area under the graph'. To convince ourselves in general that $\int_a^b f$ really does represent the shaded area we should also like to know that we can choose φ in S_f so that $\varphi(x)$ approximates $f(x)$ 'to within ε' at each individual point x. This is plausible, and indeed true, but we do not need it or prove it until much later (see 9.2).

We now check—a little belatedly—that our definition of $\int_a^b f$ makes sense for any $f \in \mathrm{C}[a,b]$. The Boundedness Theorem for continuous functions, 2.23, tells us that there exists a finite constant M such that $-M \leqslant f(x) \leqslant M$ for $x \in [a,b]$. We deduce that S_f has the following properties:

(i) $-M\chi_{[a,b]} \in S_f$ (so S_f is non-empty), and

(ii) for all $\varphi \in S_f$, $\varphi \leqslant M\chi_{[a,b]}$ and so $\int \varphi \leqslant M(b-a)$.

Consequently, by the Completeness Property, 2.5, $\sup\{\int \varphi \mid \varphi \in S_f\}$ exists in $\mathbb{R}$. We have shown that our definition of the integral on $\mathrm{C}[a,b]$ is indeed a legitimate one.

Note that the definition of $\int_a^b f$ implies that $\int_a^a f = 0$ since in this case every $\varphi \in S_f$ is zero except possibly at a. When $b > a$ we set $\int_b^a f := -\int_a^b f$, as a notational convenience.

Our principal goal in the development of the integral on $\mathrm{C}[a,b]$ will be the Fundamental Theorem of Calculus and its companion the Indefinite Integral Theorem. On the way to these key theorems we establish for functions in $\mathrm{C}[a,b]$ some of the Basic Properties. We first present a batch of very useful order properties. These stem from the fact that, by construction, $\int_a^b f \geqslant \int \varphi$ for every $\varphi \in S_f$.

5.2 Integrals and inequalities.

(a) Let $f, g \in \mathrm{C}[a,b]$ with $f \geqslant g$. Then $\int_a^b f \geqslant \int_a^b g$.

(b) Let $f \in \mathrm{C}[a,b]$ be such that $m \leqslant f \leqslant M$, where m and M are constants. Then

$$m(b-a) \leqslant \int_a^b f \leqslant M(b-a).$$

(c) **Positivity property (P) for** $\mathrm{C}[a,b]$: if $f \in \mathrm{C}[a,b]$ with $f \geqslant 0$, then $\int_a^b f \geqslant 0$.

(d) Let $f \in \mathrm{C}[a,b]$, $f \geqslant 0$, and $\int_a^b f = 0$. Then $f \equiv 0$.

Proof. Consider (a). Assume $f \geqslant g$. Then $\varphi \leqslant g$ implies $\varphi \leqslant f$, so $S_g \subseteq S_f$, whence $\int_a^b f \geqslant \int_a^b g$ (the bigger the set, the bigger the supremum). For (c) just put $g = 0$. For (b) observe that $m \leqslant f \leqslant M$ means $m\chi_{[a,b]} \leqslant f \leqslant M\chi_{[a,b]}$ and apply (a).

To prove (d) we assume for a contradiction that there exists $c \in [a,b]$ such that $f(c) > 0$. By Lemma 2.22, we can find $\delta > 0$ such that $x \in [a,b]$, $|x-c| < \delta$ implies $f(x) > f(c)/2$. Define

$$\varphi := \frac{1}{2} f(c)\chi_I, \quad \text{where } I := [a,b] \cap (c-\delta, c+\delta).$$

By construction, $f \geqslant \varphi$ (see Fig. 5.2). Therefore, by (a),

$$\int f \geqslant \int \varphi = \frac{1}{2} f(c)\ell(I) > 0,$$

the required contradiction. $\square$

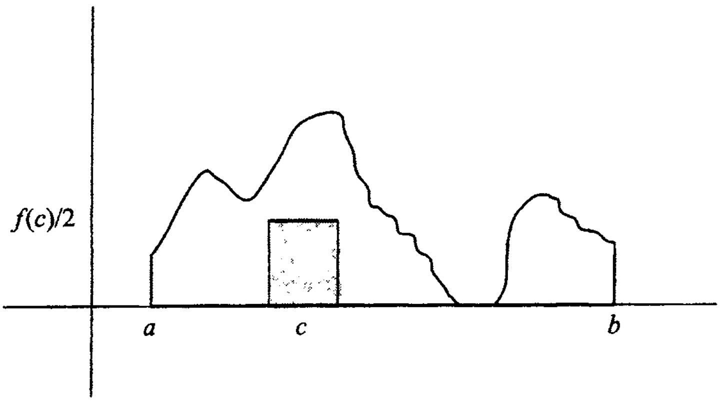


Figure 5.2

For L^{step} the linearity property (L) was immediate. This is not so for $C[a,b]$. Try to do it!—can you relate S_{f+g} to S_f and S_g, or S_{-f} to S_f? We make do temporarily with a special form of linearity, showing that integrals over different intervals fit together in the way we would expect.

5.3 Subdividing $[a,b]$. Let $f \in C[a,b]$ and let $a < c < b$. Then f is continuous on $[a,c]$ and on $[c,b]$ and

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

Proof. For any $\varphi \in L^{step}[a,b]$,

$$\varphi = \varphi\chi_{[a,c]} + \varphi\chi_{(c,b]} \quad \text{and} \quad \int \varphi = \int \varphi\chi_{[a,c]} + \int \varphi\chi_{(c,b]},$$

by linearity of the L^{step} integral (4.3). Hence

$$\int \varphi \leqslant \int_a^c f + \int_c^b f \quad \text{for all } \varphi \in S_f.$$

Taking the supremum over all $\varphi \in S_f$ we get

$$\int_a^b f \leqslant \int_a^c f + \int_c^b f.$$

For the reverse inequality we use the fact that if $\psi \in L^{step}[a,c]$ and $\theta \in L^{step}[c,b]$ with $\psi \leqslant f$ on $[a,c]$ and $\theta \leqslant f$ on $[c,b]$ then $\psi + \varphi \in L^{step}[a,b]$ and $\psi + \varphi \leqslant f$

on $[a,b]$, except possibly at c. By adjusting the value at c we can get a step function $\varphi \in S_f$ such that $\int \varphi = \int(\psi+\theta) = \int \psi + \int \theta$. Then

$$\int_a^b f \geqslant \int \varphi = \int \psi + \int \theta.$$

Taking the supremum over all possible ψ and θ gives

$$\int_a^b f \geqslant \int_a^c f + \int_c^b f. \qquad \square$$

5.4 Indefinite Integral Theorem I. Let $g \in \mathrm{C}[a,b]$. Define the *indefinite integral of* g to be G, where

$$G(x) := \int_a^x g.$$

Then G is differentiable on $[a,b]$ (with one-sided derivatives at the endpoints) and $G' = g$ on $[a,b]$.

[More sophisticated theorems on indefinite integrals come later (12.15 and 24.7.]

Proof. We have to show that, for any $x \in [a,b]$, and h such that $x+h \in [a,b]$,

$$\frac{G(x+h)-G(x)}{h} \to g(x) \quad \text{as } h \to 0$$

(the restriction on $x+h$ automatically takes care of one-sided approach when x is a or b). For $x, x+h \in [a,b]$ and $h > 0$, we have, by 5.3 with $a,b,c = a,x,x+h$,

$$G(x+h) - G(x) = \int_x^{x+h} g.$$

Now 5.2 implies that

$$\inf\{\, g(t) \mid t \in [a,a+h]\,\} \leqslant \frac{1}{h}\int_x^{x+h} g \leqslant \sup\{\, g(t) \mid t \in [a,a+h]\,\}.$$

By the remark following the Intermediate Value Theorem, 2.24, $\int_x^{x+h} g = hg(\xi)$ for some $\xi \in [x, x+h]$. Similarly, for $h < 0$,

$$G(x) - G(x+h) = \int_{x+h}^{x} g = -hg(\xi) \quad \text{for some } \xi \in [x+h, x].$$

Either way, there exists $\xi \in [a,b]$ with $|x-\xi| \leqslant |h|$ such that

$$\frac{G(x+h)-G(x)}{h} = g(\xi).$$

As $h \to 0$, the right-hand side tends to $g(x)$, by continuity of g. $\square$

5.5 The Fundamental Theorem of Calculus. Let $f \in \mathrm{C}^1[a,b]$ (that is, assume f' exists and is continuous on $[a,b]$, with one-sided derivatives at the endpoints). Then

$$\int_a^b f' = f(b) - f(a).$$

Proof. Define $F(x) := \int_a^x f'$. By the Indefinite Integral Theorem, $F' = f'$ on $[a,b]$. Hence $F - f$ is a constant (see 2.26(3)(b)). Since $F(a) = 0$ and $F(b) = \int_a^b f'$ the theorem follows. □

A friend is often known by a shortened name. As a mark of such status we henceforth call the Fundamental Theorem of Calculus the FTC.

5.6 Remarks. When we consider specific functions, such as powers or trigonometric functions, we customarily include a variable, writing, for example, $\sin x$ to denote the function f for which $f(x) = \sin x$ for each x. Therefore we shall allow

$$\int_a^b f(x)\,dx \quad \text{as an alternative way of writing} \quad \int_a^b f.$$

Note that x here is a 'dummy variable' (or, in logicians' terminology, a 'bound variable'). We could equally well have used a different letter, t, y, $\ldots$. In different contexts different symbols are adopted as standard (for example t to represent time, or r radial distance). Therefore we shan't attempt to retain a uniform usage.

We have adopted the notation $\int_a^b f$ for our integral of $f \in \mathrm{C}[a,b]$. Including the limits of integration a, b reminds us which interval we are working on. Later, when we consider principally functions defined on $\mathbb{R}$ it will be convenient to extend $f \in \mathrm{C}[a,b]$ to a function $\tilde{f}$ on $\mathbb{R}$ by defining

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in [a,b], \\ 0 & \text{otherwise.} \end{cases}$$

Of course, $\tilde{f}$ will not in general be continuous at a and b. We shall define

$$\int \tilde{f} := \int_a^b f.$$

Here, as with $\mathrm{L}^{\mathrm{step}}$, we use the notation $\int g$, with the integral symbol unadorned, to denote the integral of a function $g \colon \mathbb{R} \to \mathbb{R}$. We stress that $\int g$ is a **real number**, representing the (signed) area under the graph of g. It is a *definite integral*. We refer to g as the *integrand*.

Remembering the convention adopted in 5.1 that $\int_a^b f := -\int_b^a f$ for $a > b$ it is easy to check that both the Indefinite Integral Theorem and the Fundamental Theorem of Calculus remain valid for integrals with limits 'the wrong way round'; see Exercise 5.6. Provided we are careful with signs this observation can often spare us having to separate the cases $a < b$ and $b \leqslant a$ when dealing with $\int_a^b f$.

5.7 Integrals and amtiderivatives. A function F such that $F' = g$ is called an *antiderivative* or a *primitive* for the function g. If $g \in \mathrm{C}[a,b]$ then any two antiderivatives for g differ by a constant (see 2.26(3)(b)). When students first meet integrals they are often taught to write, for example, '$\int \sin x = -\cos x$' to mean 'the antiderivative of sin is cos, that is, $(\mathrm{d}/\mathrm{d}x)\cos x = -\sin x$'. We outlaw this usage completely. It is at variance with the notation adopted in 5.6, and is apt to cause confusion. Correctly, what the FTC gives us is

$$\int_a^x \cos t \,\mathrm{d}t = \sin x - \sin a.$$

Here each side is a function of x. Notice that $\int_a^x \cos x \,\mathrm{d}x$ is nonsensical, because we are using x simultaneously in two different ways, as an endpoint of the interval over which we are integrating and as a dummy variable. To see how confusion can arise you are invited to consider what $x + \int \cos x \,\mathrm{d}x$ is intended to mean, and to rewrite the expression correctly. it is now worth looking at how differential equations are solved by integration. To take a very simple example, suppose that $y'(x) = 2x$, with $y(0) = 1$. To avoid using x in two ways at once we switch to a new dummy variable t, and write the equation as $y'(t) = 2t$. Observe that y' is continuous. Hence, by the FTC,

$$y(x) = y(0) + \int_0^x y'(t)\,\mathrm{d}t = \int_0^x 2t\,\mathrm{d}t = \left[t^2\right]_0^x = x^2.$$

Taking account of the condition $y(0) = 1$, we conclude that $y(x) = 1 + x^2$. See 6.17 and 8.15 for illustrations of the integration of a derivative in action.

After this little homily, we set some exercises on to reinforce the points made.

5.8 Exercise example. Use the FTC to verify that

(i) $\displaystyle\int_a^b x\,\mathrm{d}x = (b^2 - a^2)/2,$ (ii) $\displaystyle\int_0^\pi \sin\theta\,\mathrm{d}\theta = 2,$

(iii) $\displaystyle\int_1^x t^{-100}\,\mathrm{d}t = (1 - x^{-99})/99,$ (iv) $\displaystyle\int_0^x (1 + y + 3y^2)\,\mathrm{d}y = x + \frac{1}{2}x^2 + x^3,$

(v) $\displaystyle\int_0^x \sin t\,\mathrm{d}t = 1 - \cos x,$ (vi) $\displaystyle\int_0^x (1 + t^2)^{-1}\,\mathrm{d}t = \tan^{-1} x.$

(In (vi) we take the principal value of the inverse tangent, with values in $(-\pi/2, \pi/2)$.)

It is now possible rapidly to develop further properties of our integral. The next proof may seem devious. If you worked through the preceding exercise, in part (iv) you will either have tacitly assumed the linearity property (L) (not yet allowed!) or have made use of the linearity of differentiation, just as we do below.

5.9 Linearity property (L) for $\mathrm{C}[a, b]$. The integral on $\mathrm{C}[a, b]$ is linear, that is, for $f, g \in \mathrm{C}[a, b]$ and $\lambda \in \mathbb{R}$,

$$\int_a^b (f + \lambda g) = \int_a^b f + \lambda \int_a^b g.$$

Proof. By linearity of differentiation and the Indefinite Integral Theorem,
$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_a^x f + \lambda\int_a^x g\right) = \frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f + \lambda\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x g = f(x) + \lambda g(x) = \frac{\mathrm{d}}{\mathrm{d}x}\int_a^x (f+\lambda g).$$
Hence
$$\int_a^x (f+\lambda g) - \left(\int_a^x f + \lambda\int_a^x g\right) = C,$$
where C is a constant independent of x. Put $x = a$ to get $C = 0$. □

5.10 Lemma. Let $f, g \in \mathrm{C}[a,b]$. Then $k \in \mathrm{C}[a,b]$, and so $\int_a^b k$ is defined, when k is any of

(a) fg and $1/f$ (the latter if $f(x) \neq 0$ for all $x \in [a,b]$), and

(b) $f \vee g$, $f \wedge g$, and $|f|$.

Further, $\left|\int_a^b f\right| \leqslant \int_a^b |f|$.

Proof. Concerning the continuity assertions, see 2.20. To prove the inequality, we argue just as we did for $\mathrm{L}^{\mathrm{step}}$ in 4.5, considering $|f| \pm f$ (note Exercise 2.3). □

5.11 Lemma.

(a) **Translation-invariance property (T) for** $\mathrm{C}[a,b]$. Let $f \in \mathrm{C}[a,b]$. Then
$$\int_a^b f(x+d)\,\mathrm{d}x = \int_{a+d}^{b+d} f(y)\,\mathrm{d}y \quad (d \in \mathbb{R}).$$

(b) Let $f \in \mathrm{C}[a,b]$ and let $c \neq 0$. Then
$$c\int_a^b f(cx)\,\mathrm{d}x = \begin{cases} \displaystyle\int_{ac}^{bc} f(y)\,\mathrm{d}y & (c > 0), \\ \displaystyle\int_{bc}^{ac} f(y)\,\mathrm{d}y & (c < 0). \end{cases}$$

Proof. Consider (a). Since $x \in [a,b]$ if and only if $x + d \in [a+d, b+d]$,
$$\{\, \psi \in \mathrm{L}^{\mathrm{step}}[a+d, b+d] \mid \psi \leqslant f_d \,\} = \{\, \varphi_d \mid \varphi \in S_f \,\}.$$
By property (T) for $\mathrm{L}^{\mathrm{step}}$, proved in 4.7,
$$\int_a^b f(x+d)\,\mathrm{d}x = \sup_{\psi\in S_{f_d}} \int \psi = \sup_{\varphi\in S_f} \int \varphi_d = \sup_{\varphi\in S_f} \int \varphi = \int_{a+d}^{b+d} f(y)\,\mathrm{d}y.$$
We leave (b) as an exercise. □

We have now completed the task of setting up the integral on $\mathrm{C}[a,b]$ and establishing the Basic Properties. Of course, Lemma 5.11 is just a formal way of saying that a linear change of variable $x \mapsto cx + d$ ($c \neq 0$) is admissible. We consider substitution in general in the next chapter.

Finally in this chapter we assemble properties of the logarithm, defined as an indefinite integral. We only ever consider $\log_e$, and use the symbol log for this rather than the alternative ln.

5.12 The logarithm function. We define

$$\log x := \int_1^x \frac{1}{t}\,\mathrm{d}t \quad (x > 0)$$

(remember that this means $-\int_x^1 t^{-1}\,dt$ if $0 < x < 1$). Then

(a) $(\mathrm{d}/\mathrm{d}x)\log x = 1/x$ for any $x > 0$;

(b) $\log x < 0$ for $0 < x < 1$, $\log 1 = 0$, and $\log x > 0$ for $x > 1$;

(c) $x \leqslant y$ implies $\log x \leqslant \log y$ and $\log x \leqslant x - 1$ for any $x \geqslant 1$;

(d) $\log(xy) = \log x + \log y$ for any $x, y > 0$, and $\log(1/x) = -\log x$;

(e) As $x \to \infty$, $\log x \to \infty$ and as $x \to 0+$, $\log x \to -\infty$;

(f) As $x \to \infty$, $x^{-1}\log x \to 0$ and as $x \to 0+$, $x\log x \to 0$.

Proof. (a) is immediate from the Indefinite Integral Theorem, while (b) and (c) come from property (P) and 5.2(d). For (d) assume without loss of generality that $x \leqslant y$. First note that if $x, y \geqslant 1$ then, with the aid of 5.11,

$$\log(xy) = \int_1^{xy} \frac{1}{t}\,\mathrm{d}t = \int_1^x \frac{1}{t}\,\mathrm{d}t + \int_x^{xy} \frac{1}{t}\,\mathrm{d}t = \int_1^x \frac{1}{t}\,\mathrm{d}t + \int_1^y \frac{1}{u}\,\mathrm{d}u = \log x + \log y.$$

The same argument applies if $x, y \leqslant 1$; just reverse all the limits. Now assume $0 < x \leqslant 1 \leqslant y$. Then (check separately the cases $xy \geqslant 1$ and $xy < 1$),

$$\log x = -\int_x^1 \frac{1}{t}\,\mathrm{d}t = -\int_{xy}^y \frac{1}{u}\,\mathrm{d}u = -\int_1^y \frac{1}{u}\,\mathrm{d}u + \int_1^{xy} \frac{1}{u}\,\mathrm{d}u.$$

This proves the first part of (d). Put $y = 1/x$ to get the second assertion.

There are many ways to prove (e). Here is one, based on (c) and (d). Given x, choose $n \in \mathbb{N}$ such that $2^n \leqslant x$ (possible since $2^n \to \infty$ as $n \to \infty$). Then

$$n \log 2 = \log(2^n) \leqslant \log x.$$

Since $n \to \infty$ as $x \to \infty$ the first part of (e) follows. Again put $y = 1/x$ to get the second assertion.

Finally consider (f). We have, for $x \geqslant 1$,

$$\frac{\log x}{x} = \frac{2}{\sqrt{x}}\left(\frac{\log \sqrt{x}}{\sqrt{x}}\right) \leqslant \frac{2}{\sqrt{x}},$$

by (c), applied with $\sqrt{x}$ in place of x. This gives the first part of (f) and the second follows from (d). $\square$

Exercises

5.1 The definition of $\int_a^b f$ in 5.1 is asymmetric in that it uses approximation from below. We might alternatively have used approximation from above. For $f \in \mathrm{C}[a,b]$ define

$$D := \inf\{ \int \psi \mid \psi \in \mathrm{L}^{\mathrm{step}}[a,b],\ \psi \geqslant f \}.$$

Show that $D = \int_a^b f$.

5.2 Let $f \in \mathrm{C}[a,b]$. Prove that $\left|\int_a^b f\right| = \int_a^b |f|$ if and only if either $f \geqslant 0$ on $[a,b]$ or $f \leqslant 0$ on $[a,b]$.

5.3 Let $g, h \in \mathrm{C}[a,b]$. By considering $\int_a^b (g \pm h)^2$, prove the **Cauchy–Schwarz inequality**

$$\left(\int_a^b gh\right)^2 \leqslant \int_a^b g^2 \int_a^b h^2.$$

Let $f \in \mathrm{C}[a,b]$ be such that $f \geqslant 0$. By taking (i) $g(x) = f(x)^{3/2}$, $h(x) = f(x)^{1/2}$, and (ii) $g(x) = f(x)$, $h(x) = 1$, prove that

$$\left(\int_a^b f\right)^3 \leqslant (b-a)^2 \int_a^b f^3.$$

5.4 Let $f(x) = a_0 + a_1 x + \cdots + a_m x^m$ be a polynomial $(a_0, \ldots, a_m \in \mathbb{R})$. Assume that

$$\int_a^b f(x) x^n \,\mathrm{d}x = 0 \quad (n = 0, 1, \ldots).$$

Prove that $f \equiv 0$.

5.5 Let $g \in \mathrm{C}[a,b]$. Prove that there exists $c \in [a,b]$ such that $\int_a^c g = \int_c^b g$. [Hint: apply the Intermediate Value Theorem, 2.24, to G, where $G(x) = \int_a^x g$.]

5.6 Let $g \in \mathrm{C}[a,b]$ and define $H(x) := \int_x^b g$ for $x \in [a,b]$. By appealing to 5.3 or to 5.11(b) prove that $H' = -g$.

5.7 Let $\varphi \in \mathrm{L}^{\mathrm{step}}$ and define

$$\Phi(x) := \int_0^x \varphi \quad \left(:= \int \varphi \chi_{[0,x]}\right).$$

Describe Φ graphically. At which points x is it true that $\Phi'(x) = \varphi(x)$?

5.8 What is wrong with the following argument?

$$\int_{-1}^1 x^{-2} \,\mathrm{d}x = \left[-x^{-1}\right]_{-1}^1 = (-1) - (-(-1)) = -2.$$

5.9 Let $f \in \mathrm{C}[0,b]$ $(b > 0)$. Define

$$\Phi(x) := \frac{1}{x} \int_0^x f(t) \,\mathrm{d}t \quad (x > 0).$$

Prove that

$$\Phi'(x) = \frac{1}{x^2}\int_0^x (f(x) - f(t))\,dt.$$

Deduce from 2.26(c)(a) that if f is non-negative and increasing on $(0, b]$ then the same is true of Φ.

5.10 (a) Let $\eta\colon \mathbb{R} \to \mathbb{R}$ be a continuous function with the properties that $\eta \geqslant 0$, $\eta(x) = 0$ for $|x| \geqslant 1$, and $\int \eta = 1$. Let $\eta_n(x) := n\eta(nx)$. Prove that if $f \in \mathrm{C}[-1, 1]$ then

$$\lim_{n\to\infty} \int_{-1}^{1} f\eta_n = f(0).$$

(b) Let $\psi_n := n^2\chi_{[0,1/n]} - n^2\chi_{[-1/n,0)}$. Prove that if $f \in \mathrm{C}[-1, 1]$ and $f'(0)$ exists then

$$\lim_{n\to\infty} \int_{-1}^{1} f\psi_n = f'(0).$$

5.11 Let $g \in \mathrm{C}^2(\mathbb{R})$ and assume that $g'' \geqslant 0$. Use the FTC to prove that g is *convex*, that is, for any $u, v \in \mathbb{R}$,

$$g\big(\tfrac{1}{2}(u+v)\big) \leqslant \frac{1}{2}\left(g(u) + g(v)\right),$$

and interpret this result graphically. [The converse is true too: apply L'Hôpital's rule to $(g(x+h) + g(x-h) - 2g(x))/h^2$.]

6 Techniques of integration I

This chapter links the theory developed in Chapter 5 to the tools of practical integration. We summarize the most useful techniques, for revision and reference. The methods—recognizing the integrand as a derivative, integration by parts, and integration by substitution—can seem haphazard to beginners in calculus. A review at this stage may help to dispel any lingering mystery.

The exhaustive accounts of 'systematic integration' in older textbooks have been rendered largely obsolete by sophisticated computer algebra packages which can work out a wide range of integrals exactly, just as might be required for a calculus assignment. Nonetheless, to gain a real understanding of the processes involved, and of their limitations, there is no substitute for the experience that comes from pencil-and-paper calculations. All the examples below come within the scope of the theory of Chapter 5. Extensions to integrals over unbounded intervals, for example, are given in Chapter 16.

Our primary concern will be the evaluation of $\int_a^b f$ for a variety of functions $f \in \mathrm{C}[a,b]$. Our principal tool is the Fundamental Theorem of Calculus, 5.5, which may be presented as follows.

6.1 The Fundamental Theorem of Calculus (restated). Let $f \in \mathrm{C}[a,b]$. If F is such that $F' = f$ on $[a,b]$ then $\int_a^b f = F(b) - F(a)$.

6.2 Recognizing antiderivatives. To evaluate $\int_a^b f$ by the FTC we need to find F such that $F'(x) = f(x)$ for all $x \in [a,b]$. An initial list of antiderivatives comes from known derivatives: for $f(x) = x^p$ we may take $F(x) = x^{p+1}/(p+1)$ ($p \in \mathbb{R}$, $p \neq -1$), for $f(x) = \sec^2 x$ we take $F(x) = \tan x$, and for $f(x) = \mathrm{e}^x$ we take $F(x) = \mathrm{e}^x$, and so on. We must ensure that our functions are well defined—no square roots or logarithms of negative numbers! Thus x^{-1} has antiderivative $\log x$ for $x > 0$ and $-\log|x|$ for $x < 0$. A list of the most fundamental antiderivatives is given in Appendix II. Some of these antiderivatives, and many others too, come from the chain rule for differentiation, which tells us that $g'(f(x))f'(x)$ is the derivative of $g(f(x))$ (assuming the former expression makes sense). As special cases we have

(a) $(f(x))^2$ is an antiderivative of $2f(x)f'(x)$;

(b) $-1/f(x)$ is an antiderivative of $f'(x)/(f(x))^2$ $(f(x) \neq 0)$;

(c) $\mathrm{e}^{f(x)}$ is an antiderivative of $f'(x)\mathrm{e}^{f(x)}$;

(d) if $f > 0$, $\log f(x)$ is an antiderivative of $f'(x)/f(x)$.

For example,

(i) $-\frac{1}{2}\cos^2 x$ is an antiderivative of $\sin x \cos x$,

(ii) $-e^{-\frac{1}{2}x^2}$ is an antiderivative of $xe^{-\frac{1}{2}x^2}$,
(iii) $\log\sin x$ is an antiderivative of $\cot x$ (when $\sin x > 0$), and
(iv) $\log(\log x)$ is an antiderivative of $1/(x\log x)$ (for $x > 1$).

6.3 Exercise example. Write down antiderivatives for

(i) $x\sin(x^2)$, (ii) $x^3e^{-x^4}$, (iii) $\sinh x\cosh^{13}x$, (iv) $(1+x)^{-3}$,
(v) $\dfrac{x}{(1+x^2)^4}$, (vi) $\sin x\sec^2 x$, (vii) $\dfrac{\tan^{-1}x}{1+x^2}$, (viii) $\dfrac{1}{x(\log x)^2}$.

In each case indicate any values of x which must be debarred.

Convince yourself that the technique of 6.2 will not directly supply antiderivatives for the following (but see also 6.6, 6.11, and the remarks in 6.16):

(i) $\sin^n x$ $(n \geqslant 2)$, (ii) $(1+x^2)^{-2}$, (iii) $e^{\sin x}$, (iv) $\cos(x^2)$.

6.4 Recognizing antiderivatives (continued). Frequently a function needs to be expressed in a different form for an antiderivative to be spotted. Below are samples of commonly used types of manipulation.

Trigonometric and hyperbolic functions.

(a) Using the formulae $\sin 2x = 2\sin x\cos x$ and $\cos 2x = 1 - 2\sin^2 x$:

$$\sin^2 x\cos^2 x = \frac{1}{4}\sin^2 2x = \frac{1}{8}(1-\cos 4x) = \frac{d}{dx}\left(\frac{1}{8}x - \frac{1}{32}\sin 4x\right).$$

(b) Using the formula $\sin A\cos B = \frac{1}{2}(\sin(A+B) + \sin(A-B))$:

$$\sin 5x\cos 3x = \frac{1}{2}(\sin 8x + \sin 2x) = \frac{d}{dx}\left(-\frac{1}{16}\cos 8x - \frac{1}{4}\cos 2x\right).$$

(c) Using $\cosh^2 x - \sinh^2 x = 1$:

$$\sinh^5 x\cosh^3 x = \sinh^5 x\cosh x + \sinh^7 x\cosh x = \frac{d}{dx}\left(\frac{1}{6}\sinh^6 x + \frac{1}{8}\sinh^8 x\right).$$

Rational functions. Note the ways in which heavy computations are avoided.

(a) Removal of improper fractions:

(i) $x(x+2)^{-1} = 1 - 2(x+2)^{-1}$, which is the derivative of $x - 2\log(x+2)$ if $x > -2$ and of $x + 2\log|x+2|$ of $x < -2$.

(ii) $x^3(x+2)^{-1} = ((x+2)-2)^3(x+2)^{-1} = (x+2)^2 - 6(x+2) + 12 - 8(x+2)^{-1}$, for which an antiderivative can be written down.

Long division can always be applied to top-heavy fractions but, as (ii) shows, it is not needed in every case.

(b) Partial fractions:

(i) $\dfrac{1}{4-x^2} = \dfrac{1}{(2-x)(2+x)} = -\dfrac{1}{4}\left(\dfrac{1}{x-2} - \dfrac{1}{2+x}\right) = F'(x)$, where $F(x)$ is $-\frac{1}{4}\log((x-2)(x+2))$ for $|x| > 2$ and $\frac{1}{4}\log((2-x)(x+2))$ ($|x| < 2$).

(ii) $\dfrac{1}{x^4-1} = \dfrac{1}{(x^2-1)(x^2+1)} = \dfrac{1}{2(x^2-1)} - \dfrac{1}{2(x^2+1)}$; the first term can be treated as in (i), and the second has antiderivative $\frac{1}{2}\tan^{-1} x$.

(c) Building a divisor of the denominator:

$$\frac{x}{(x+2)^3} = \frac{x+2}{(x+2)^3} - \frac{2}{(x+2)^3} = \frac{\mathrm{d}}{\mathrm{d}x}\left(-(x+2)^{-1} + (x+2)^{-2}\right).$$

(d) Completing the square in an irreducible quadratic factor:

$$\frac{1}{1+x+x^2} = \frac{1}{\frac{3}{4} + (x+\frac{1}{2})^2} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{2}{\sqrt{3}}\tan^{-1}((2x+1)/\sqrt{3})\right).$$

(e) Building the derivative of the denominator:

$$\frac{x}{1+x+x^2} = \frac{2x+1}{2(1+x+x^2)} - \frac{1}{2(1+x+x^2)},$$

of which the first term is the derivative of $\frac{1}{2}\log(1+x+x^2)$; note that the calculation in (d) shows that $1 + x + x^2 > 0$ for all x.

6.5 Exercise example. Find antiderivatives for

(i) $\dfrac{x^2}{x^2+1}$, (ii) $\dfrac{1}{(x+2)(x+4)}$, (iii) $\dfrac{x^2+1}{(x+2)^2}$, (iv) $\dfrac{x}{x^2+3x+2}$,

(v) $\dfrac{1}{x(x^2+4)}$, (vi) $\dfrac{1}{x^2+6x+25}$, (vii) $\sin^3 x \cos^3 x$, (viii) $\sin^2 x \cos^4 x$.

Closely related to 6.2 is the method of substitution, in which we convert a given integral written in terms of a variable x into a more amenable integral in terms of some new variable $t = g(x)$, putting '$\mathrm{d}t = g'(x)\mathrm{d}x$'. This rule of thumb is brought within our theory with the aid of the chain rule and the Indefinite Integral Theorem, as follows.

6.6 Integration by substitution. Let $f \in \mathrm{C}[a,b]$ and suppose that $g\colon [c,d] \to [a,b]$ is such that g' exists and is positive and continuous on $[c,d]$. Then

$$\int_{g(c)}^{g(d)} f(t)\,\mathrm{d}t = \int_c^d f(g(x))g'(x)\,\mathrm{d}x.$$

Proof. Note that $(f \circ g)(t) := f(g(t))$ is defined for each $t \in [c,d]$. Continuity of $f \circ g$ is a standard result about continuous functions. Let $F(t) := \int_{g(c)}^{t} f(x)\,\mathrm{d}x$, so that $F' = f$ by the Indefinite Integral Theorem. Now consider $u := F \circ g$. By the chain rule for derivatives,

$$u'(t) = (F \circ g)'(t) = F'(g(t))g'(t) = f(g(t))g'(t) \quad (t \in [c,d]).$$

Hence, by the FTC,

$$\begin{aligned}\int_c^d f(g(t))g'(t)\,\mathrm{d}t &= u(d) - u(c) = F(g(d)) - F(g(c)) \\ &= \int_{g(c)}^{g(d)} F'(x)\,\mathrm{d}x = \int_{g(c)}^{g(d)} f(x)\,\mathrm{d}x. \qquad \square\end{aligned}$$

6.7 Remarks. Care is needed with signs in making substitutions. The restriction that g' be positive is not necessary for the proof of the substitution formula. However it does guarantee that g is strictly increasing and so that g maps $[c,d]$ onto $[g(c), g(d)]$, with the limits automatically coming out 'the right way round'. Further, it ensures that $x \mapsto t := g(x)$ has a well-defined inverse function $t := g^{-1}(x)$, and it is this that is used in practice to convert an integral $\int_a^b f(x)\,\mathrm{d}x$ to a more tractable form.

Injudicious substitution can lead to nonsense. For example if we write $t = g(x) := x^{-1}$ in $\int_{-1}^{1}(1+x^2)^{-1}\,\mathrm{d}x$ $(= I$, say), we appear to get $-\int_{-1}^{1}(t^2+1)^{-1}\,\mathrm{d}t$, which would imply $I = 0$. This is wrong. Since $(1+x^2)^{-1}$ is continuous and is an antiderivative of $\tan^{-1} x$, the FTC gives, correctly, $I = \pi/2$. The error was in an invalid choice of g: irrespective of the value assigned to it at $x = 0$, x^{-1} does not have a continuous derivative on $[-1,1]$.

6.8 Examples. In each of the examples below 6.6 applies, with the function $g(x)$ having a continuous derivative on the required interval.

(1) By substituting $t = g(x) := \sqrt{x+2}$,

$$\begin{aligned}\int_2^7 \frac{1}{(x+1)\sqrt{x+2}}\,\mathrm{d}x &= \int_2^3 \frac{2t}{(t^2-1)t}\,\mathrm{d}t \\ &= \int_2^3 \frac{1}{t-1} - \frac{1}{t+1}\,\mathrm{d}t = [\log(t-1) - \log(t+1]_2^3 = \log(3/2).\end{aligned}$$

(2) By substituting $x = \sin t$,

$$\int_{1/\sqrt{2}}^{1} \frac{1}{x^2\sqrt{1-x^2}}\,\mathrm{d}x = \int_{\pi/4}^{\pi/2} \frac{\cos t}{\sin^2 t \cos t}\,\mathrm{d}t = [-\cot t]_{\pi/4}^{\pi/2} = 1.$$

Note that the sign of the square root is chosen correctly here: $\sqrt{1-\sin^2 t}$ is indeed $+\cos t$ on $[\pi/4, \pi/2]$.

(3) By substituting $t = \tan^{-1} x$,

$$\int_0^1 \frac{1}{(1+x^2)^2}\,dx = \int_0^{\pi/4} \frac{\sec^2 t}{(\sec^2 t)^2}\,dt$$
$$= \int_0^{\pi/4} \cos^2 t\,dt = \frac{1}{2}\int_0^{\pi/4} (1+\cos 2t)\,dt = \frac{1}{4}\left(\frac{\pi}{2}+1\right).$$

(4) By substituting $t = \tan\frac{1}{2}x$, and noting that $dx/dt = 2/(1+t^2)$ and $\sin x = 2t/(1+t^2)$, we have

$$\int_{\pi/3}^{2\pi/3} \operatorname{cosec} x\,dx = \int_{1/\sqrt{3}}^{\sqrt{3}} \frac{1+t^2}{2t}\cdot\frac{2}{1+t^2}\,dt = [\log t]_{1/\sqrt{3}}^{\sqrt{3}} = \log 3.$$

(5) Again by the substitution $t = \tan\frac{1}{2}x$, noting that $\cos x = (1-t^2)/(1+t^2)$,

$$\int_0^{\pi/2} \frac{\sin x}{\sqrt{1+\cos x}}\,dx = \int_0^1 \frac{2t}{1+t^2}\cdot\frac{1}{\sqrt{1+(1-t^2)(1+t^2)^{-1}}}\cdot\frac{2}{1+t^2}\,dt$$
$$= \int_0^1 \frac{4t}{\sqrt{2}(1+t^2)^{3/2}}\,dt = \left[\frac{-4t}{\sqrt{2}(1+t^2)^{1/2}}\right]_0^1 = 2.$$

In (1) we get rid of the unappealing square root, hoping that we will not exchange one awkward integral for another: fortunately we don't. The factor $\sqrt{1-x^2}$ in (2) is a signal to try $x = \sin t$; had it been $\sqrt{x^2-1}$ ($\sqrt{x^2+1}$) we would try $x = \cosh t$ ($x = \sinh t$) instead. The $t = \tan\frac{1}{2}x$ substitution in (4) and (5) is useful for converting rational functions of $\sin x$ and $\cos x$ into rational functions of t. For rational functions of $\cos^2 x$ and $\sin^2 x$ the substitution $t = \tan x$ is preferable.

6.9 Exercise example. Evaluate the following integrals by making the indicated substitution (or otherwise):

(i) $\displaystyle\int_0^1 \frac{x}{1+\sqrt{x}}\,dx \quad (t = \sqrt{x})$, (ii) $\displaystyle\int_0^1 \frac{1}{1+e^x}\,dx \quad (t = e^x)$,

(iii) $\displaystyle\int_0^1 \tan^3 x\,dx \quad (t = \tan x)$, (iv) $\displaystyle\int_0^{\pi/2} \frac{\cos x}{\sqrt{1+\cos x}}\,dx \quad (t = \tan\frac{1}{2}x)$,

(v) $\displaystyle\int_{\log 2}^{\log 3} \frac{x^2}{\sqrt{x^2-1}}\,dx \quad (\cosh t = x)$, (vi) $\displaystyle\int_1^{\sqrt{3}} \frac{1}{x^2\sqrt{1+x^2}}\,dx \quad (\sinh t = x)$.

6.10 Exercise example.

(a) By making the substitution $t = \tan\frac{1}{2}x$ show that

$$\int_0^{\pi/2} \frac{1}{1+\sin x+\cos x}\,dx = \frac{1}{2}\log 2.$$

(b) By making the substitution $t = \tan x$ (or $t = (b/a)\tan x$) prove that

$$\int_{-\pi/4}^{\pi/4} \frac{1}{a^2\cos^2 x + b^2\sin^2 x}\, dx = \frac{2}{ab}\tan^{-1}\left(\frac{b}{a}\right) \quad (a, b > 0).$$

6.11 Integration by parts. Let u, v have continuous derivatives on $[a, b]$. Then

$$\int_a^b uv' = [uv]_a^b - \int_a^b u'v.$$

Proof. Apply the FTC to $(uv)'$; this is $uv' + u'v$, and is continuous. $\square$

6.12 Example. The following are typical applications of 6.11. In all of them the chosen functions u and v clearly satisfy the conditions of 6.11.

(1) $\int_0^\pi x\cos x\, dx = [x\sin x]_0^\pi - \int_0^\pi \sin x\, dx = [\cos x]_0^\pi = -2.$

(2) Let $I = \int_0^1 e^x \sin \pi x\, dx$. Integrating by parts twice,

$$\begin{aligned} I &= [e^x \sin \pi x]_0^1 - \frac{1}{\pi}\int_0^1 e^x \cos \pi x\, dx \\ &= -\frac{1}{\pi}[e^x \cos(\pi x)]_0^1 - \frac{1}{\pi^2}\int_0^1 e^x \sin \pi x\, dx = \frac{e+1}{\pi} + \frac{1}{\pi^2}I. \end{aligned}$$

Hence $I = \pi(e+1)/(\pi - 1)$.

(3) $$\begin{aligned} \int_0^1 x^3 e^{-x^2/2}\, dx = \int_0^1 x^2 \cdot xe^{-x^2/2}\, dx &= \left[x^2 \cdot -e^{-x^2/2}\right]_0^1 + \int_0^1 2x.e^{-x^2/2}\, dx \\ &= -e^{-1/2} + \left[-2e^{-x^2/2}\right]_0^1 = (1 - 3e^{-1/2}). \end{aligned}$$

(4) $$\begin{aligned} \int_0^1 \tan^{-1} x\, dx = \int_0^1 1 \cdot \tan^{-1} x\, dx &= [x\tan^{-1} x]_0^1 - \int_0^1 \frac{x}{1+x^2}\, dx \\ &= \frac{\pi}{4} - \frac{1}{2}\log 2. \end{aligned}$$

Note the various ideas in play here. In (1) we have a product of two functions, for each of which we can write down both the derivative and an antiderivative. Differentiating x eliminates it and gives an integral to which the FTC applies. For (3) we would like, as in (1), to reduce the power of x by differentiating it. However $u = x^3$ and $v' = e^{-x^2/2}$ won't work, because we cannot calculate an antiderivative v. Attaching a factor of x to $e^{-x^2/2}$ to make the derivative of $-e^{-x^2/2}$ solves the problem.

In (2) we can at the first stage choose $u = \sin \pi x$ and $v = e^x$ (as we did), or vice versa. At the second stage we retain v as e^x, taking $u = \cos(\pi x)$, and reach a soluble equation for I; the opposite choice leads unhelpfully back to our starting point. We have in (4) a single function which we can differentiate but not directly integrate. The trick is to take $v = 1$.

6.13 Exercise example. Apply integration by parts to evaluate the following:

$$\text{(i)} \int_1^2 x^p \log x \,\mathrm{d}x \quad (p \neq -1), \quad \text{(ii)} \int_1^a \log x \,\mathrm{d}x \quad (a > 1), \quad \text{(iii)} \int_0^1 x^2 \sin^2 x \,\mathrm{d}x,$$

$$\text{(iv)} \int_0^1 \frac{x}{1+x^3}\,\mathrm{d}x, \qquad \text{(v)} \int_0^1 \frac{x^2}{(x^2+1)^{3/2}}\,\mathrm{d}x, \qquad \text{(vi)} \int_0^3 \sqrt{9-x^2}\,\mathrm{d}x.$$

6.14 Reduction formulae. This technique is better illustrated by example than presented in general terms.

Consider $I_n = \int_0^1 x^n \mathrm{e}^{-x}\,\mathrm{d}x$ $(n = 0, 1, 2, 3, \ldots)$. By the FTC, $I_0 = 1 - \mathrm{e}^{-1}$. For $n \geqslant 1$, we integrate by parts with $u(x) = x^n$, $v(x) = -\mathrm{e}^{-x}$ to get

$$I_n = \left[-x^n \mathrm{e}^{-x}\right]_0^1 + n\int_0^1 x^{n-1}\mathrm{e}^{-x}\,\mathrm{d}x = -\mathrm{e}^{-1} + nI_{n-1}.$$

This gives in turn $I_1 = 1 - 2\mathrm{e}^{-1}$, $I_2 = 2 - 5\mathrm{e}^{-1}$, $I_3 = 6 - 16\mathrm{e}^{-1}$, and so on.

Here is a more complicated example.

$$I_n := \int_0^{\pi/2} \cos^n x \,\mathrm{d}x.$$

We can calculate I_0 and I_1 immediately by the FTC. For $n \geqslant 2$,

$$\int_0^{\pi/2} \cos x \cos^{n-1} x \,\mathrm{d}x = \left[\sin x \cos^{n-1} x\right]_0^{\pi/2} +$$

$$(n-1)\int_0^{\pi/2} \sin x \ \sin x \ \cos^{n-2} x \,\mathrm{d}x = (n-1)\int_0^{\pi/2} (1-\cos^2 x)\cos^{n-2} x \,\mathrm{d}x.$$

Hence $nI_n = (n-1)I_{n-2}$. So $I_n = (n-1)(n-3)I_{n-4}/(n(n-2))$, and so on. Finally,

$$I_{2n} = \frac{(2n-1)(2n-3)\ldots 1}{2n(2n-2)\ldots 2} I_0 = \frac{(2n)!}{2^{n+1}n!^2}\pi,$$

$$I_{2n+1} = \frac{2n(2n-2)\ldots 2}{(2n+1)(2n-1)\ldots 3} I_1 = \frac{(2^n n!)^2}{(2n+1)(2n)!^2}.$$

The strategy was to derive a *reduction formula* linking I_n with I_m, where $m < n$. Iterating downwards we ultimately reach an integral whose value is easy to find, and are then able to obtain I_n in general.

Many other integrals depending on a parameter n $(n \geqslant 1$, say) can be handled in a similar way by obtaining a difference equation, provided one or more of the integrals is already known. In simple cases the difference equation can be explicitly solved; at the least, it can be used step-by-step to give explicit answers for particular values of n. For an example, see 6.15(b).

6.15 Exercise example.

(a) Let I_n be defined by $n!I_n := \int_0^\pi (x-\pi)^n \sin x \, \mathrm{d}x$. Evaluate I_0 and I_1. Prove that $I_n = -I_{n-2}$ $(n = 2, 3, \dots)$, and hence evaluate I_n for all $n \geqslant 2$.

(b) Prove that if a is a positive constant and I_n is defined by

$$I_n = \int_0^\pi \mathrm{e}^{-a\cos x} \sin^n x \, \mathrm{d}x \quad (n = 0, 1, 2, \dots)$$

then I_n satisfies

$$a^2 I_n = -(n-1)(n-2)I_{n-2} + (n-1)(n-3)I_{n-4} \quad (n \geqslant 4).$$

Evaluate I_1, I_3, I_5, and I_7.

6.16 Stocktaking. It is possible to write down elaborate procedures for integrating functions of various general types, such as rational functions and polynomial functions of sines and cosines. However the majority of applications of integration involve a relatively small number of forms. Therefore a 'be prepared for anything' attitude is not necessary. The methods described above suffice for integrating those combinations of standard functions which arise regularly.

The techniques of this chapter are not all powerful. Not every continuous function can be exhibited as the derivative of a function built up from familiar functions such as polynomials, trigonometric functions, and exponentials. For example, no amount of ingenuity will produce such a recipe for an antiderivative of e^{x^2} or of $\sqrt{(1-x^2)(1-k^2x^2)}$ $(0 < k < 1)$. What can be done in such cases?

First of all, it is often only integrals over certain particular intervals that are needed, for example over $[0, 2\pi]$ or $(-\infty, \infty)$, and can apply techniques quite different from those used here. We have in mind in particular the methods of complex analysis, as described for example in [**13**].

Another approach is as old as integral calculus itself. To evaluate $\int_a^b f$ we would represent f by its Maclaurin expansion and integrate the terms of the sum one at a time. Quite a lot of work is involved in validating this procedure, under appropriate technical conditions. See Chapters 8 and 17, where we explain why 'term-by-term' integration of an infinite series is not always allowed, and discuss when it is.

We also note that differentiating with respect to a parameter may yield new integrals from old ones; see Chapter 20.

Finally we might evaluate an integral approximately, using methods such as Simpson's rule. We discuss numerical integration in the next chapter.

So far we have relied on the FTC, with a passing mention in 6.6 of the Indefinite Integral Theorem. We now present some applications of the latter.

6.17 Integral equations. Consider the problem of finding a solution to

$$f(x) - \lambda \int_0^x t f(t)\,\mathrm{d}t = g(x) \quad (x \in [0, b]),$$

where g is a given differentiable function on $[0, b]$, λ is some constant, and the continuous function f is to be found. By the Indefinite Integral Theorem, $\int_0^x t f(t)\,\mathrm{d}t$ is differentiable, with derivative $x f(x)$. Putting $x = 0$ we obtain $f(0) = g(0)$. The given equation f implies that f is differentiable too. Thus

$$f'(x) - \lambda x f(x) = g'(x),$$

subject to the initial condition $f(0) = g(0)$. To solve this differential equation we switch to a dummy variable t and multiply by the integrating factor $\mathrm{e}^{-\lambda t}$. Then by the FTC we have

$$f(x)\mathrm{e}^{-\lambda x} = g(x) + \lambda \int_0^x \mathrm{e}^{\lambda(x^2 - t^2)/2} g(t)\,\mathrm{d}t \quad (x \in [0, b]).$$

This is an example of a *Volterra integral equation*, of the form

$$f(x) - \lambda \int_a^x K(x, t) f(t)\,\mathrm{d}t = g(x) \quad (x \in [a, b]).$$

Such equations, and variants thereof, occur widely in applied mathematics. Our simple example shows the conversion of an integral equation to a differential equation. This process, which is reversible, holds under quite weak conditions for Volterra equations in general, and the symbiotic relationship between integral and differential equations underlies Picard's proof of the existence theorem for solutions of first-order differential equations. This is outlined in Exercise 8.6.

Integral equations form too specialized and too large a topic for us to consider them systematically in this book, though we do provide in later chapters many of the technical tools for handling integral operators. For an elementary account see for example **[8]**, and for a more advanced treatment see **[9]**.

Exercises

6.1 Find an antiderivative F for the each of the following functions f:

(i) $\dfrac{x}{\sqrt{x^2 - 4}}$, (ii) $\dfrac{1}{x(x+2)^2}$, (iii) $\dfrac{\sin x + \cos x}{\sin x - \cos x}$, (iv) $\dfrac{1}{x^2 + 5x + 6}$,

(v) $\dfrac{(\log x)^2}{x}$, (vi) $\dfrac{\mathrm{e}^x}{1 + \mathrm{e}^x}$, (vii) $\sin^2 x \cos^3 x$, (viii) $\dfrac{1}{x^2 + 4x + 13}$,

(ix) $\dfrac{1}{\sqrt{x(4 - x)}}$, (x) $x\sqrt{9 + x^2}$, (xi) $\dfrac{1}{1 - x + x^2}$, (xii) $\operatorname{cosec} x \sin 3x$.

Indicate any restrictions on $[a, b]$ that must be imposed in order that the FTC should yield $\int_a^b f = F(b) - F(a)$.

6.2 Prove that, for m, n non-negative integers,

$$\int_{-\pi}^{\pi} \cos mx \sin nx \, \mathrm{d}x = 0, \qquad \int_{-\pi}^{\pi} \cos mx \cos nx \, \mathrm{d}x = \begin{cases} 0 & \text{if } m \neq n, \\ 2\pi & \text{if } m = n = 0, \\ \pi & \text{if } m = n \geqslant 1, \end{cases}$$

$$\int_{-\pi}^{\pi} \sin mx \sin nx \, \mathrm{d}x = \begin{cases} 0 & \text{if } m \neq n \text{ or } m = n = 0, \\ \pi & \text{if } m = n \geqslant 1. \end{cases}$$

[These formulae underlie the theory of Fourier series; see Chapter 29.]

6.3 Let $f \in \mathrm{C}[0, a]$. Prove that $\int_0^a f(x)\, \mathrm{d}x = \int_0^a f(a - x)\, \mathrm{d}x$. Hence prove that

$$\int_0^{\pi} \frac{x \sin x}{1 + \cos^2 x} \, \mathrm{d}x = \frac{1}{4}\pi^2.$$

6.4 Let $g \in \mathrm{C}[0, \alpha]$ have a strictly positive derivative and satisfy $g(0) = 0$ and $g(\alpha) = \beta$. Prove that

$$\int_0^{\beta} f(t) \, \mathrm{d}t + \int_0^{\alpha} g(s) \, \mathrm{d}s = \alpha\beta,$$

where f is the inverse function to g (whose continuity you may assume).

6.5 Use repeated integration by parts to prove that

$$\int_{-1}^{1} (1 - x^2)^n \, \mathrm{d}x = \frac{2^{2n+1}(n!)^2}{(2n+1)(2n)!} \quad (n = 0, 1, 2, \ldots).$$

[Hint: write $(1 - x^2)^n$ as $(1 + x)^n (1 - x)^n$.]

6.6 Let $f \in \mathrm{C}^2[a, b]$. Prove that $\displaystyle\int_a^b x f''(x) \, \mathrm{d}x = (bf'(b) - f(b)) - (af'(a) - f(a))$.

6.7 Checking that any applications you make of 6.6 or 6.11 are valid, prove that

$$\int_0^{1/2} \frac{x \mathrm{e}^{\sin^{-1} x}}{\sqrt{1 - x^2}} \, \mathrm{d}x = \frac{1}{4}\left(2 + \mathrm{e}^{\pi/6}(1 - \sqrt{3})\right).$$

6.8 Prove that, for $0 < a < b < \infty$,

$$\left| \int_a^b \sin(x^2) \, \mathrm{d}x \right| \leqslant a^{-1}.$$

[Hint: regard the integrand as $(x \sin(x^2))/x$ and integrate by parts.]

6.9 Let $I_n = \int_0^{\pi/2} \sin^n x \, \mathrm{d}x$.

(a) Prove that $nI_n = (n - 1)I_{n-2}$ and hence prove by induction that

$$I_{2n} = \frac{(2n)!}{(2^n n!)^2} \cdot \frac{\pi}{2} \quad \text{and} \quad I_{2n+1} = \frac{(2^n n!)^2}{(2n+1)!}.$$

(b) Prove that

$$\frac{2n}{2n+1} \leqslant \frac{I_{2n+1}}{I_{2n}} \leqslant 1 \quad (n \geqslant 1).$$

(c) Deduce Wallis's formula, *viz.*

$$\lim_{n\to\infty} \frac{1}{2n+1} \cdot \frac{2^{4n+1}(n!)^4}{(2n)!^4} = \pi.$$

6.10 For $m, n = 0, 1, 2, \ldots$ let

$$I_n = \int_{-\pi/2}^{\pi/2} e^{-mx} \cos^n x \, dx.$$

Prove that, for each m,

$$(m^2 + n^2) I_n = n(n-1) I_{n-2} \quad (n \geqslant 2)$$

and hence evaluate I_n.

6.11 For $m, n \geqslant 1$, let

$$I_{m,n} = \int_0^\pi \cos^m x \sin nx \, dx.$$

Prove that

$$(m+n) I_{m,n} = (-1)^{m+n} - 1 + m I_{m-1,n-1}$$

and deduce that $I_{m,n} = 0$ whenever m, n are both even or both odd.

6.12 For $n = 0, 1, 2, \ldots$, let $I_n = \int_0^1 x^n \sqrt{1-x} \, dx$. Prove that

$$I_{n+1} - I_n = -\frac{3}{2(n+1)} I_{n+1}.$$

Deduce that $(2n+5) I_{n+1} = 2(n+1) I_n$ and hence show that

$$I_n = \frac{2^{2n+2} n!(n+1)!}{(2n+3)!} \quad (n \geqslant 0).$$

6.13 (a) Solve $f(x) = 1 + \int_0^x f(t) \, dt$,

(b) Solve $f(x) = 1 - \int_0^x \sin t f(t) \, dt$.

(c) Integrate by parts to solve $f(x) = 1 + \int_0^x (t-x) f(t) \, dt$.

6.14 Let $f \in C[0,1]$ satisfy

$$f(x) - \lambda \int_0^1 (1 - 3xt) f(t) \, dt = g(x) \quad (\lambda > 0).$$

Write $a = \int_0^1 f(t) \, dt$ and $b = \int_0^1 t f(t) \, dt$, so that $f(x) = \lambda a - 3\lambda b x + g(x)$.

(a) Prove that a and b must satisfy

$$(1-\lambda)+\frac{3}{2}\lambda b=\int_0^1 g(x)\,\mathrm{d}x, \quad \text{and} \quad -\frac{1}{2}\lambda a+(1+\lambda)b=\int_0^1 xg(x)\,\mathrm{d}x.$$

(b) Prove that there is a unique solution for a and b, and hence for f, provided $\lambda \neq 2$.

(c) Prove that if $\lambda = 2$ the integral equation has a solution only if

$$\int_0^1 g(x)\,\mathrm{d}x = \pm\int_0^1 xg(x)\,\mathrm{d}x.$$

[Here we have a *Fredholm integral equation* (in which the integral term takes the form $\int_a^b K(x,t)f(t)\,\mathrm{d}t$). This simple example shows that the behaviour of such equations is quite different from that of equations of Volterra type.]

6.15 Let $f \in \mathrm{C}[0,1]$. Let

$$f(x)=\int_0^1 g(t)(\min\{x,t\}-xt\})\,\mathrm{d}t.$$

By splitting the interval $[0,1]$ into the intervals $[0,x]$ and $[x,1]$ prove that f satisfies $f''(x)=g(x)$, $f(0)=f(1)=0$, and hence find f.

7 Approximations

Suppose you are given a particular continuous function f on a compact interval $[a, b]$ and want to work out $\int_a^b f$ but can't find an antiderivative for f. Your calculator will, in many cases, present you with an answer, to an impressive number of decimal places. Methods for estimating integrals are obviously programmed into calculators—how is this done and how accurate are the results? Numerical integration is a specialized topic, and we do no more here than indicate some rudimentary techniques and show how our theory gives error estimates.

7.1 Linear approximation and the simple trapezium rule. Let $f \in \mathrm{C}^2[a, b]$ and let $0 \leqslant h \leqslant b - a$. Consider

$$\begin{aligned} h^2 \int_0^1 (1-t) f''(a+th)\,\mathrm{d}t &= [h(1-t)f'(a+th)]_0^1 + h\int_0^1 f'(a+th)\,\mathrm{d}t \quad \text{(by 6.11)} \\ &= -hf'(a) + f(a+h) - f(a) \quad \text{(by FTC)}. \end{aligned}$$

Observe that we may rewrite this as

$$f(a+h) = f(a) + hf'(a) + h^2 \int_0^1 (1-t) f''(a+th)\,\mathrm{d}t,$$

or, putting $x = a + h$, as

$$(\dagger) \qquad f(x) = f(a) + (x-a)f'(a) + (x-a)^2 \int_0^1 (1-t) f''(a+th)\,\mathrm{d}t.$$

We may view this as giving a linear approximation, $f(a) + (x-a)f'(a)$, to $f(x)$, with $(x-a)^2 \int_0^1 (1-t)f''(a+th)\,\mathrm{d}t$ as the error. Said another way, we have found the first two terms in a Taylor series for f, with the remainder expressed as an integral.

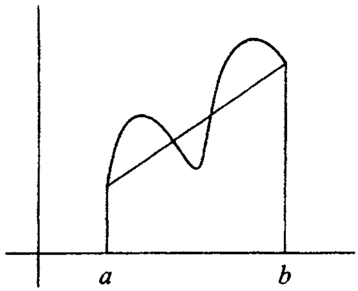
Now consider

$$\begin{aligned} h^2 \int_0^1 t(1-t) f''(a+th)\,\mathrm{d}t &= \big[ht(1-t)f'(a+th)\big]_0^1 - h\int_0^1 (1-2t) f'(a+th)\,\mathrm{d}t \\ &= -\Big[(1-2t)f(a+th)\Big]_0^1 - 2\int_0^1 f(a+th)\,\mathrm{d}t. \end{aligned}$$

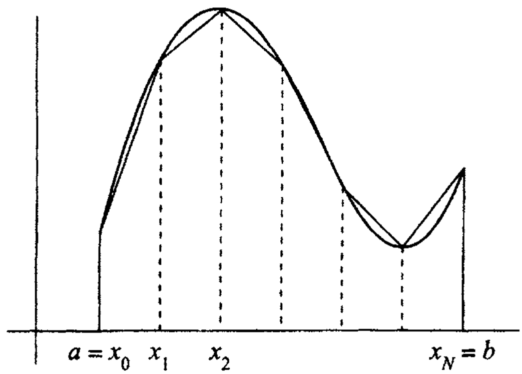
Hence, putting $h = b - a$,

$$(\ddagger) \qquad \int_a^b f(x)\,\mathrm{d}x = \frac{1}{2}h\,(f(a) + f(b)) - \frac{1}{2}h^3 \int_0^1 t(1-t) f''(a+th)\,\mathrm{d}t.$$

Look at Fig. 7.1(a): $(b-a)(f(a) + f(b))/2$ is the area of the shaded trapezium. It is the exact integral of the linear approximation $f(a) + (x-a)f'(a)$ to $f(x)$



(a) simple trapezium rule



(b) N-step trapezium rule

Figure 7.1

on $[a, b]$. Clearly this provides an estimate to $\int_a^b f$, albeit a crude one in general. What our analysis has done is to give a formula for the error, as an integral involving f''. Let us estimate this error, assuming that $|f''| \leqslant M$ on $[a, b]$. Note that $t(1-t) \geqslant 0$ on $[0, 1]$. Hence

$$-Mt(1-t) \leqslant t(1-t)f''(a+th) \leqslant Mt(1-t),$$

from which we get, by property (P),

$$\left|\frac{1}{2}h^3 \int_0^1 t(1-t)f''(a+th)\,\mathrm{d}t\right| \leqslant \frac{1}{2}h^3 M \int_0^1 t(1-t)\,\mathrm{d}t = \frac{1}{12}Mh^3.$$

In a similar way we can show that the error $h^2 \int_0^1 (1-t)f''(a+th)\,\mathrm{d}t$ in our Taylor expansion is bounded by $Mh^2/2$.

Therefore in both the linear approximation to $f(x)$ in (†) and the trapezium rule approximation to $\int_a^b f$ in (‡) the goodness of the approximation depends on two factors: the length, h, of the interval $[a, b]$ over which the approximation is taken, and the bound M on $|f''|$. Not surprisingly the most accurate results are obtained when h is small and when f does not vary too rapidly, as measured by the rate of change of the gradient f' of f. We shall see below that we can improve on (†) and (‡) by making stronger differentiability assumptions on f.

7.2 Example. To illustrate, we apply the trapezium rule to $\int_0^1 (1+x^2)^{-1}\,\mathrm{d}x$, whose correct value is $\tan^{-1} 1$, equal to $\pi/4$, or approximately 0.785. The trapezium rule supplies the approximate value 0.75.

For $f(x) = (1+x^2)^{-1}$ we have $f''(x) = 8x^2(1+x^2)^{-3} - 2(1+x^2)^{-2}$. It is easy to check that $f''' \geqslant 0$ on $[0, 1]$, so on this interval f'' attains its maximum and minimum at $x = 1$ and $x = 0$, respectively. Hence $-2 \leqslant f'' \leqslant 1/2$ on

$[0,1]$. Therefore the error in the trapezium rule is bounded by $1/6$. This is not particularly impressive, and more refined methods are clearly needed.

Before extending the results above we give a mean value theorem for integrals which refines 5.2(b) and is useful in particular for error estimation.

7.3 Integral Mean Value Theorem. Let $f, g \in C[a,b]$, with $g \geqslant 0$. Then there exists $\xi \in [a,b]$ such that

$$\int_a^b fg = f(\xi)\int_a^b g.$$

[The continuity of g is used in the proof only to ensure that fg is integrable; once we have developed the necessary theory we may assume merely that g is integrable on $[a,b]$ (see 23.2(a)). Continuity of f is essential.]

Proof. The Boundedness Theorem, 2.23, ensures that f is bounded. Let $m := \inf f$, $M := \sup f$ on $[a,b]$. Then $m \leqslant f \leqslant M$ and $g \geqslant 0$ imply

$$mg \leqslant fg \leqslant Mg,$$

whence by 5.2 and 5.9

$$m\int_a^b g \leqslant \int_a^b fg \leqslant M\int_a^b g.$$

If $\int_a^b g = 0$ then $\int_a^b fg = 0$ too, and the required result holds trivially. Otherwise, $(\int_a^b fg)/(\int_a^b g) \in [m,M]$ and so is of the form $f(\xi)$ for some $\xi \in [a,b]$, by the remark following the Intermediate Value Theorem, 2.24. □

7.4 Example. For any $x \in [-1,1)$ there exists $\tau \in [0,1]$ (depending on x) such that

$$I_n(x) := \int_0^1 \frac{(1-t)^n}{(1+xt)^{n+1}}\,\mathrm{d}t = (1+x\tau)^{-(n+1)}\int_0^1 (1-t)^n\,\mathrm{d}t = \frac{1}{n+1}(1+x\tau)^{-(n+1)}.$$

Here we have applied the Integral MVT with $f(t) = (1+xt)^{-(n+1)}$ and $g(t) = (1-t)^n$; both $f, g \in C[0,1]$ and $g \geqslant 0$. Since we do not know what τ is, we seek lower and upper bounds for $I_n(x)$ which do not involve τ. We see easily that

$$0 \leqslant \int_0^1 \frac{(1-t)^n}{(1+xt)^{n+1}}\,\mathrm{d}t \leqslant \begin{cases} (n+1)^{-1} & (0 \leqslant x \leqslant 1), \\ (n+1)^{-1}(1+x)^{-(n+1)} & (-1 < x < 0). \end{cases}$$

If we are interested in the behaviour of $I_n(x)$ as $n \to \infty$ the above bound is no help for $-1 < x < 0$ since then $(n+1)^{-1}(1+x)^{-(n+1)} \to \infty$ as $n \to \infty$. However we can argue differently: $0 \leqslant t \leqslant 1$ and $-1 < x < 0$ imply $(1-t) < (1+xt)$ and $(1+xt)^{-1} \leqslant (1+x)^{-1}$, so, by 5.2,

$$0 \leqslant \int_0^1 \frac{(1-t)^n}{(1+xt)^{n+1}}\,\mathrm{d}t \leqslant (1+x)^{-1} \quad (-1 < x < 0).$$

7.5 Taylor's Theorem with integral remainder. Let $f\colon [a, a+h] \to \mathbb{R}$ be such that $f \in \mathrm{C}^{(n+1)}[a, a+h]$ (replace $[a, a+h]$ by $[a+h, a]$ if $h < 0$). Then

$$f(a+h) = f(a) + hf'(a) + \cdots + \frac{h^n}{n!} f^{(n)}(a) + R_n,$$

where the remainder term R_n is given by

$$R_n = \frac{h^{n+1}}{n!} \int_0^1 (1-t)^n f^{(n+1)}(a+th)\,\mathrm{d}t.$$

Proof. We use induction. For $n = 0$ we have, by the FTC,

$$R_0 = \int_0^1 f'(a+th)\,\mathrm{d}t = \left[\frac{1}{h} f(a+th)\right]_{t=0}^1 = \frac{1}{h}(f(a+h) - f(a)),$$

so the result is true in this case. Assume that it is true for $n = k-1$ ($k \geqslant 1$). Integrating by parts and using the inductive hypothesis,

$$\begin{aligned} R_k &= \frac{h^{k+1}}{k!} \int_0^1 (1-t)^k f^{(k+1)}(a+th)\,\mathrm{d}t \\ &= \left[\frac{h^k}{k!}(1-t)^k f^{(k)}(a+th)\right]_0^1 + \frac{h^k}{(k-1)!} \int_0^1 (1-t)^{k-1} f^{(k)}(a+th)\,\mathrm{d}t \\ &= f(a+h) - \Big(f(a) + hf'(a) + \cdots + \frac{h^k}{k!} f^{(k)}(a)\Big). \end{aligned}$$

□

7.6 Taylor expansions: remarks. Make the same assumptions as in 7.5. Writing x in place of $a+h$ we may regard the formula in 7.5 as giving a polynomial approximation

$$p(x) := f(a) + (x-a)f'(a) + \cdots + \frac{1}{n!}(x-a)^n f^{(n)}(a)$$

to $f(x)$ with error

$$R_n := \frac{h^{n+1}}{n!} \int_0^1 (1-t)^n f^{(n+1)}(a+th)\,\mathrm{d}t,$$

where $h = x - a$; here R_n depends on f, a, and h (or x). By the Integral MVT, there exists $\xi \in [a, a+h]$ such that

$$R_n = \frac{h^{n+1}}{n!} f^{(n+1)}(\xi) \int_0^1 (1-t)^n\,\mathrm{d}t = \frac{h^{n+1}}{(n+1)!} f^{(n+1)}(\xi).$$

This re-expresses the remainder in the form due to Lagrange.

For $f \in \mathrm{C}^\infty[a, a+h]$ (or $\mathrm{C}^\infty[a+h, a]$ for $h < 0$) we obtain as infinite series expansion for $f(a+h)$ in powers of h,

$$f(a+h) = \sum_{k=0}^\infty \frac{h^k}{k!} f^{(k)}(a), \quad \text{provided } R_n \to 0 \text{ as } n \to \infty.$$

Examples 7.7 and 7.8 are instances of these processes.

7.7 Example. We apply the preceding results to the logarithm. Putting $f(x) = \log(1+x)$ we have $f^{(k)}(x) = (-1)^{k-1}(k-1)!(1+x)^{-k}$. Hence, for $n = 1, 2, \dots$ and $-1 < x \leqslant 1$,

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^n \frac{x^n}{n} + (-1)^n x^{n+1} \int_0^1 \frac{(1-t)^n}{(1+xt)^{n+1}}\, \mathrm{d}t.$$

The integral in the remainder term, R_n, was estimated in 7.4. Using the bounds obtained there,

$$|R_n| \leqslant \begin{cases} (n+1)^{-1}|x|^{n+1} & (0 \leqslant x \leqslant 1), \\ (1+x)^{-1}|x|^{n+1} & (-1 < x < 0), \end{cases}$$

In either case we deduce that $R_n \to 0$ as $n \to \infty$. Hence

$$\log(1+x) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k} \quad (-1 < x \leqslant 1).$$

7.8 Exercise example. Prove that, for $n = 1, 2, \dots,$

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots + (-1)^{n-1} \frac{x^{2n-1}}{2n-1} + (-1)^n \int_0^x \frac{t^{2n}}{1+t^2}\, \mathrm{d}t.$$

By applying the Integral MVT or by estimating the integral directly deduce that

$$\tan^{-1} x = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^{2k-1}}{2k-1} \quad (|x| \leqslant 1).$$

We now return to the numerical evaluation of integrals, and describe ways of improving on the simple trapezium rule.

7.9 The N-step trapezium rule. Let $f \in \mathrm{C}^2[a,b]$. We can refine (‡) in 7.1 by subdividing the interval $[a,b]$ into N equal subintervals, applying the simple trapezium rule on each subinterval I_r, and combining the results; see Fig. 7.1(b). Specifically, let $x_r := a + rh$ $(r = 0, 1, \dots, N)$, and $I_r := [x_{r-1}, x_r]$ $(r = 1, \dots, N)$, where $h := (b-a)/N$. Let

$$M_r := \sup\{\, |f''(x)| \mid x \in I_r \,\}.$$

Then, with $h = (b-a)/N$,

$$\Big| \int_a^b f(x)\, \mathrm{d}x - \frac{h}{2}\Big(f(a) + 2f(a+h) + 2f(a+2h) + \dots$$
$$+ 2f(a+(N-1)h)) + f(a+Nh)\Big)\Big| \leqslant \frac{h^3}{12}(M_1 + \cdots + M_N).$$

A more convenient, but less tight, error bound is $M(b-a)^3/(12N^2)$, in which we use the universal constant $M := \sup\{\,|f''(x)| \mid x \in [a,b]\,\}$ to bound f'' on each interval I_r. Look at $f(x) = \log x$ on $[1, N+1]$, and split this interval into N subintervals of length 1. Here $M_r = r^{-2}$ and $M = 1$. The tighter error bound is $\left(1 + 2^{-2} + \cdots + N^{-2}\right)/12$ while that using M is $N/12$. The latter is not bounded as $N \to \infty$ but the former is bounded—a fact we exploit in proving Stirling's formula in 7.16.

7.10 Exercise example. Use the N-step trapezium rule with $N = 2$ and $N = 4$ to approximate $\int_0^1 (1+x^2)^{-1}\,\mathrm{d}x$, and give an estimate for the error in each case.

7.11 Remarks. We can always break an integral over a compact interval $[a,b]$ into the sum of a finite number of integrals over equal intervals of arbitrarily small length. Thus our prime concern is to derive good approximations for $\int_a^b f$ when $(b-a)$ is small. We aim for an error bound of the form $C(b-a)^k$, where C is some constant and k is as large as possible. In the trapezium rule we achieved $k = 3$, with $C = M/12$, where M is a bound on $|f''|$. Can we do better?

In 7.1 we approximated $f \in \mathrm{C}^2[a,b]$ by a linear polynomial. Our discussion of Taylor series in 7.5 showed that for functions with sufficiently many derivatives we can improve our approximation by using polynomials of higher degree. Likewise for $\int_a^b f$: the better-tempered f is, the better an approximation we may expect. For $f \in \mathrm{C}^4[a,b]$ we have Simpson's rule, which employs quadratic approximation and gives very accurate results. See for example **[12]** for illustrations.

7.12 Simpson's rule. Let $f \in \mathrm{C}^4[a,b]$ and assume $\left|f^{(4)}(x)\right| \leqslant M$ on $[a,b]$. Let $c = (a+b)/2$ and $h = (b-a)/2$ so $a = c-h$, $b = c+h$. By repeated integration by parts we can show that

$$\int_c^b \Big(\frac{1}{24}(b-x)^4 - \frac{1}{18}h(b-x)^3\Big) f^{(4)}(x)\,\mathrm{d}x$$
$$= \frac{1}{72}h^4 f^{(3)}(c) - \frac{1}{6}h^2 f(c) - \frac{1}{3}h\Big(f(b) + 2f(c)\Big) + \int_c^b f(x)\,\mathrm{d}x.$$

The Integral MVT may be applied to the left-hand side of the displayed equation to show that it is bounded by $Mh^5/180$. Likewise,

$$\int_a^c \Big(\frac{1}{24}(a-x)^4 + \frac{1}{18}h(b-x)^3\Big) f^{(4)}(x)\,\mathrm{d}x$$
$$= -\frac{1}{72}h^4 f^{(3)}(c) + \frac{1}{6}h^2 f(c) - \frac{1}{3}h\Big(f(a) + 2f(c)\Big) + \int_a^c f(x)\,\mathrm{d}x,$$

with the left-hand side bounded by $Mh^5/180$. We conclude that

$$\left|\int_a^b f(x)\,\mathrm{d}x - \frac{1}{6}(b-a)\Big(f(a) + 4f\big(\frac{\mathrm{a}+\mathrm{b}}{2}\big) + f(b)\Big)\right| \leqslant \frac{1}{2880}M(b-a)^5.$$

We next present a simple example of a different approach to approximate integration, known as Gaussian quadrature. We reveal the theory that underlies the method in Chapter 31.

7.13 Example. We shall approximate $\int_0^1 f$ by $a_1 f(x_1) + a_2 f(x_2)$, choosing the constants a_1, a_2 and the interpolation points x_1, x_2 so that the error is zero when f is a polynomial of degree $\leqslant 3$. We first assume that this is possible and show how to estimate the error, E, when f is an element of $C^4[0,1]$ with $|f^{(4)}(x)| \leqslant M$ on $[0,1]$. We may use Taylor's Theorem with Lagrange's form of the remainder to write

$$f(x) = p(x) + \frac{x^4}{4!} f^{(4)}(\xi_x) \quad \text{where } p(x) := f(0) + xf'(0) + \frac{x^2}{2!} f''(0) + \frac{x^3}{3!} f'''(0),$$

for some $\xi_x \in (0,1)$. Here $p(x)$ is a polynomial of degree $\leqslant 3$ and $|f(x) - p(x)| \leqslant M/24$ for every $x \in [0,1]$. We now have

$$\int_0^1 f(x)\,\mathrm{d}x = a_1 f(x_1) + a_2 f(x_2) + E,$$
$$\int_0^1 p(x)\,\mathrm{d}x = a_1 p(x_1) + a_2 p(x_2).$$

Hence, by subtraction, and using the Integral MVT on the integral term,

$$|E| \leqslant |a_1|\,|f(x_1) - p(x_1)| + |a_2|\,|f(x_2) - p(x_2)| + \frac{1}{24}\left|\int_0^1 x^4 f^{(4)}(\xi_x)\,\mathrm{d}x\right|$$
$$\leqslant \frac{1}{24} M\Big(|a_1| + |a_2| + \frac{1}{5}\Big).$$

We now show that a_1, a_2, x_1, x_2 can be chosen so that $\int_0^1 f = a_1 f(x_1) + a_2 f(x_2)$ whenever f is a polynomial of degree $\leqslant 3$. In particular we require the formula to be exact for $f(x) = (x - x_1)(x - x_2)$ and for $f(x) = x(x - x_1)(x - x_2)$, both of which are zero at x_1 and x_2. This means

$$0 = \int_0^1 x^2\,\mathrm{d}x - \alpha \int_0^1 x\,\mathrm{d}x + \beta \int_0^1 1\,\mathrm{d}x = \frac{1}{3} - \frac{1}{2}\alpha + \beta,$$
$$0 = \int_0^1 x^3\,\mathrm{d}x - \alpha \int_0^1 x^2\,\mathrm{d}x + \beta \int_0^1 x\,\mathrm{d}x = \frac{1}{4} - \frac{1}{3}\alpha + \frac{1}{2}\beta,$$

where we have written $\alpha = x_1 + x_2$ and $\beta = x_1 x_2$. Solving for α and β gives $\alpha = 1$ and $\beta = 1/6$. Therefore x_1 and x_2 are uniquely determined as the roots of the quadratic equation

$$0 = t^2 - (x_1 + x_2)t + x_1 x_2 = 0 = t^2 - \alpha t + \beta = t^2 - t + \frac{1}{6}.$$

We can now put $f(x) = (x - x_1)$ and $f(x) = (x - x_2)$ in turn to get

$$a_2(x_1 - x_2) = \frac{1}{2} - x_1 \quad \text{and} \quad a_1(x_2 - x_1) = \frac{1}{2} - x_2.$$

A straightforward calculation gives the values

$$x_1 = \frac{3+\sqrt{3}}{6}, \quad x_2 = \frac{3-\sqrt{3}}{6}, \quad a_1 = \frac{1}{2}, \quad a_2 = \frac{1}{2}.$$

We have indicated how to find approximations to integrals we cannot evaluate explicitly. Contrariwise, there are cases in which the integral is known and its value provides an estimate of an expression which approximates. We illustrate by presenting two famous limits.

7.14 Sums and integrals compared: the Integral Test. Let $f\colon [1,\infty) \to \mathbb{R}^+$ be continuous and decreasing. Define

$$I_n := \int_1^n f \quad \text{and} \quad s_n := \sum_{k=1}^{n} f(k).$$

Because f is decreasing,

$$(*) \qquad 1 \cdot f(k+1) \leqslant \int_k^{k+1} f \leqslant 1 \cdot f(k) \quad (k = 1, 2, \ldots)$$

by 5.2; see Fig. 7.2.

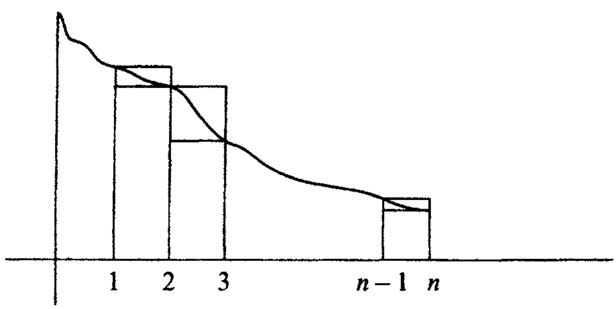


Figure 7.2

Adding these inequalities for $k = 1, \ldots, n-1$,

$$f(2) + \cdots + f(n) \leqslant \int_1^n f \leqslant f(1) + \cdots + f(n-1),$$

that is,

$$(**) \qquad s_n - f(1) \leqslant I_n \leqslant s_n - f(n).$$

Consider $\sigma_n := s_n - I_n$. By $(**)$, $f(1) \geqslant \sigma_n \geqslant f(n)$, so $\sigma_n \geqslant 0$ for all n. Also

$$\sigma_n - \sigma_{n-1} = (s_n - s_{n-1}) - (I_n - I_{n-1}) = f(n) - \int_{n-1}^n f \leqslant 0,$$

by $(*)$ with $k = n - 1$. By the Monotonic Sequence Theorem, $\{\sigma_n\}$ converges, to a limit σ such that $0 \leqslant \sigma \leqslant f(1)$. It follows that $\{s_n\}$ and $\{I_n\}$ either both converge to a finite limit or both tend to ∞.

This result can be used in various ways. For functions f for which I_n can be calculated it provides a convergence test for $\sum f(k)$—the **Integral Test**. Taking $f(x) = x^{-p}$, which certainly satisfies the stated conditions for $p \geqslant 0$, we have

$$I_n = \begin{cases} \left(n^{1-p} - 1\right)/(1-p) & (p \neq 1), \\ \log n & (p = 1), \end{cases}$$

which has a finite limit if and only if $p > 1$. Thus the Integral Test provides one proof of the fact that $\sum k^{-p}$ converges if and only if $p > 1$.

7.15 Euler's constant. The special case $f(x) = x^{-1}$ in 7.14 is important. From it we see that the nth partial sum of the divergent series $\sum k^{-1}$ tends to ∞ at the same rate as does $\log n$:

$$\gamma_n := \left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) - \log n$$

converges as $n \to \infty$. The limit, γ, is known as *Euler's constant*. Its value is approximately 0.5772156.

7.16 Stirling's formula. Let $f(x) = \log x$ $(x \in [1, n])$. Then $f''(x) = -x^{-2}$. We can apply the N-step trapezium rule to $\log x$ on $[1, n]$ with $N = (n-1)$ $(n \geqslant 2)$. In this case $h = 1$ and the rth interval is $[r, r+1]$ and M_r is $1/r^2$. We obtain

$$\int_1^n \log x \, dx - \left(\log 2 + \cdots + \log(n-1) + \frac{1}{2}\log n\right)$$
$$= e_1 + \cdots + e_{n-1}, \quad \text{where } |e_r| \leqslant \frac{1}{12} r^{-2} \quad (1 \leqslant r < n).$$

The integral above can be explicitly calculated by integrating by parts:

$$\int_1^n \log x \, dx = [x \log x]_1^n - \int_1^n \frac{1}{x} \, dx = n \log n - n + 1.$$

By properties of the logarithm, $\log 2 + \cdots + \log(n-1) + \frac{1}{2}\log n$ can be written more compactly as $\log(n!) - \frac{1}{2}\log n$. Therefore we can write

$$\log(n!) = (n + \frac{1}{2}) \log n - n + E_n,$$

where E_n tends to a limit as $n \to \infty$ because the series $\sum 1/k^2$ converges. We now take exponentials and deduce that there exists a finite non-zero constant C such that

$$a_n := \frac{n!}{\sqrt{n}(n/\mathrm{e})^n} \to C \quad \text{as } n \to \infty.$$

Finally we use a trick to compute C. By properties of limits of sequences, $a_n^2/a_{2n} \to C$. But

$$\frac{a_n^2}{a_{2n}} = \frac{(n!)^2}{n(n/\mathrm{e})^{2n}} \cdot \frac{\sqrt{2n}(2n/\mathrm{e})^{2n}}{(2n)!} = \sqrt{2}\frac{2^{2n} n!^2}{\sqrt{n}(2n)!},$$

which converges to $\sqrt{2\pi}$ by Wallis's formula, *viz.*

$$\lim_{n\to\infty} \frac{1}{2n+1} \cdot \frac{2^{4n+1} n!^4}{(2n)!^2} = \pi.$$

(Exercise 6.9). We have proved Stirling's formula:

$$\frac{n!}{\sqrt{2\pi n}(n/\mathrm{e})^n} \to 1 \quad \text{as } n \to \infty.$$

This is widely used in probability and elsewhere to give an asymptotic estimate for $n!$.

Exercises

7.1 Let $f \in \mathrm{C}[0,1]$. By applying the Integral MVT on $[0, 1/\sqrt{n}]$ and $[1/\sqrt{n}, 1]$ prove that

$$\int_0^1 \frac{nf(x)}{1+n^2x^2}\,\mathrm{d}x \to \frac{\pi}{2}f(0) \quad \text{as } n \to \infty.$$

7.2 Assume that $f\colon \mathbb{R} \to \mathbb{R}$ is differentiable and such that $f(0) = 1$ and $f'(x) = f(x)$ for all x. Prove that $f \in \mathrm{C}^\infty(\mathbb{R})$ and that

$$f(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (x \in \mathbb{R}).$$

[This characterizes the function defined by the exponential series.]

7.3 Show that the remainder term R_n in 7.5 may be expressed in the form

$$\frac{h^n}{(n-1)!}(1-\theta)^{n-1} f^{(n)}(a+\theta h) \quad \text{for some } \theta \in [0,1].$$

[Cauchy's form of the remainder.]

7.4 Let $f \in \mathrm{C}^2[0,1]$.

(a) Let $x \in (0,1)$. By applying Rolle's Theorem to the functions F and F' on $[0,1]$, where $F(t) = tf(1) - (t-1)f(0) + At(t-1)$ and A is chosen so that $F(x) = 0$, prove that there exists $\xi \in (0,1)$ such that

$$f(x) = xf(1) - (x-1)f(0) - \frac{1}{2}(x(x-1))f''(\xi).$$

Interpret this result graphically.

(b) Derive from (a) an alternative proof of the trapezium rule on $[0, 1]$.

7.5 (a) Obtain an approximate value of $\int_0^1 e^{-x^2/2}\,dx$, and give am estimate of the error using (i) the simple trapezium rule and (ii) Simpson's rule, giving a bound on the error in each case.
(b) Repeat (a) with $[0, 1]$ subdivided into four equal subintervals.

7.6 Use Simpson's rule, with four steps, to estimate $\int_1^2 e^{-x}/x\,dx$, and give a bound on the error.

7.7 (a) Show that $\sum 1/(k \log k)$ diverges.
(b) Find for which real values of $\alpha \neq 1$ the series $\sum 1/(k(\log k)^\alpha)$ converges.

8 Uniform convergence and power series

As far as the development of integration theory is concerned much of this chapter is a side-track. It rounds off the study of continuous functions on compact intervals, investigating limits and infinite sums of such functions, and in particular power series. The arguments are quite elementary, but firmly rooted in the ε-δ tradition. Our main technique is that of uniform convergence. We stress that uniform convergence rarely occurs (except locally) for sequences of functions on unbounded intervals. This makes it of limited value in our general theory. The powerful and widely applicable convergence theorems proved in Chapter 14 subsume the elementary results on limits of integrals we derive in this chapter.

We shall be concerned with the way limits interact with other concepts involving limits, either directly or indirectly. In particular, does a limit commute with another limit or with an integral? Sadly, the answer in general is 'no'. As a warning consider the double sequence $\{a_{mn}\}_{m,n\geqslant 1}$, where

$$a_{mn} = \begin{cases} 1 & \text{if } m < n, \\ 0 & \text{if } m \geqslant n. \end{cases}$$

We have

$$\lim_{m\to\infty}\left\{\lim_{n\to\infty} a_{mn}\right\} = 1 \quad \text{and} \quad \lim_{n\to\infty}\left\{\lim_{m\to\infty} a_{mn}\right\} = 0,$$

so the order in which the limits are taken affects the limiting value. Moral: limits must be handled with care.

8.1 Definition. Suppose that $\{f_n\}$ is a sequence of real-valued functions defined on some set S. We say $\{f_n\}$ *converges pointwise* to the function $f\colon S \to \mathbb{R}$ (and write $f = \lim f_n$ or $f_n \to f$ on S) if for each $x \in S$ the sequence $\{f_n(x)\}$ converges to $f(x)$.

Pointwise convergence is nothing unfamiliar. In saying, for example,

$$e^x = 1 + x + \frac{x^2}{2!} + \dots \qquad \text{on } \mathbb{R}$$

we mean precisely that the partial sums of the series on the right-hand side converge pointwise to the exponential function.

8.2 Limits and integrals: preliminary remarks. Consider the following sequences $\{f_n\}$ on $[0,1]$; you are encouraged to sketch the graphs of the first few terms in each sequence.

(1) Let $f_n(x) = (n+x)^{-2}$. For each fixed x, $\lim_{n\to\infty} f_n(x) = 0$ and, by the FTC,

$$\lim_{n\to\infty} \int f_n = \lim_{n\to\infty} \left[-(n+x)^{-1}\right]_0^1 = 0 \quad \text{and} \quad \int \lim_{n\to\infty} f_n = 0.$$

(2) Let $f_n(x) = x^n$. Then $f_n(x) \to 0$ as $n \to \infty$ for each $x \neq 1$ and $f_n(1) = 1 \to 1$. Hence $\lim f_n = \chi_{\{1\}}$ and

$$\lim_{n\to\infty} \int f_n = \lim_{n\to\infty} \left[\frac{x^{n+1}}{n+1}\right]_0^1 = 0 \quad \text{and} \quad \int \lim_{n\to\infty} f_n = 0.$$

(3) Let $f_n(x) = n^{-1}\chi_{((1/n-1/n!),1/n)}(x)$. If $x = 0$ then $f_n(x) = 0$ for all n. If $0 < x \leqslant 1$ then $f_n(x) = 0$ for $n > 1/x$. Thus $\lim f_n = 0$ and

$$\lim_{n\to\infty} \int f_n = \lim_{n\to\infty} \frac{1}{n.n!} = 0 \quad \text{and} \quad \int_{n\to\infty} \lim f_n = 0.$$

(4) Let $f_n = n^2\chi_{(0,1/n)}$. As in (3) we have $\lim f_n = 0$. This time

$$\lim_{n\to\infty} \int f_n = \lim_{n\to\infty} n = \infty \quad \text{and} \quad \int \lim_{n\to\infty} f_n = 0.$$

8.3 Exercise example. Consider the following functions f_n:

(i) $(1+n^2x^2)^{-1}$, (ii) $x(1+nx)^{-1}$, (iii) $\sqrt{n}\chi_{(0,1/n)}$, (iv) $(-1)^n n\chi_{(0,1/n]}$.

In each case calculate the pointwise limit f of $\{f_n\}$, sketch the graphs of the first few terms of the sequence, and evaluate $\lim \int_0^1 f_n$ and $\int_0^1 f$ ($:= \int_0^1 \lim f_n$).

8.4 Limits and integrals: further remarks. We are forced by 8.2(4) (and 8.3) to acknowledge that in general

Z

$$\lim \int f_n \neq \int \lim f_n.$$

This is a pity. However, encouraged by examples (1)–(3) in 8.2, we can ask under what circumstances it is true that $\lim \int f_n = \int \lim f_n$.

Let $\{f_n\}$ be a sequence of integrable functions on a compact interval $[a,b]$ converging pointwise to an integrable function f. (At this stage, this means that the functions must be step functions or be continuous.) Then

$$\begin{aligned} \left|\int_a^b f_n - \int_a^b f\right| &= \left|\int_a^b (f_n - f)\right| && \text{(by property (L))} \\ &\leqslant \int_a^b |f_n - f| && \text{(by property (M))} \\ &\leqslant (b-a)\sup|f_n - f| && \text{(by 5.2)} \end{aligned}$$

and this tends to 0 as $n \to \infty$ provided

$$(*) \qquad \alpha_n = \sup|f_n - f| := \sup\{\,|f_n(x) - f(x)| \mid x \in [a,b]\,\} \to 0 \text{ as } n \to \infty.$$

Thus $(*)$ is a sufficient condition for $\lim \int_a^b f_n = \int_a^b \lim f_n$. Let's review the examples in 8.2.

In (1), $\alpha_n = \sup\{\,(n+x)^{-2} \mid x \in [0,1]\,\} = n^{-2} \to 0,$	so $(*)$ holds.
In (2), $\alpha_n = \sup\{\,x^n \mid x \in [0,1)\,\} = 1$ for all n,	so $(*)$ fails.
In (3), $\alpha_n = \sup\{\,n^{-1} \mid x \in ((1/n - 1/n!), 1/n)\,\} = n^{-1} \to 0,$	so $(*)$ holds.
In (4), $\alpha_n = \sup\{\,n^2 \mid x \in (0, 1/n]\,\} = n^2 \to \infty,$	so $(*)$ fails.

This agrees with our general argument above. Example (2) shows that $(*)$ is not necessary for $\lim \int_a^b f_n = \int_a^b \lim f_n$ to occur. Nevertheless, $(*)$ is worth pursuing.

8.5 Uniform convergence of sequences. Suppose that $\{f_n\}$ is a sequence of real-valued [or complex-valued] functions defined on some set S. Then we say $\{f_n\}$ *converges uniformly* to f on S (and write $f_n \xrightarrow{\text{u}} f$ on S) if

$$\alpha_n := \sup\{\,|f_n(x) - f(x)| \mid x \in S\,\} \to 0 \text{ as } n \to \infty.$$

It is important to appreciate the difference between pointwise and uniform convergence. To understand it, we have to go into ε-ology:
$f_n \to f$ pointwise on S if

$$(\forall x \in S)\ (\forall \varepsilon > 0)\ (\exists N_\varepsilon(x) \in \mathbb{N})\ (n \geqslant N_\varepsilon(x) \implies |f_n(x) - f(x)| < \varepsilon).$$

while $f_n \to f$ uniformly on S if

$$(\forall \varepsilon > 0)(\exists N_\varepsilon \in \mathbb{N})\ (n \geqslant N_\varepsilon \implies \sup\{\,|f_n(x) - f(x)| \mid x \in S\,\} < \varepsilon)$$

or, equivalently,

$$(\forall \varepsilon > 0)(\exists N_\varepsilon \in \mathbb{N})\ (n \geqslant N_\varepsilon \implies (\forall x \in S)\ |f_n(x) - f(x)| < \varepsilon).$$

Obviously $f_n \xrightarrow{\text{u}} f$ on S implies $f_n \to f$ on S. The key difference between the two modes of convergence is this: in uniform convergence there is a single N_ε which serves as $N_\varepsilon(x)$ for all x—the sequence converges at a 'uniform rate' over S. [There may appear to be a sleight of hand in the claim of equivalence above: the supremum of a set of numbers all $< \varepsilon$ may equal ε. This is in fact immaterial, since in limit definitions we may use strict or non-strict inequalities interchangeably.]

8.6 Examples. Our earlier calculations have shown that, of the sequences in 8.2, those in (1) and (3) converge uniformly on $[0,1]$, while those in (2) and (4) do not. We now consider some further examples.

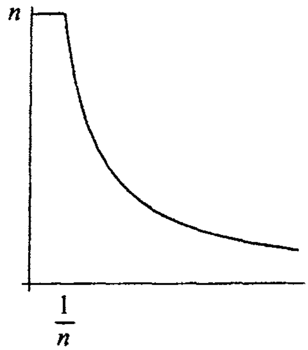


Figure 8.1

(1) On $[0,1]$, let $f_n(x) = \min\{1/x, n\}$ for $x \neq 0$, $f_n(0) = 0$, as shown in Fig. 8.1. Here we have $f_n(x) = 1/x$ if $n > 1/x$ $(x \neq 0)$. So $f_n \to f$, where $f(x) = 1/x$ for $x \neq 0$ and $f(0) = 0$. Hence

$$\alpha_n = \sup_{x \in (0,1/n)} |n - 1/x| = \infty.$$

Therefore $\{f_n\}$ does not converge uniformly on $[0,1]$.

(2) Consider $f_n(x) := nx^3 \mathrm{e}^{-nx^2}$ on $[0,1]$. For $x = 0$, $f_n(x) = 0$ for all n. For $x \neq 0$ we have, looking ahead to 8.16,

$$0 \leqslant f_n(x) = nx^3/(1 + nx^2 + \frac{(nx^2)^2}{2!} + \dots) \leqslant \frac{2}{nx} \to 0 \text{ as } n \to \infty.$$

We now fix n and compute $\alpha_n := \sup\{\, nx^3 \mathrm{e}^{-nx^2} \mid x \in [0,1] \,\}$. The continuous function f_n attains a maximum value on $[0,1]$, which we can find by differentiation. We have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(nx^3\mathrm{e}^{-nx^2}\right) = 3nx^2\mathrm{e}^{-nx^2} - 2n^2x^4\mathrm{e}^{-nx^2}$$

and this is zero when $x = 0$ (giving a minimum) and when $2nx^2 = 3$ (giving a maximum). Hence

$$\alpha_n = nx^3\left[\mathrm{e}^{-nx^2}\right]_{x=\sqrt{3/2n}} = C\sqrt{3/2n},$$

where C is a constant independent of n, and so $\alpha_n \to 0$ as $n \to \infty$. Therefore $f_n \xrightarrow{\mathrm{u}} 0$ on $[0,1]$. See Fig. 8.2.

(3) Let $f_n(x) = n^{-2}(n - x)\chi_{[0,n]}(x)$ on $\mathbb{R}$. Clearly $\lim f_n = 0$ and $\alpha_n = n^{-1}$ when $\alpha_n := \sup\{\, |f_n(x)| \mid x \in \mathbb{R} \,\}$. Hence $f_n \xrightarrow{\mathrm{u}} 0$ on $\mathbb{R}$. But

$$\int f_n = \int_0^n n^{-2}(n - x)\,\mathrm{d}x = \left[n^{-2}(nx - \frac{1}{2}x^2)\right]_0^n \to \frac{1}{2}.$$

Moral: uniform convergence of $\{f_n\}$ on the unbounded set $\mathbb{R}$ is not sufficient to guarantee $\lim \int f_n = \int \lim f_n$.

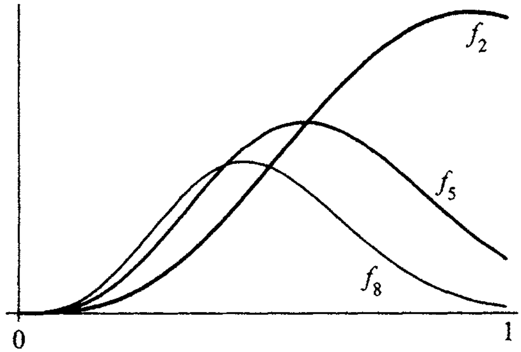


Figure 8.2

8.7 Exercise example. Which of the sequences in 8.3 converge uniformly on $[0,1]$?

8.8 Limits and continuity. Look again at Example (2) in 8.2, in which $f_n(x) = x^n$ on $[0,1]$. Here we have another example of non-commuting limits:

$$\lim_{x\to 1}\left\{\lim_{n\to\infty} f_n(x)\right\} = 0 \quad \text{and} \quad \lim_{n\to\infty}\left\{\lim_{x\to 1} f_n(x)\right\} = 1.$$

Each function f_n is continuous, yet the limit function $\lim f_n = \chi_{\{1\}}$ is not continuous. It is no coincidence that this arose for a sequence $\{f_n\}$ which is not uniformly convergent.

Let $\{f_n\}$ be a sequence of continuous functions on S and assume $f_n \xrightarrow{u} f$ on S. We claim that f is continuous. To prove this we fix the obligatory $\varepsilon > 0$ and let $x \in S$. By uniform convergence we can find $N \in \mathbb{N}$ such that

$$n \geqslant N \implies (\forall y \in S)\ |f_n(y) - f(y)| < \varepsilon.$$

By continuity of f_N at x there exists $\delta > 0$ such that

$$|x - y| < \delta \implies |f_N(x) - f_N(y)| < \varepsilon$$

(δ depending on N—but N is fixed). Hence for $|x - y| < \delta$,

$$\begin{aligned}|f(x) - f(y)| &= |f(x) - f_N(x) + f_N(x) - f_N(y) + f_N(y) - f(y)| \\ &\leqslant |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| < 3\varepsilon.\end{aligned}$$

This suffices to prove our claim.

Observe that when discussing limits and integrals in 8.4 we assumed the limit, f, of the sequence $\{f_n\}$ to be integrable on $[a,b]$. We now see that this is automatic when the functions f_n are continuous and convergence is uniform.

We can sum up as follows.

8.9 Proposition. Let $\{f_n\}$ be a sequence of continuous functions converging uniformly to f on $[a,b]$. Then $f \in \mathrm{C}[a,b]$ and

$$\lim \int_a^b f_n = \int_a^b \lim f_n = \int_a^b f.$$

8.10 Uniform convergence of series. As usual, we handle a series by considering its sequence of partial sums. Accordingly, given a sequence of real-valued functions $\{u_k\}$ on a set S we say that the series $\sum u_k$ *converges pointwise (uniformly) on* S if $\{f_n\}$ converges pointwise (uniformly) on S, where $f_n := u_1 + \cdots + u_n$.

We derive as a corollary of Proposition 8.9 the following: if the functions $u_k \in \mathrm{C}[a,b]$ are continuous and the series $\sum u_k$ converges uniformly on $[a,b]$, then its sum is continuous and

$$\int_a^b \sum_{k=1}^{\infty} u_k = \sum_{k=1}^{\infty} \int_a^b u_k.$$

To verify this, note that $f_n := u_1 + \cdots + u_n$ is continuous for each n and that

$$\int_a^b \sum_{k=1}^{\infty} u_k = \int_a^b \lim_{n\to\infty} f_n = \lim_{n\to\infty} \int_a^b f_n = \sum_{k=1}^{\infty} \int_a^b u_k.$$

Thus we can legitimately integrate a series of continuous functions 'term-by-term' over $[a,b]$ provided it is uniformly convergent there. Term-by-term integration of series is not universally valid. A **finite** sum and an integral can always be interchanged, by linearity. However an **infinite** sum involves a limiting process, and limits and integrals do not always commute, as we illustrated earlier.

There is a user-friendly sufficient condition for uniform convergence of a series. It is not a necessary condition.

8.11 Weierstrass' M-test. The series $\sum u_k$ converges uniformly on S if there exist real numbers M_k such that

$$(\forall k)\ |u_k(x)| \leqslant M_k \text{ for all } x \in S \quad \text{and} \quad \sum M_k \text{ converges.}$$

Proof. We invoke the Cauchy Convergence Principle stating that a sequence $\{x_n\}$ of real numbers converges if and only if it satisfies the Cauchy condition:

$$(\forall \varepsilon > 0)(\exists N \in \mathbb{N})\ (m,n \geqslant N \implies |x_m - x_n| < \varepsilon).$$

Let $f_n := u_1 + \cdots + u_n$. For each $x \in S$ and $n > m$,

$$|f_m(x) - f_n(x)| = |u_{m+1}(x) + \cdots + u_n(x)| \leqslant M_{m+1} + \cdots + M_n \to 0 \text{ as } m,n \to \infty$$

by the Cauchy condition applied to the partial sums of the series $\sum M_n$. Hence $\{f_n(x)\}$ satisfies the Cauchy condition and so converges, to $f(x)$, say. Thus the series $\sum u_n$ converges pointwise. To check that convergence is uniform, take the limit as $m \to \infty$ in the displayed line (with x fixed) to get

$$(\forall x \in S)\ |f(x) - f_n(x)| \leqslant \sum_{k=n+1}^{\infty} M_k \to 0 \text{ as } n \to \infty. \qquad \square$$

8.12 Examples.

(1) Let $u_k(x) = \dfrac{x^p}{1+k^2x^2}$ (p constant, $p \geqslant 2$). For $x \in [0,1]$,

$$|u_k(x)| \leqslant \frac{x^{p-2}}{k^2} \leqslant \frac{1}{k^2}.$$

Since $\sum k^{-2}$ converges, $\sum u_k(x)$ converges uniformly on $[0,1]$. [Actually convergence is uniform for any $p > 1$ but this is harder to prove.]

(2) Consider $\sum_{k=0}^{\infty}(-1)^k x^{2k+1}/(2k+1)!$ on $[-M, M]$. On this interval

$$\left|(-1)^k \frac{x^{2k+1}k}{(2k+1)!}\right| \leqslant M_k := \frac{M^{2k+1}}{(2k+1)!}$$

and $\sum M_k$ converges. So the M-test applies and we conclude that the given series converges uniformly on any bounded interval.

(3) Consider $\sum_{k=0}^{\infty} x^k$. This converges if and only if $|x| < 1$. We show it is not uniformly convergent on $(-1,1)$ by looking at partial sums. We have

$$f_n(x) := \sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x} \to f(x) := \frac{1}{1-x} \text{ for } |x| < 1.$$

Hence

$$\alpha_n = \sup_{|x|<1} |f_n(x) - f(x)| = \sup_{|x|<1} \left|(1-x)^{-1}x^{n+1}\right| = \infty.$$

8.13 Power series. A power series is a function defined as a sum $\sum_{k=0}^{\infty} a_k x^k$ where the constants a_k are real or complex numbers. For simplicity we shall work here with the real case. Power series are important because they define, or can be used to represent, the fundamental functions of mathematics: exponential, logarithm, and trigonometric functions, for example. We get a power series expansion of any infinitely differentiable function which has a Taylor expansion with a remainder term which tends to zero (recall 7.5), and power series solutions exist for differential equations of suitable type.

We assume some familiarity with power series. These are developed, for example, in [1]. By definition $\sum_{k=0}^{\infty} a_k x^k$ has *radius of convergence*

$$R := \sup\{\, |x| \mid \sum \left|a_k x^k\right| \text{ converges}\,\}.$$

Here R is a non-negative real number or ∞. It can be computed by using any of the well-known tests for convergence of series of positive terms, with d'Alembert's ratio test or Cauchy's nth root test handling the majority of series arising in applications. The most familiar power series and their radii of convergence are given in Appendix I.

A power series $\sum_{k=0}^{\infty} a_k x^k$ with radius of convergence R has the following properties:

(R1) $\sum |a_k x^k|$ converges for $|x| < R$ and diverges for $|x| > R$;

(R2) $\sum a_k x^k$ converges for $|x| < R$ and diverges for $|x| > R$.

Further, the following properties hold.

(R3) $\sum a_k x^k$ converges uniformly on $[-R+\delta, R-\delta]$ for any $\delta > 0$ (or on any bounded interval if $R = \infty$); in general $\sum a_k x^k$ does **not** converge uniformly on $(-R, R)$.

(R4) $f(x) := \sum_{k=0}^{\infty} a_k x^k$ defines a continuous function f on $(-R, R)$.

The first part of (R3) is immediate from (R1) and Weierstrass' M-test applied with $M_k := |a_k(R-\delta)^k|$. For the second, consider $\sum_{k=0}^{\infty} x^k$ (see 8.12(2)).

To prove (R4), we need to prove that f is continuous at each point $x \in (-R, R)$. Let $|x| < R$ and choose $\delta > 0$ such that $|x| < R - \delta$. Then by (R3) and 8.8 f is continuous on $[-R+\delta, R-\delta]$ and in particular at x.

2 The subtlety in property (R3)—a nasty trap for the unwary—begins to expose the limitations of uniform convergence as a tool. Fortunately the Lebesgue theory is good at accommodating bad behaviour such as the divergence of a power series integrand at one or more of the endpoints $\pm R$ of the interval of convergence $(-R, R)$ (cf. 4.6, 12.9).

8.14 Integration of power series. Let $\sum_{k=0}^{\infty} a_k x^k$ be a power series with radius of convergence $R > 0$. Then, for any x with $|x| < R$,

$$\int_0^x \sum_{k=0}^{\infty} a_k t^k \, dt = \sum_{k=0}^{\infty} \int_0^x a_k t^k \, dt = \sum_{k=0}^{\infty} \frac{1}{k+1} a_k x^{k+1}.$$

Proof. Fix x with $|x| < R$. Assume without loss of generality that $x > 0$. By (R3) in 8.13, $\sum_{k=0}^{\infty} a_k t^k$ converges uniformly (with respect to the variable t) on $[0, x]$. The result follows from 8.9. ☐

(Note the way we work with definite integrals; recall the comments in 5.6.)

8.15 Differentiation of power series. Let $f(x) := \sum_{k=0}^{\infty} a_k x^k$ be a power series with radius of convergence $R > 0$. We have already recorded the fact that f is continuous on $(-R, R)$. It is well known that we can do better: f is differentiable, with its derivative obtained by differentiating 'term-by-term'. This is often proved by a rather messy and technical argument (see, for example, [**13**], 2.12). The result can be obtained more neatly as a spin-off from 8.14.

We wish to integrate the differentiated series $\sum_{k=1}^{\infty} ka_k x^{k-1}$. Assume *pro tem* that this converges for $|x| < R$. By 8.14 applied to $\sum_{k=0}^{\infty} ka_k x^{k-1}$ (with the $k = 0$ term interpreted as 0),

$$\int_0^x \sum_{k=0}^{\infty} ka_k t^{k-1}\,\mathrm{d}t = \sum_{k=0}^{\infty} a_k x^k.$$

By the Indefinite Integral Theorem, the left-hand side is differentiable at x, with derivative $\sum_{k=1}^{\infty} ra_k x^{k-1}$. Therefore the right-hand side is differentiable and

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^{\infty} a_k x^k = \sum_{k=1}^{\infty} ka_k x^{k-1} = \sum_{k=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}x} a_k x^k \qquad (|x| < R).$$

We must now check that $\sum_{k=1}^{\infty} ka_k x^{k-1}$ converges for $|x| < R$. Some technical juggling is needed here. Fix y with $|x| < |y| < R$, so $\rho := |x/y| < 1$. Then

$$|ka_k x^{k-1}| = \frac{k}{|y|}\left|\left(\frac{x}{y}\right)^{k-1} a_k y^k\right| \leqslant c_k := \frac{k}{|y|}\rho^{k-1},$$

if k is sufficiently large (because $\sum a_k y^k$ converges, $a_k y^k \to 0$). The series $\sum c_k$ converges, whence by comparison $\sum ka_k x^{k-1}$ does too. [In fact, $\sum a_k x^k$, $\sum ka_k x^{k-1}$, and $\sum a_k x^{k+1}/(k+1)$ all have the same radius of convergence.]

8.16 The exponential function. The exponential function is given by

$$\mathrm{e}^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \qquad (x \in \mathbb{R}).$$

This may be taken as the definition of e^x, or as the Maclaurin expansion of the function f satisfying $f'(x) = f(x)$ on $\mathbb{R}$. We can derive the following familiar properties.

(a) $(\mathrm{d}/\mathrm{d}x)\mathrm{e}^x = \mathrm{e}^x$ for all $x \in \mathbb{R}$.

(b) $\mathrm{e}^{x+y} = \mathrm{e}^x\mathrm{e}^y$ for all $x, y \in \mathbb{R}$ and $\mathrm{e}^{-x} = 1/\mathrm{e}^x$ for all x.

(c) $x^k\mathrm{e}^{-x} \to 0$ as $x \to \infty$ for any $k \in \mathbb{R}$.

Proof. The power series has $R = \infty$. Part (a) is immediate from 8.15. The second part of (b) follows from the first because $\mathrm{e}^0 = 1$. To prove the first part we differentiate

$$f(u) := \mathrm{e}^u\mathrm{e}^{c-u}$$

with respect to u (with c fixed). We get $f'(u) = 0$, whence f is a constant depending on c. Putting $u = c$ we see this constant is e^c. Now take $c = x + y$ and put $u = x$.

For (c) observe that for $x > 0$ and $m > k$, $m \in \mathbb{N}$,

$$0 \leqslant x^k\mathrm{e}^{-x} = \frac{x^k}{\mathrm{e}^x} = \frac{x^k}{1 + x + \cdots + x^m/m! + \ldots} \leqslant m!x^{k-m}.$$

8.17 Exponentials and logarithms. We defined the logarithm in 5.12 by

$$\log x := \int_1^x \frac{1}{t}\,\mathrm{d}t \quad (x > 0).$$

We can now confirm that log so defined acts as the inverse to the exponential function and thence examine its rate of growth.

(a) $\log(\mathrm{e}^x) = x$ $(x \in \mathbb{R})$ and $\mathrm{e}^{\log x} = x$ $(x > 0)$.

(b) As $x \to \infty$, $x^{-k}\log x \to 0$ for any $k > 0$.

(c) As $x \to 0+$, $x^k \log x \to 0$ for any $k > 0$.

Proof. By the chain rule for derivatives and 8.16 $(\mathrm{d}/\mathrm{d}x)(\log \mathrm{e}^x) = \mathrm{e}^x/\mathrm{e}^x = 1$, whence $\log(\mathrm{e}^x) = x + C$, where the constant C is seen to be 0 by putting $x = 0$. This proves the first assertion in (a) and the other is proved similarly.

For (b), write $y = \log x$ so that $x = \mathrm{e}^y$. Then $x^{-k}\log x = y\mathrm{e}^{-ky} \to 0$ as $y \to \infty$ (by 8.16) and hence as $x \to \infty$ (by 5.12(e)). Part (c) then follows from 5.12(d). □

Exercises

8.1 Consider the following functions f_n:

$$\text{(i)}\ \frac{nx}{1+n^2x^2}, \qquad \text{(ii)}\ nx^n(1-x), \qquad \text{(iii)}\ n^2x^n(1-x), \qquad \text{(iv)}\ \frac{x}{1+nx^2}.$$

In which cases does $\{f_n\}$ converge uniformly on $[0,1]$? In which cases is it true that $\lim \int f_n = \int \lim f_n$?

8.2 Prove that

$$\int_0^1 \frac{(1-x^2)^2}{1-x}\,\mathrm{d}x = \sum_{k=1}^{\infty} \frac{8}{k(k+2)(k+4)}$$

and hence evaluate the sum.

8.3 (a) Prove that $\sum u_k$ converges uniformly on $[0,1]$ when

$$u_k(x) := (1-x^2)^k x^3.$$

(b) Prove that $\sum u_k$ does not converge uniformly on $[0,1]$ when

$$u_k(x) := x(1-x)^k.$$

(c) For the series in (a) and (b) calculate $\sum_{k=0}^{\infty}\int_0^1 u_k$ and $\int_0^1 \sum_{k=0}^{\infty} u_k$. Comment on your answers.

8.4 Prove that the following series converge uniformly:

$$\text{(i)}\ \sum \frac{\sin kx}{k^2} \text{ on } \mathbb{R}, \quad \text{(ii)} \sum (x\log)^k \text{ on } (0,1], \quad \text{(iii)}\ \sum \frac{x}{(1+kx)^k} \text{ on } \mathbb{R}^+.$$

8.5 Let $\{f_n\}$ be a sequence of functions in $\mathrm{C}^1[a,b]$, assume that $\{f'_n\}$ converges uniformly on $[a,b]$, to a function g and that $\{f_n(\alpha)\}$ converges for some $\alpha \in [a,b]$. Prove that there exists $f \in \mathrm{C}[a,b]$ such that $f_n \xrightarrow{\mathrm{u}} f$ on $[a,b]$ and $f' = g$. [Hint: consider $F_n(x) := \int_\alpha^x f_n$; cf. 8.15.]

8.6 Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a continuous function such that $|f(x,y)| \leqslant 1$ for all x, y and suppose that, for some constant c and for all y, z,

$$|f(x,y) - f(x,z)| \leqslant |y - z| \quad (0 \leqslant x \leqslant c).$$

For $n = 1, 2, \ldots$, define $g_n\colon [0,c] \to \mathbb{R}$ by

$$g_1(x) = 0, \quad g_{n+1}(x) = \int_0^x f(t, g_n(t))\,\mathrm{d}t \quad (n \geqslant 1).$$

Assuming that $\{g_n\}$ converges uniformly on $[0,c]$ to a function g, prove that g satisfies the differential equation $g'(x) = f(x, g(x))$ subject to the initial condition $g(0) = 0$.

9 Building foundations

We have developed an integral on $\mathrm{L}^{\mathrm{step}}$ and from it an integral defined for any continuous function on a compact interval. Now we examine an alternative way to define the integral of $f \in \mathrm{C}[a, b]$. This revised approach, via increasing sequences of step functions, extends to a much wider class of functions. We want to bring in some powerful machinery and it is necessary first to construct the foundations to support it. Thus this chapter is in two parts. The first sections discuss the integration of continuous functions on compact intervals and the remainder present some highly technical results on sequences of step functions which underpin the later theory. We recommend that the proofs of these results be omitted on a first reading.

Our definition of the integral of $f \in \mathrm{C}[a, b]$ as

$$\sup\{ \int \varphi \mid \varphi \in \mathrm{L}^{\mathrm{step}}[a, b],\ \varphi \leqslant f \}$$

has led us successfully to the Fundamental Theorem of Calculus for continuous functions which we can recognize as derivatives. However not every such function on $[a, b]$ can be seen to be of this form, and the defining formula for $\int f$ is clearly impractical for evaluating the integral in general. We should prefer—and in many ways this would be more natural—to define $\int f$ as $\lim \int \varphi_n$, where $\{\varphi_n\}$ is a sequence of step functions approximating f from below. For this we elect to use monotonic approximations. This allows us to take advantage of the Monotonic Sequence Theorem.

9.1 Constructing step function approximations: preamble. As a simple example, consider $f(x) = x$ on $[0, 1]$. To approximate f by a step function it is natural to subdivide $[0, 1]$ into N subintervals $I_1, \ldots, I_N$ each of length $1/N$ and to take the Nth approximation ψ_N to be of the form $\sum_{r=1}^{N} c_r \chi_{I_r}$. To ensure that ψ_N approximates f from below we choose $c_r := \inf\{ f(x) \mid x \in I_r \}$. The approximations so obtained for $N = 2$ and $N = 3$ are shown in Fig. 9.1(a). It is not true that $\psi_2 \leqslant \psi_3$. Consider instead ψ_2, ψ_4, and ψ_8, as shown in Fig. 9.1(b): we do have $\psi_2 \leqslant \psi_4 \leqslant \psi_8$. This suggests that at stage n we should subdivide $[0, 1]$ into $N = 2^n$ subintervals $I_1, \ldots, I_N$ each of length $1/2^n$ and define

$$\varphi_n = \sum_{r=1}^{2^n} c_r \chi_{I_r} \quad \text{where } c_r = \inf\{ f(x) \mid x \in I_r \}.$$

At each point $x \in [0, 1]$ we have (see the figure)

$$0 \leqslant f(x) - \varphi_n(x) \leqslant 2^{-n} \to 0 \quad \text{as } n \to \infty.$$

Explicitly, we take

$$\varphi_n(x) = \sum_{r=1}^{2^n} \frac{(r-1)}{2^n} \chi_{((r-1)/2^n, r/2^n]}(x) \quad \text{for } x \in [0,1].$$

We have

$$\int \varphi_n = \sum_{r=1}^{2^n} \frac{(r-1)}{2^n} \cdot \frac{1}{2^n} = \frac{2^n(2^n-1)}{2^{2n+1}} \to \frac{1}{2}.$$

(We have used the formula $\sum_{k=1}^{m} k = m(m+1)/2$.) This calculation is messy but reassuring. It tells us that $\lim \int \varphi_n$ is $1/2$, the value of $\int_0^1 x \, dx$ supplied by the FTC.)

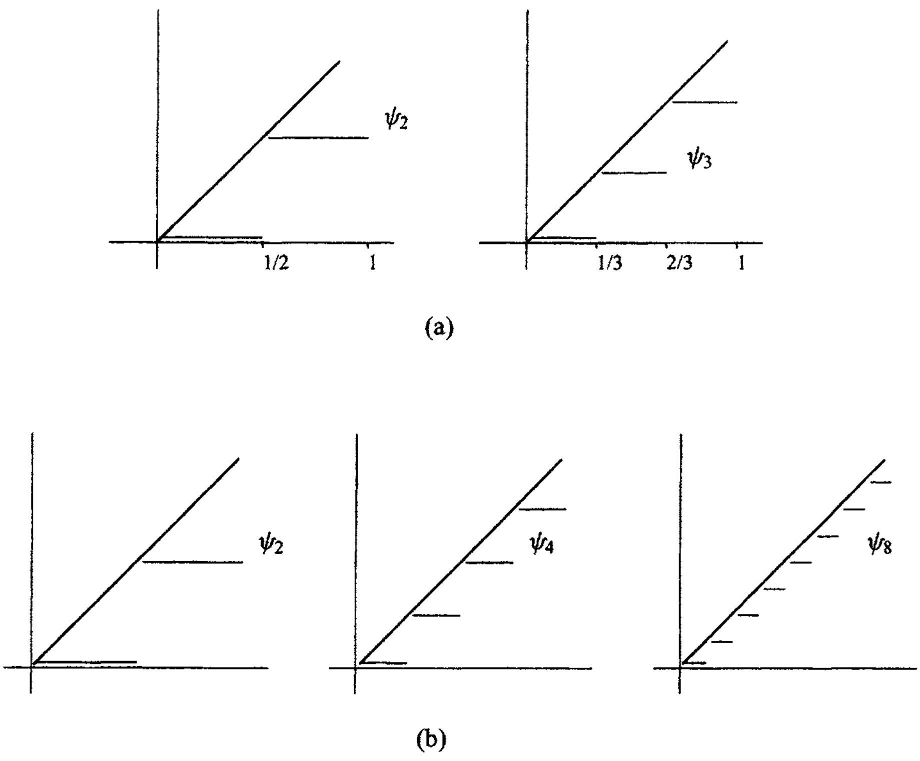


Figure 9.1

9.2 Approximation of continuous functions by step functions. Take $f \in \mathrm{C}[a,b]$. At stage n we partition $[a,b]$ into $N = 2^n$ subintervals each of length $h := (b-a)/N$, *viz.*,

$$I_1 := [a, a+h] \quad \text{and} \quad I_r := (a+(r-1)h, a+rh] \quad (r = 2, \ldots, N).$$

This choice ensures that the intervals slot together without overlapping and that their union is $[a, b]$. We denote the corresponding closed intervals by

$$\overline{I}_r = [a + (r-1)h, a + rh] \quad (r = 1, \ldots, N).$$

On each of these compact intervals $\overline{I}_r$ the continuous function f is bounded below and attains its infimum; let

$$c_r := \min\{ f(x) \mid x \in \overline{I}_r \} \quad (= f(\xi_r), \text{ say}).$$

Now define our approximation at stage n to be

$$\varphi_n := \sum_{r=1}^{2^n} c_r \chi_{I_r}.$$

Then φ_n is a step function and $\varphi_n \leqslant f$ by construction. Further, because of the way we chose N, each interval used at stage $n+1$ is a subset of one used at stage n. This ensures that $\varphi_n(x) \leqslant \varphi_{n+1}(x)$ for each x.

Fix $x \in [a, b]$ and assume that r is such that $x \in I_r$. By continuity of f at x,

$$(\forall \varepsilon > 0)(\exists \delta_x > 0) \ |x - y| < \delta_x \implies |f(x) - f(y)| < \varepsilon.$$

Choose N such that $(b - a)/N < \delta_x$. Then for $n \geqslant N$, $y \in \overline{I}_r$ implies that $|x - y| < \delta_x$. Hence

$$|f(x) - \varphi_n(x)| = |f(x) - f(\xi_r)| < \varepsilon.$$

We have proved that $\varphi_n(x) \to f(x)$ as $n \to \infty$.

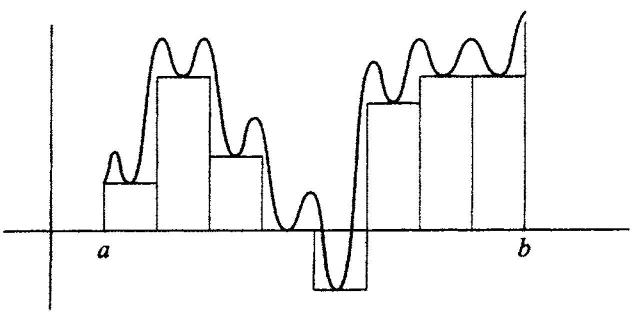


Figure 9.2

9.3 Exercise example. Given $f \in \mathrm{C}[a, b]$, let $\{\varphi_n\}$ be as defined in 9.2. For the following functions sketch the graphs of f, φ_1, φ_2, and φ_3:

(i) x^3 on $[-1, 1]$, (ii) $x\mathrm{e}^{-x}$ on $[0, 4]$, (iii) $\sin x$ on $[0, 2\pi]$.

9.4 Exercise example (for masochists). Carry out the step function approximation process in 9.2 on

$$\text{(i) } e^x \text{ on } [1,2], \qquad \text{(ii) } x^2 \text{ on } [-1,2].$$

In each case calculate $\lim \int \varphi_n$ for the sequence $\{\varphi_n\}$ of step functions you construct.
(The purpose of this exercise is to convince you that
- (i) the answers are those you expect, and
- (ii) the method is **not** to be recommended as a practical means of integration.)

We should like to prove that $\int_a^b f = \lim \int \varphi_n$ for any $f \in C[a,b]$, where $\{\varphi_n\}$ is the sequence of approximating step functions constructed as in 9.2. This is indeed true, but before we can confirm it we need a consequence of the Heine–Borel Theorem which will allow us to show that $\varphi_n \xrightarrow{u} f$.

9.5 Continuity vs. uniform continuity. Let $f \in C[a,b]$. Continuity of f means that

$$(\forall \varepsilon > 0)(\forall x \in [a,b])(\exists \delta_x > 0)\ y \in [a,b],\ |x-y| < \delta_x \implies |f(x) - f(y)| < \varepsilon.$$

Here δ_x in general depends on x. We say that f is *uniformly continuous* on $[a,b]$ if

$$(\forall \varepsilon > 0)(\exists \delta > 0)\ x, y \in [a,b],\ |x-y| < \delta \implies |f(x) - f(y)| < \varepsilon.$$

This time we require $\delta > 0$ to be **independent** of x. The difference between continuity and uniform continuity parallels that between pointwise convergence and uniform convergence discussed in Chapter 8.

We claim that $f \in C[a,b]$ implies that f is uniformly continuous on $[a,b]$. Take δ_x as in the continuity definition and define $I(x) := (x - \delta_x, x + \delta_x)$. The union of the open intervals $\{I(x)\}_{x \in [a,b]}$ contains $[a,b]$. By the Heine–Borel Theorem we can cover $[a,b]$ with **finitely many** of the intervals $I(x)$, say $I(x_1), \ldots, I(x_p)$. Then $\delta := \min\{\delta_{x_1}, \ldots, \delta_{x_p}\} > 0$ and serves as the universal δ demanded in the definition of uniform continuity.

The argument in the previous paragraph relies crucially on $[a,b]$ being compact. Consider $f(x) = x^{-1}$ on $(0,1]$. We can see easily that δ_x is forced to be smaller and smaller as x approaches 0, and that no universal δ can be found.

9.6 $\int_a^b f$ reassessed. Given $f \in C[a,b]$, construct a sequence $\{\varphi_n\}$ in L^{step} in the way described in 9.2. We showed there that
- (i) $\varphi_n \leqslant f$ for all n,
- (ii) $\varphi_n(x) \to f(x)$ for each x, and
- (iii) $\varphi_n \leqslant \varphi_{n+1}$ for all n.

We now claim that we can call on uniform continuity to strengthen (ii) to (ii)$'$ $\varphi_n \to f$ uniformly on $[a,b]$, that is,

$$(\forall \varepsilon > 0)(\exists N \in \mathbb{N})\ (\forall x \in [a,b])\ (n \geqslant N \implies |f(x) - \varphi_n(x)| < \varepsilon).$$

To do this, choose $\delta > 0$ so that $|x-y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$. Now choose N so that the intervals I_r have length $< \delta$. Remembering that φ_n takes value $f(\xi_r)$ on I_r for some $\xi_r \in \overline{I}_r$, we see that (ii)$'$ holds.

We can now prove that $\int_a^b f = \lim \int \varphi_n$, where the integral on the left-hand side is that defined in Chapter 5. Because of (iii) and 4.4,

$$\int \varphi_n \leqslant \int \varphi_{n+1} \quad \text{for all } n.$$

Also, there exists M such that $f(x) \leqslant M$ for all $x \in [a,b]$ by the Boundedness Theorem, 2.23. This implies, by (i), that $\varphi_n(x) \leqslant M$ for all n and x, so that

$$\int \varphi_n \leqslant M(b-a) \quad \text{for all } n.$$

Thus $\{\int \varphi_n\}$ is an increasing sequence of real numbers which is bounded above. Hence $\lim \int \varphi_n$ exists and is finite. We wish to show that

$$\lim \int \varphi_n = \int f := \sup\{ \int \varphi \mid \varphi \in \mathrm{L}^{\mathrm{step}},\ \varphi \leqslant f \}.$$

Certainly $\lim \int \varphi_n \leqslant \int_a^b f$, since each φ_n is a step function below f.

For the reverse inequality we fix a step function $\varphi \leqslant f$ and $\varepsilon > 0$. By (ii)$'$ there exists N such that $\varphi - \varepsilon \leqslant f - \varepsilon < \varphi_n$ for all $n \geqslant N$. Hence $\int \varphi \leqslant \varepsilon(b-a) + \int \varphi_n$ for all n sufficiently large. We deduce that, for any $\varepsilon > 0$,

$$\int \varphi \leqslant \varepsilon(b-a) + \lim \int \varphi_n.$$

Taking the supremum over all φ we see that

$$\int f \leqslant \varepsilon(b-a) + \lim \int \varphi_n.$$

Since ε is arbitrary, we conclude that $\int f \leqslant \lim \int \varphi_n$, as claimed.

9.7 Branching out. A step function has domain $\mathbb{R}$. A function $f \in \mathrm{C}[a,b]$ has domain $[a,b]$. As in 5.6 we extend f to a function $\tilde{f}$ with domain $\mathbb{R}$ by setting $\tilde{f} = f$ on I and $\tilde{f} = 0$ on $\mathbb{R} \setminus [a,b]$. As foreshadowed in 5.6 define $\int \tilde{f} := \int_a^b f$ for $f \in \mathrm{C}[a,b]$. We now define

$$\mathrm{L}^{\mathrm{C}} := \{ \tilde{f} \mid f \in \mathrm{C}[a,b] \text{ and } -\infty < a \leqslant b < \infty \}$$

to be the set of all functions arising in this way.

Therefore the functions on $\mathbb{R}$ which we can currently include in our class of integrable functions is $\mathrm{L}^{\mathrm{step}} \cup \mathrm{L}^{\mathrm{C}}$. Every function in this class is *of compact support* in the sense that it takes value 0 outside some compact interval. We certainly want to integrate functions which are not of compact support: e^{-x^2} for instance.

We have shown above that for any $f \in \mathrm{L^C}$ we can find a sequence $\{\varphi_n\}$ such that

(i) $\varphi_n \in \mathrm{L^{step}}$ for all n,
(ii) $\varphi_n \leqslant \varphi_{n+1}$ for all n,
(iii) $\exists K$ such that $\int \varphi_n \leqslant K$ for all n, where K is a finite constant independent of n,

and $\varphi_n \to f$. Conditions (i)–(iii) ensure that $\lim \int \varphi_n$ exists (and is finite), and 9.6 reconciled this with our earlier definition of the integral of f. With a view to extending our class of integrable functions, let's start at the other end, and consider a sequence $\{\varphi_n\}$ satisfying (i)–(iii). Then we might reasonably **define** $\lim \int \varphi_n$ to be the integral of f where $f(x) := \lim \varphi_n(x)$; the sequence $\{\varphi_n(x)\}$ is monotonic increasing, so the limit exists, provided we allow $f(x) = \infty$ when $\{\varphi_n(x)\}$ is not bounded above. There is no presumption that f must be continuous or that f must be of compact support. However we would not normally allow a function to take the value infinity. Define

$$E := \{\, x \in \mathbb{R} \mid \varphi_n(x) \nearrow \infty \,\}.$$

Suppose E is finite. Then we should not be unduly concerned. We have already seen that we can alter the value of a step function on a finite set without altering its integral (see 4.6). So, redefining each φ_n to take value 0 on E, we get a finite-valued limit function everywhere for $\{\varphi_n\}$, to which we would ascribe the integral $\lim \int \varphi_n$. We are led to ask what we can say about the set E in general. It turns out that, even if E is not finite, it is a set which we can regard as being of 'zero length' and so negligible for integration purposes.

The key to understanding the set E above is a very simple observation. Given an upper bound K for the area, we can construct a rectangle whose height can be made as big as we please, so long as we compensate by making its base length suitably small. See Fig. 9.3. We formalize this in a lemma.

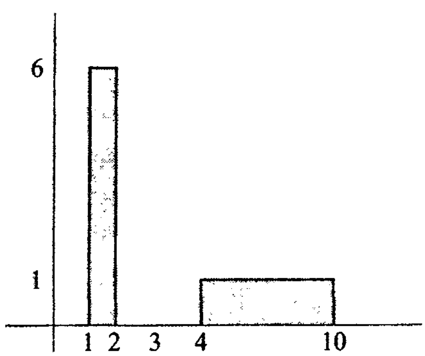


Figure 9.3

9.8 Lemma. Let φ be a step function, expressed as $\sum_{r=1}^{n} c_r \chi_{I_r}$, where the bounded intervals $I_1, \dots, I_n$ are disjoint, and let $\int \varphi \leqslant K$, where $K > 0$. Then for any $m \in \mathbb{N}$, $S^{Km} := \{ x \mid \varphi(x) \geqslant Km \}$ is the disjoint union of finitely many intervals of total length $\leqslant 1/m$.

Proof. Since the intervals are assumed disjoint, $\varphi(x) \geqslant Km$ if and only if $x \in I_r$ for some r for which $c_r \geqslant Km$. Thus S^{Km} is the union of a (possibly empty) subcollection of the intervals $I_1, \dots, I_n$, and

$$K \geqslant \int \varphi = \sum_{r=1}^{n} c_r \ell(I_r) \geqslant Km\,\ell(S^{Km}). \qquad \square$$

We can now present the first of our technical theorems.

9.9 Technical Theorem I. Let $\{\varphi_n\}$ be an increasing sequence of step functions for which there exists K such that $\int \varphi_n \leqslant K$ for all n and let

$$E := \{ x \mid \varphi_n(x) \nearrow \infty \}.$$

Then, given $m \in \mathbb{N}$, there exists a (finite or infinite) sequence $\{J_1, J_2, \dots\}$ of bounded open intervals such that

$$E \subseteq \bigcup_i J_i \quad \text{and} \quad \sum_i \ell(J_i) \leqslant 1/m.$$

[In the terminology introduced below, $\{\varphi_n(x)\}$ converges almost everywhere.]

Proof. Fix m. Let $E_n := \{ x \mid \varphi_n(x) \geqslant 2Km \}$ (the reason for the factor 2 will emerge). By Lemma 9.8, E_n is a union of a finite collection $\mathcal{I}_n$ of disjoint intervals of total length $\leqslant 1/(2m)$. Since $\varphi_n \leqslant \varphi_{n+1}$, we have $E_n \subseteq E_{n+1}$. By refinement if necessary we may assume that $\mathcal{I}_n \subseteq \mathcal{I}_{n+1}$ for all n. We may arrange all the intervals in a single sequence, $\{J_1, J_2, \dots\}$, by first listing those in $\mathcal{T}_1$, then those in $\mathcal{T}_2$, and so on (possible duplication does not matter).

Let $x \in E$. For all n sufficiently large, $\varphi_n(x) \geqslant Km$, so that $x \in E_n$. Therefore $E \subseteq \bigcup_i J_i$. Also, for any N, the family of those J_i with $i \leqslant N$ is contained in some $\mathcal{I}_n$ and hence the sums of the lengths of the intervals is $\leqslant 1/(2m)$. Taking the limit as $N \to \infty$, we get $\sum_i \ell(J_i) \leqslant 1/(2m)$. We have proved the stated result, except in one particular: our intervals J_i need not be open. However this is easily remedied. Any bounded interval I lies inside an open interval I', with $\ell(I') - \ell(I)$ as small as we please. By enlarging the ith interval J_i so that the length increases by at most $2^{-i-1}/m$, we replace the original intervals of total length $\leqslant 1/(2m)$ by a new sequence of open intervals of total length at most

$$1/(2m) + \sum_{i=1}^{\infty} 2^{-i-1}/m = 1/(2m) + 1/(2m) = 1/m.$$

This suffices to complete the proof. $\square$

9.10 Null sets. In the preceding theorem we showed that the 'bad' set E on which $\{\varphi_n(x)\}$ diverges to ∞ can be put inside a union of a sequence $\{J_1, J_2, \dots\}$

of open intervals such that $\sum_i \ell(J_i) \leqslant 1/m$ where m can be made as large as we choose. Assuming we could assign a length, $\ell(E)$, to E we would naturally expect this to have the property that $0 \leqslant \ell(E) \leqslant \sum_i \ell(J_i)$. Thus we would be forced to conclude that $0 \leqslant \ell(E) < 1/m$ for all $m \in \mathbb{N}$, whence $\ell(E) = 0$.

We evade until Chapter 22 the question 'can the length of a general subset of $\mathbb{R}$ be defined?' For the moment, we can get by quite nicely with length only defined for finite unions of disjoint intervals (as in 5.3) and for sets of 'zero length'. We define a set $E \subseteq \mathbb{R}$ to be a *null set* if for any $\varepsilon > 0$ there exists a sequence $\{J_1, J_2, \dots\}$ of open intervals such that

$$E \subseteq \bigcup_i J_i \quad \text{and} \quad \sum_i \ell(J_i) \leqslant \varepsilon.$$

This provides a workable definition of a set of 'zero length' which relies solely on the definition of the length of a bounded interval, and which accords with the way we should like length to behave.

If some property $P(x)$ about real numbers holds for all $x \notin E$, where E is some null set then we say that $P(x)$ holds *almost everywhere* (abbreviated to a.e.) or *for almost all* x. In particular we may write, for example, $\varphi_n \xrightarrow{\text{a.e.}} f$ if there exists a null set E such that $\varphi_n(x) \to f(x)$ for all $x \notin E$.

Theorem 9.9 may be restated: an increasing sequence $\{\varphi_n\}$ of step functions converges a.e. if $\{\int \varphi_n\}$ is bounded (or equivalently convergent). The converse holds too, and adds evidence in favour of studying null sets. We need this result much later on, in Chapter 25.

9.11 Technical Theorem II. Let E be a null set. Then there exists an increasing sequence $\{\varphi_n\}$ of step functions such that $\{\int \varphi_n\}$ is bounded and $\varphi_n(x) \nearrow \infty$ for each $x \in E$.

Proof. As E is null we can find, for each m, a sequence $\{J_{1,m}, J_{2,m}, \dots\}$ of bounded open intervals covering E and such that $\sum_i \ell(J_{i,m}) \leqslant 2^{-m}$. We may arrange the intervals $\{J_{i,m} \mid i, m = 1, 2, \dots\}$ as a single sequence $\{K_1, K_2, \dots\}$. (One way to do this is to put the sets into a square array and count along the diagonals (cf. [**18**]).) The sum of the lengths of the intervals $\{K_i\}$ is at most

$$2^{-1} + 2^{-2} + \cdots + 2^{-m} + \cdots = 1.$$

Now define

$$\varphi_n := \sum_{r=1}^{n} \chi_{K_r}.$$

Certainly $\{\varphi_n\}$ is an increasing sequence of step functions. Also

$$\int \varphi_n = \sum_{r=1}^{n} \ell(K_r) \leqslant 1 \quad \text{for all } n.$$

Hence $\{\int \varphi_n\}$ is increasing and bounded above, and so converges. Finally, fix $x \in E$. We must show $\varphi_n(x) \nearrow \infty$. For each m there is some i_m such that

$x \in J_{i_m,m}$. The interval $J_{i_m,m}$ occurs somewhere in the sequence $\{K_1, K_2, \dots\}$, as K_{k_m} say. For any fixed n, let $N \geqslant k_1 + \cdots + k_n$. Then

$$\varphi_N(x) = \sum_{r=1}^{N} \chi_{K_r}(x) \geqslant \sum_{m=1}^{n} \chi_{K_{k_m}}(x) = n.$$

This proves our claim. $\square$

We shall need one further technical theorem in the same spirit as the two preceding ones. This is given at the end of the next chapter, when we have developed the properties of null sets used in its proof.

Exercises

9.1 Calculate $\int_{-1}^{1} |x| \, dx$ by constructing an approximating sequence of step functions. Could you apply the FTC to evaluate this integral?

9.2 Let $\varphi_n := n^2 \chi_{(0,1/n)}$. Prove that
(i) $\varphi_n(x) \to 0$ for every $x \in \mathbb{R}$, and
(ii) $\int \varphi_n \nearrow \infty$.

9.3 Let $\{\alpha_n\}$ be the sequence

$$1, 2, 1, 2, 3, 1, 2, 3, 4, \dots, (k-1), 1, 2, \dots, k, 1, \dots.$$

Let $\varphi_n := \chi_{(\alpha_n, \alpha_{n+1})}$. Prove that
(i) $\{\varphi_n(x)\}$ does not converge for any x, and
(ii) $\lim \int \varphi_n = 0$.

9.4 Construct a sequence $\{\varphi_n\}$ of step functions such that $\varphi_n(x) \to \infty$ for all $x \in E$ and $\{\int \varphi_n\}$ converges when E is

$$\text{(i) } \mathbb{Z}, \qquad \text{(ii) } \{\, 2^{-k} + 3^{-m} \mid k, m = 1, 2, \dots \}.$$

10 Null sets

In Chapter 9 we introduced null sets as the sets which are negligible in an integration-theoretic sense. On a null set the values of an integrable function f may be altered without changing $\int f$, and on a null set delinquent behaviour is condoned. Certainly any finite set must be null. In general, null sets can be very complicated, reflecting the rich structure of the real line. We do not explore far in this direction, since our emphasis is on functions (modulo their behaviour on negligible sets), rather than on sets *per se*. The harder sections of this chapter and many of the exercises may without detriment be omitted or delayed.

10.1 A refresher on countable sets. We assume that readers are familiar with the notion of countability, at least to the extent of knowing that the real numbers are uncountable and the rationals countable. For more details than we give here, see for example [1].

A set S is countable if its elements can be listed as a sequence, $s_1, s_2, \ldots$, so that the members of S can be labelled by natural numbers. Formally this means that there exists a one-to-one function $\eta\colon s_i \mapsto i$ mapping S into $\mathbb{N}$. According to this definition, every finite set is countable. We call a set countably infinite if it is countable but not finite. Familiar countable sets are $\mathbb{N}$ itself, $\mathbb{Z}$, and $\mathbb{Q}$; on the other hand, $\mathbb{R}$ and any interval $\langle a, b\rangle$ with $a < b$ are uncountable.

The importance for us of countable sets is that they are null; see 10.4. We shall want to combine null sets to create new null sets, and to combine into a single sequence a countable collection of sequences of intervals, as we did in the last chapter. For this the following lemma is very useful. It includes as special cases the statements that the union of two (or any finite number) of countable sets is countable.

10.2 Lemma. Let $\{S_n\}_{n=1,2,\ldots}$ be a sequence of countable sets. Then $S := \bigcup_n S_n$ is countable.

Sketch proof. We may label the elements of S_n as

$$s_{1,n}, s_{2,n}, \ldots, s_{m,n}, \ldots.$$

Define $\eta\colon \bigcup_n S_n \to \mathbb{N}$ by

$$\eta(s_{m,n}) := 2^m 3^n \quad (m, n = 1, 2, \ldots).$$

Because a natural number > 1 can be factorized as a product of prime powers in just one way, $\eta(s_{m.n}) = \eta(s_{p,q})$ only if $m = p$ and $n = q$. Thus η assigns a natural number label to each point of the union S. $\square$

10.3 Null sets (recap). In the preceding chapter we defined a subset E of $\mathbb{R}$ to be a *null set* if, for any given $\varepsilon > 0$, there exists a sequence $\{J_1, J_2, \dots\}$ of bounded open intervals such that

(N1) $E \subseteq \bigcup_i J_i$ and

(N2) $\sum_i \ell(J_i) \leqslant \varepsilon$.

We shall refer to a sequence $\{J_1, J_2, \dots\}$ satisfying (N1) and (N2) as an *ε-cover* for E. Here we allow the possibility that there are only finitely many sets J_i. It is important to stress that the family $\{J_i\}$ is required to be countable.

A technical note: in (N2), we could equivalently have demanded $\sum_i \ell(J_i) < \varepsilon$ or $\sum_i \ell(J_i) \leqslant C\varepsilon$ for a fixed constant C. Such tricks should be familiar from elementary analysis: for example, we may conclude that a sequence $\{\alpha_n\}$ converges to 0 if for each $\varepsilon > 0$ we can find N such that $|\alpha_n| \leqslant 5\varepsilon$ for $n \geqslant N$.

10.4 Lemma.
(a) Any subset of a null set is null.
(b) Any finite set is null.
(c) Any countable set is null.

Proof. (a) is immediate from the definition. Now let $E = \{\alpha_1, \dots, \alpha_m\}$. Fix $\varepsilon > 0$. Define

$$J_i := (\alpha_i - \varepsilon, \alpha_i + \varepsilon) \quad (i = 1, \dots, m).$$

Then $E \subseteq \bigcup_{i=1}^m J_i$ and $\ell(J_i) = 2\varepsilon$ so $\{J_1, \dots, J_m\}$ is a $2m\varepsilon$-cover for E. By our technical note above this suffices to prove (b).

For (c), write $E = \{\alpha_1, \alpha_2, \dots\}$. Again we can satisfy (N1) by assigning each point of E an open interval which covers it, but we can no longer take intervals of equal length. We define

$$J_i := (\alpha_i - 2^{-i-1}\varepsilon,\ \alpha_i + 2^{-i-1}\varepsilon) \quad (i = 1, 2, \dots).$$

Then $\ell(J_i) = 2^{-i}\varepsilon$. Because the geometric series $2^{-1} + 2^{-2} + \dots$ converges to 1, $\{J_1, J_2, \dots\}$ is an ε-cover for E. Thus (c) holds. $\square$

10.5 Examples. The following are null sets:

$$\mathbb{N},\ \ \mathbb{Z},\ \ \mathbb{Q},\ \ \{1/n \mid n \in N\},\ \ \{r/2^n \mid r \in \mathbb{Z},\ n \in \mathbb{N}\}.$$

As we have already indicated, we are going to extend the integral beyond L^{step} and L^{C} using monotonic sequences of step functions which will be allowed to diverge to ∞ on null sets. Indeed, our choice of the letter E to denote a general null set is to reflect the connotation of an 'exceptional set'. Frequently we shall have two different 'bad sets' in play. For example we might have $f = g$ except on one null set E_1 and $\varphi_n \nearrow f$ except on a different null set E_2. Then $\varphi_n \nearrow g$ off $E_1 \cup E_2$. Is this null?

10.6 Lemma.
(a) Let $E_1, E_2, \ldots, E_m$ be null sets. Then $E_1 \cup E_2 \cup \cdots \cup E_m$ is null.
(b) Let $\{E_n\}_{n\in\mathbb{N}}$ be a countable collection of null sets. Then $\bigcup E_n$ is null.

Proof. The lemma specializes to Lemma 10.4 in case each set E_n is a single point, and the proof is similar. We sketch a proof of (b). Fix $\varepsilon > 0$. For each n take a $(2^{-n}\varepsilon)$-cover $J_{n,1}, J_{n,2}, \ldots$ for E_n. Since a countable union of countable sets is countable we may list the sets $\{J_{n,i} \mid n, i \geqslant 1\}$ as a sequence $\{K_1, K_2, \ldots\}$, with $\sum_i \ell(K_i) \leqslant \sum_{n=1}^{\infty} 2^{-n}\varepsilon$. Hence $\{K_1, K_2, \ldots\}$ is an ε-cover for $\bigcup E_n$. □

The first part of the next result is used in the transfer of the translation-invariance property (T) beyond $\mathrm{L}^{\mathrm{step}}$.

10.7 Exercise example. Let E be null. Prove that

$$E + d := \{\, x + d \mid x \in E \,\} \quad \text{and} \quad dE := \{\, dx \mid x \in E \,\}$$

are null for any $d \in \mathbb{R}$.

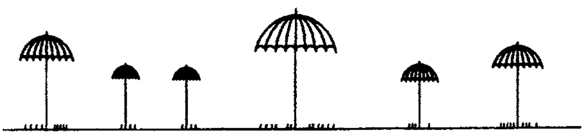


Figure 10.1

10.8 Remarks. It is a common misconception that every null set must be finite or countably infinite. This is not so: we shall give in 10.9 an example of a famous set (the Cantor set) which is both null and uncountable. The misconception arises from a misunderstanding of the null set definition. To explain this further, it will be helpful to think of the intervals in an ε-cover $\{J_1.J_2, \ldots\}$ as a family of umbrellas sheltering the points of E, as in 10.1. Remember that we insist that the members of this family are labelled by the natural numbers (or a subset thereof). In Lemma 10.4 we were able to allow each point of our finite or countable set to carry its own individual umbrella. In general, clusters of points have to share umbrellas.

We now give the promised example to dispel the myth that every null set is countable. The construction is intricate—at least, it seems so on first encounter. We shall not need it later, except peripherally in Chapter 22, where Cantor-like sets with even more exotic properties are paraded.

10.9 The Cantor set: an uncountable null set. Using the notation of 10.7, let

$$C_0 := [0, 1], \qquad C_{n+1} := C_n \cap (1/3C_n \cup (1/3C_n + 2/3)) \quad (n \geqslant 0).$$

The Cantor set is $C := \bigcap_n C_n$. In more detail,

$C_1 = [0, 1/3] \cup [2/3, 3/3]$—2 disjoint closed intervals each of length $1/3$,
$C_2 = [0, 1/3^2] \cup [2/3^2, 3/3^2] \cup [6/3^2, 7/3^2] \cup [8/3^2, 9/3^2]$—$2^2$ disjoint closed intervals each of length $1/3^2$,

and so on. An easy but instructive proof by induction gives, for all n,
(a) $C_{n+1} \subset C_n$;
(b) C_n is the disjoint union of 2^n intervals each of length $1/3^n$;
(c) $\ell(C_n) = (2/3)^n$.

Because $C \subseteq C_n$ for any n, C can be covered by finitely many open intervals of total length $(2/3)^{n-1}$ (just cover each constituent closed interval of C_n by a slightly larger open interval). Since $(2/3)^{n-1} \to 0$ as $n \to \infty$, we conclude that C must be null. The complement $[0,1] \setminus C$ is the union of the sequence $\{O_n\}$ of open intervals

$$\begin{array}{lllll} (1/3, 2/3), & & & & \\ (1/3^2, 2/3^2), & (4/3^2, 5/3^2), & (7/3^2, 8/3^2), & & \\ (1/3^3, 2/3^3), & (4/3^3, 5/3^3), & (7/3^3, 8/3^3), & (10/3^3, 11/3^3), & \ldots . \end{array}$$

We now claim C is uncountable. As is customary with proofs of uncountability, we argue by contradiction. So suppose we could list all the points of C as $x_1, x_2, x_3, \ldots$. To reach a given point in C we may imagine being given signposts from C_0 into C_1, C_1 into C_2, and so on. At stage n the signpost says either L (turn left into $(1/3)C_n$) or R (turn right into $(1/3)C_n + (2/3)$). If we are perverse, and at stage k go left if the signpost to x_k is R, and right if the signpost to x_k is L, then we ultimately reach a point x_0 in C (this is not obvious; see Exercise 10.3), but this point cannot be x_k for any $k \geqslant 1$. This gives the required contradiction. [This, in a slightly informal presentation, is the proof of uncountability of those real numbers expressible in the form $\sum_{k=1}^{\infty} a_k 3^{-k}$ ($a_k = 0$ or 2); compare with the usual proof of uncountability of $[0,1]$ via decimal expansions.]

The next result is intuitively clear. Perhaps surprisingly its proof relies on a special case of a major result about the structure of $\mathbb{R}$—the Heine–Borel Theorem (stated in 2.16(3)).

10.10 Proposition. Let $a < b$. Then the interval $I = \langle a, b \rangle$ is not null.

Proof. Since a subset of a null set is null we may assume without loss of generality that I is bounded. Assume for a contradiction that I is null. Let $\{J_1, J_2, \ldots\}$ be an ε-cover for I. By adding open intervals of length ε centered on the endpoints a and b, we obtain a 3ε-cover, $\{K_1, K_2, \ldots\}$ say, for $[a, b]$. By the Heine–Borel Theorem, we may cover $[a, b]$ with finitely many of the open intervals K_i, say $K_{i_1}, \ldots, K_{i_N}$. Then $I \subseteq K_{i_1} \cup \cdots \cup K_{i_N}$, whence

$$\ell(I) \leqslant \ell(K_{i_1}) + \cdots + \ell(K_{i_N})$$

by 3.3. Therefore $0 < b - a \leqslant 3\varepsilon$. Since this must hold for all $\varepsilon > 0$ we have a contradiction. $\square$

We conclude this chapter with the last of our technical theorems. This one is used in the proof of the consistency of the extension to our integral that we define in the next chapter. Again we need the Heine–Borel Theorem.

10.11 Technical Theorem III. Let $\{\theta_n\}$ be a decreasing sequence of non-negative step functions such that $\theta_n \xrightarrow{\text{a.e.}} 0$. Then $\int \theta_n \to 0$ as $n \to \infty$.

Proof. Throughout the proof a subscript x indicates possible dependence on x. We first assemble some facts and notation about $\{\theta_n\}$. Fix $\varepsilon > 0$.

(a) Let E_0 be a null set such that $\theta_n \to 0$ off E_0. Given $x \notin E_0$, there exists N_x such that $0 \leqslant \theta_{N_x}(x) < \varepsilon$.

(b) For each $n \geqslant 1$, θ_n has at most a finite set E_n of discontinuity points. By 10.4 and 10.6, $E := \bigcup_{n \geqslant 0} E_n$ is null. Let $\{J_1, J_2, \dots\}$ be an ε-cover for E.

(c) There exist a constant M and a compact interval $[a, b]$ such that $\theta_1 \leqslant M\chi_{[a,b]}$.

(d) For $m \geqslant n \geqslant 1$, $0 \leqslant \theta_m \leqslant \theta_n \leqslant M\chi_{[a,b]}$.

Fix $x \notin E$. By arrangement, θ_{N_x} is continuous at x, so we can find an open interval $I(x) := (x - \delta_x, x + \delta_x)$ $(\delta_x > 0)$ such that

$$y \in I(x) \implies \theta_{N_x}(y) = \theta_{N_x}(x)$$

(see Exercise 3.5). Hence, by (c),

$$0 \leqslant \theta_n(y) \leqslant \theta_{N_x}(y) < \varepsilon \quad \text{for all } y \in I(x),\ n \geqslant N_x.$$

Certainly

$$[a, b] \subseteq \bigcup_i J_i \cup \bigcup_{x \notin E} I(x).$$

By the Heine–Borel Theorem, we can cover $[a, b]$ with a finite subcollection of the open intervals in the right-hand union, say $J_{i_1}, \dots, J_{i_p}, I(x_1), \dots, I(x_q)$. Let $J := J_{i_1} \cup \dots \cup J_{i_p}$ and $I := I(x_1) \cup \dots \cup I(x_q)$. Define $N := \max\{N_{x_1}, \dots, N_{x_q}\}$. For every point $y \in [a, b] \setminus J$ and any $n \geqslant N$, $0 \leqslant \varphi_n(y) < \varepsilon$. Hence

$$\int \theta_n = \int \theta_n \chi_{[a,b]} \leqslant \int (\theta_n \chi_I + \theta_n \chi_J) = \int \theta_n \chi_J + \int \varphi_n \chi_I.$$

But $\int \theta_n \chi_J \leqslant \varepsilon M$ and $\int \theta_n \chi_I \leqslant \varepsilon \sum_{s=1}^{q} \ell(I(x_s))$ (using 3.3). Hence for $n \geqslant N$ we have $\int \theta_n \leqslant C\varepsilon$ where C is a finite constant. This tells us that $\int \theta_n \to 0$. □

Exercises

10.1 Let E be a null set. Prove that $\{x \in \mathbb{R} \mid x - q \in E \text{ for some } q \in \mathbb{Q}\}$ is null.

10.2 Let E be a null set.

(a) Prove that $E^2 := \{ x^2 \mid x \in E \}$ is null. [Hint: do it first for the case that E is bounded.]

(b) Prove that $\{ x^n \mid x \in E,\ n = 1, 2, \dots \}$ is null.

10.3 Let $\{K_n\}$ be a sequence of closed subintervals of a compact interval $[a, b]$ which is such that $K_n \supseteq K_{n+1}$ for all n. Prove that $\bigcap_{n=1}^{\infty} K_n \neq \varnothing$. [Hint: argue by contradiction and apply the Heine–Borel Theorem, or use the Bolzano–Weierstrass Theorem.]

10.4 Fill in the details in the proof in 10.9 that the Cantor set is null and uncountable. [You will need the preceding exercise to show x_0 exists.]

10.5 Assume that, for each k, I_k is a subinterval of $[0, 1]$, and E is the set of points $x \in [0, 1]$ such that $x \in I_k$ for infinitely many values of k. By covering E with the intervals $I_N, I_{N+1}, \dots$ show that E is a null set provided $\sum_{k=1}^{\infty} \ell(I_k)$ is finite.

10.6 Let E_α denote the set of points $x \in [0, 1]$ which have the property that, for infinitely many integer values of q, there exist integers p with

$$\left| x - \frac{p}{q} \right| < \frac{1}{q^\alpha}.$$

Use the result of Exercise 10.5 to show that E_α is a null set if $\alpha > 2$.

10.7 Let E be a null set, let $f\colon \mathbb{R} \to \mathbb{R}$ and consider $f(E) := \{ f(x) \mid x \in E \}$.

(a) Let f have a continuous derivative. Prove that $f(E)$ is null [Hint: Exercise 10.2(b) is a special case.]

(b) Let f be differentiable. Prove that $f(E)$ is null [This is harder. Hint: consider the sets of the form

$$\{ x \in \mathbb{R} \mid |x - y| \leqslant m \implies |f(x) - f(y)| \leqslant n \}$$

for $y \in \mathbb{R}$ and $m, n \in \mathbb{N}$.]

(c) Exhibit a null set E and a continuous function f such that $f(E)$ is not null.

10.8 Let U be a union of open intervals. Let $q_1, q_2, \dots$ be an enumeration of $\mathbb{Q}$. For each $x \in U$ let U_x be the (possibly unbounded) open interval (a_x, b_x) where $a_x := \inf\{ a \in \mathbb{R} \mid (a, x) \subseteq U \}$ and $b_x := \sup\{ b \in \mathbb{R} \mid (x, b) \subseteq U \}$.

(a) Prove that $x \in U_x \subseteq U$ and that if $U_x \cap U_y \neq \varnothing$ then $U_x = U_y$. Deduce that U is the disjoint union of a family of open intervals.

(b) Associate to each $x \in U$ the rational q_{n_x} such that $q_{n_x} \in U_x$ but $q_n \notin U_x$ for $n < n_x$ (this is possible because every non-empty open interval contains a rational number). Prove that $U_x = U_y$ if and only if $n_x = n_y$, and deduce that U is the union of a countable collection of disjoint open intervals.

[A subset U of $\mathbb{R}$ is by definition *open* if $U = \varnothing$ or U is a union of open intervals. This exercise proves that a non-empty open set is expressible as the union of a countable collection of disjoint open intervals.]

11 $\mathrm{L}^{\mathrm{inc}}$ functions

In this chapter we extend our class of integrable functions significantly, aiming once again to satisfy the Basic Properties set out in Chapter 1. Remember that L^{C} denotes the class of functions on $\mathbb{R}$ covered by the integration theory developed in Chapter 5: $f\colon \mathbb{R} \to \mathbb{R}$ belongs to L^{C} if and only if, for some compact interval $[a,b]$, $f \in \mathrm{C}[a,b]$ and $f = 0$ off $[a,b]$. We have an integral defined on $\mathrm{L}^{\mathrm{step}}$ and one defined on L^{C}, but we do not yet have a definition for $\int(\varphi + f)$, where $\varphi \in \mathrm{L}^{\mathrm{step}}$ and $f \in \mathrm{L}^{\mathrm{C}}$. We shall now encompass both $\mathrm{L}^{\mathrm{step}}$ and L^{C} at the same time, and more.

Guided by the results of Chapter 9, we look first at functions which can be approximated from below by increasing sequences of step functions, the monotonicity guaranteeing a natural candidate for the value of the integral. At the same time we incorporate null sets, to take account of the fact that the behaviour of a function on a null set should not affect its integral.

11.1 The class $\mathbf{L}^{\mathrm{inc}}$. Let $f\colon \mathbb{R} \to \mathbb{R}$. We say that $f \in \mathrm{L}^{\mathrm{inc}}$ if there exists a sequence $\{\varphi_n\}$ (known as an $\mathbf{L}^{\mathrm{inc}}$*-sequence for* f) such that

(i) $\varphi_n \in \mathrm{L}^{\mathrm{step}}$ for all n,

(ii) $\varphi_n \leqslant \varphi_{n+1}$ for all n,

(iii) there exists a finite constant K such that $\int \varphi_n \leqslant K$ for all n, and

(iv) $\varphi_n \xrightarrow{\text{a.e.}} f$.

Conditions (i), (ii), and property (P) for $\mathrm{L}^{\mathrm{step}}$ imply that $\{\int \varphi_n\}$ is an increasing sequence in $\mathbb{R}$. Therefore, given (i) and (ii), condition (iii) is equivalent to

(iii′) the sequence $\{\int \varphi_n\}$ converges to a finite limit.

(Remember our remarks on sequences in 2.14.) We have an obvious candidate for the integral of f, namely $\lim \int \varphi_n$. We discuss this fully in 11.7. First we examine $\mathrm{L}^{\mathrm{inc}}$ functions *per se*.

Suppose we have x such that $\varphi_m(x) > f(x)$ for some m. Then

$$\sup_n \varphi_n(x) \geqslant \varphi_m(x) > f(x),$$

so $\varphi_n(x) \nrightarrow f(x)$. Accordingly, $\{\varphi_n\}$ approaches f from below off the null set E on which $\{\varphi_n\}$ fails to converge to f.

Part (b) of the following theorem is little more than a statement of earlier results in new terminology; (c) is new.

11.2 Theorem.
(a) A function g equal a.e. to a function f in $\mathrm{L}^{\mathrm{inc}}$ is itself in $\mathrm{L}^{\mathrm{inc}}$.
(b) $\mathrm{L}^{\mathrm{step}} \subseteq \mathrm{L}^{\mathrm{inc}}$ and $\mathrm{L}^{\mathrm{C}} \subseteq \mathrm{L}^{\mathrm{inc}}$.
(c) Let $I = \langle a, b\rangle$ be a bounded interval. Let $f : I \to \mathbb{R}$ be bounded and continuous a.e. (and extended to take value 0 outside I). Then $f \in \mathrm{L}^{\mathrm{inc}}$. [In particular, if f is bounded and monotonic on I, then $f \in \mathrm{L}^{\mathrm{inc}}$.]

Proof. For (a), let $\varphi_n \to f$ except on the null set E and $f = g$ except on the null set F. Then $\varphi_n \to g$ except on $E \cup F$, which is null by 10.6. Thus an $\mathrm{L}^{\mathrm{inc}}$-sequence $\{\varphi_n\}$ for f serves also for g.

(b) Given a step function φ, the sequence $\{\varphi_n\}$ with $\varphi_n = \varphi$ for all n is an $\mathrm{L}^{\mathrm{inc}}$-sequence for φ. Now let $f \in \mathrm{L}^{\mathrm{C}}$. The sequence $\{\varphi_n\}$ constructed from f as in 9.2 is an $\mathrm{L}^{\mathrm{inc}}$-sequence for f.

To prove (c) we construct $\{\varphi_n\}$ exactly as in 9.2. We then have that $\{\varphi_n(x)\}$ converges at each point of continuity of f, so $\varphi_n \to f$ a.e. Also, there exists M such that $f(x) \leqslant M$ for all $x \in I$ by assumption, so that we have $\int \varphi_n \leqslant (b-a)M$ as before. Thus $\{\varphi_n\}$ is an $\mathrm{L}^{\mathrm{inc}}$-sequence for f.

[The final assertion in (c) is a consequence of a theorem which states that a monotonic function has at most a countable set of discontinuities (this is proved, for example, in [**1**]). It is therefore continuous a.e.] $\square$

11.3 Examples. Consider the function $\theta = \chi_{[0,1]\setminus\mathbb{Q}}$ which was introduced in Chapter 1. Because $\mathbb{Q} \cap [0,1]$ is null, $\theta = \chi_{[0,1]}$ a.e. Hence by (a) and (b), $\theta \in \mathrm{L}^{\mathrm{inc}}$ and, according to our intended definition, $\int \theta = \int f = 1$.

The following are confirmed by 11.2(c) to be in $\mathrm{L}^{\mathrm{inc}}$:

$$\text{(i) } \sin(x^{-1}) \text{ on } (0,1], \quad \text{(ii) } x\log x \text{ on } (0,1], \quad \text{(iii) } \operatorname{sgn}(\sin(\pi x^{-1})) \text{ on } (0,1].$$

The functions in (i) and (ii) are continuous and bounded on $(0,1]$, which is bounded (but not compact). That in (iii) takes only the values 0 and ± 1, so is bounded. It is continuous except at the points of the null set $\{0\} \cup \{n^{-1} \mid n \in \mathbb{N}\}$.

We cannot use 11.2(c) to show that the following are integrable:

$$\text{(i) } x^{-1} \text{ on } (0,1], \quad \text{(ii) } \log x \text{ on } (0,1], \quad \text{(iii) } 1 + \chi_{\mathbb{Z}}.$$

Here x^{-1} is continuous on the bounded interval $(0,1]$, but is not bounded; likewise for $\log x$ on $(0,1]$. The function in (iii) is bounded and continuous a.e. but its domain is not a bounded interval. [Later, when our final definition of integrable function is in place we shall see that the function in (ii) is integrable, and the others are not.]

11.4 Remarks. Do not confuse the statements
(a) $\{\varphi_n\}$ is an increasing sequence of functions, and
(b) $\{\varphi_n\}$ is a sequence of increasing functions.
Statement (a) means that $\varphi_n \leqslant \varphi_{n+1}$ for all n, that is, $\varphi_n(x) \leqslant \varphi_{n+1}(x)$ for all x and for all n; (b) means that, for each fixed n, $x \leqslant y$ implies that $\varphi_n(y)$ $\varphi_n(x) \leqslant \varphi_n(y)$.

Do not confuse the statements

(a) f is continuous a.e.;

(b) f is equal a.e. to a continuous function.

They are not the same. For example, $\chi_{\mathbb{Q}} = 0$ except on the null set $\mathbb{Q}$, but $\chi_{\mathbb{Q}}$ is not continuous at any point. Note that this provides an example of a function in $\mathrm{L}^{\mathrm{inc}}$ which is continuous nowhere. [The functions on a bounded interval I which are bounded and continuous a.e. can be shown to be exactly the functions which are Riemann integrable on I, with the Riemann and Lebesgue integrals coinciding; see **[18]**.]

We turn to some examples. These are artificial, but instructive. They introduce techniques we shall use repeatedly and reveal that integrable functions may sometimes behave in ways we might not have expected. We also see that functions of the types listed in 11.2 do not exhaust $\mathrm{L}^{\mathrm{inc}}$; in particular, there are $\mathrm{L}^{\mathrm{inc}}$ functions which are not of compact support and $\mathrm{L}^{\mathrm{inc}}$ functions which are unbounded. Later, when we have techniques for handling them efficiently, we shall return to functions more commonly encountered in applications: polynomials, trigonometric functions, exponentials, and so on.

11.5 Examples.

(1) Define g as follows:

$$g = \sum_{k=1}^{\infty} \frac{1}{k(k+1)} \chi_{[k,k+1)}.$$

Then g is not a step function as it has infinitely many points of discontinuity; see Fig. 11.1. However we may approximate g by the step functions

$$\varphi_n := \sum_{k=1}^{n} \frac{1}{k(k+1)} \chi_{[k,k+1)}.$$

For any x we have $\varphi_n(x) = g(x)$ for large enough n ($n > x$ suffices), so $\varphi_n(x) \to g(x)$ as $n \to \infty$. Also, $\varphi_n \leqslant \varphi_{n+1}$ (the two functions are equal, except on $[n, n+1)$, and on this interval $\varphi_n = 0$ and $\varphi_{n+1} > 0$). Finally,

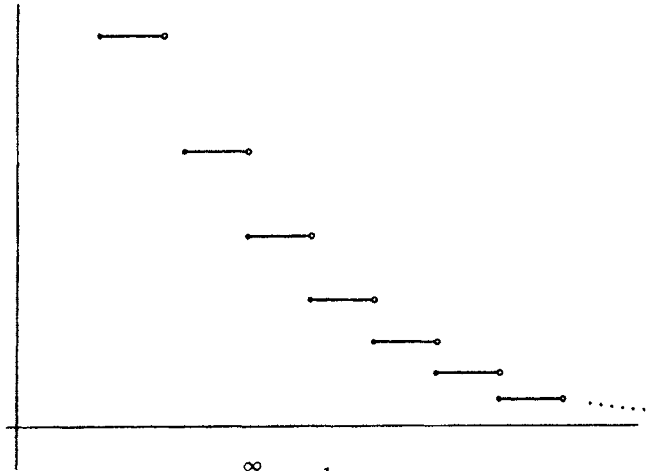
$$\int \varphi_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+1}\right) = 1 - \frac{1}{n+1} \leqslant 1.$$

Therefore $\{\varphi_n\}$ is an $\mathrm{L}^{\mathrm{inc}}$-sequence for g, with $\lim \int \varphi_n = 1$.

(2) Let h be defined by

$$h(x) = \begin{cases} k & \text{if } x \in (1/(k+1)!, 1/k!] \quad (k = 1, 2, \ldots), \\ 0 & \text{otherwise.} \end{cases}$$

Then the graph of h is as shown in Fig. 11.2. It is not a step function because it takes infinitely many different values. However we obtain a step function



$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} \chi_{[k,k+1)}$$

Figure 11.1

by truncating: take

$$\varphi_n := h\chi_{[1/(n+1)!,1]} = \sum_{k=1}^{n} (k+1)\chi_{(1/(k+1)!,1/k!]}.$$

We claim that $\{\varphi_n\}$ is an $\mathrm{L}^{\mathrm{inc}}$-sequence for h. We have already noted that φ_n is a step function. Its integral is

$$\int \varphi_n = \sum_{k=1}^{n} (k+1)\Big(\frac{1}{k!} - \frac{1}{(k+1)!}\Big) = \sum_{k=1}^{n} \frac{k(k+1)}{(k+1)!} = \sum_{k=0}^{n-1} \frac{1}{k!} \leqslant \mathrm{e}.$$

Also, for $x \notin (0,1]$, we have $\varphi_n(x) = h(x) = 0$, while for $x \in (1/(n+1)!, 1/n!]$, we have $\varphi_n(x) = h(x)$ if $n > 1/x$, so that $\varphi_n \to h$. Finally, $\varphi_n \leqslant \varphi_{n+1}$ because these functions are equal except on $(1/(n+2)!, 1/(n+1)!]$ and there $\varphi_n(x) = 0$ while $\varphi_{n+1}(x) = h(x) > 0$. We have proved that $h \in \mathrm{L}^{\mathrm{inc}}$. Note that h is not a bounded function.

11.6 Exercise example. Sketch the graph of f, where

$$f := \sum_{k=1}^{\infty} 2^k \chi_{[2^k, 2^k + 2^{-2k}]}.$$

Construct an $\mathrm{L}^{\mathrm{inc}}$-sequence $\{\varphi_n\}$ for $f \in \mathrm{L}^{\mathrm{inc}}$ and evaluate $\lim \int \varphi_n$.

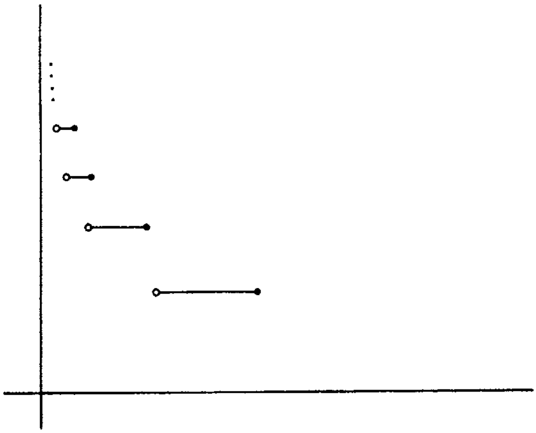


Figure 11.2

11.7 The integral on L^{inc}. Let $\{\varphi_n\}$ be an L^{inc}-sequence for $f \in L^{inc}$. As we have already noted, $\{\int \varphi_n\}$ converges to a finite limit, and we may define

$$\int f := \lim \int \varphi_n.$$

Observe that our definition of $\int f$ has been contrived so that

$$\lim \int \varphi_n = \int f = \int \lim \varphi_n.$$

However there is a possible snag: the same function f may have many different L^{inc}-sequences, and different sequences might give different values for $\int f$. To rescue our definition we have to prove that any two L^{inc}-sequences $\{\varphi_n\}$, $\{\psi_n\}$ for f give the same value for $\int f$, that is, that

$$\lim \int \varphi_n = \lim \int \psi_n.$$

We straightaway derive the technical lemma, 11.8, which yields this, and more (see 11.10). It relies on Technical Theorem III from Chapter 10.

11.8 Consistency lemma for L^{inc}. Suppose $f, g \in L^{inc}$, with L^{inc}-sequences $\{\varphi_n\}$, $\{\psi_n\}$, respectively, and that $f \geqslant g$ a.e. Then

$$\lim \int \varphi_n \geqslant \lim \int \psi_n.$$

Proof. Assume that $f \geqslant g$ except on the null set F. We have $\varphi_n \nearrow f$ off a null set E_1 and $\psi_n \nearrow g$ off a null set E_2. For each fixed k, the sequence $\{\psi_k - \varphi_n\}_{n \geqslant 1}$ is decreasing, with limit $\psi_k - f$, off the null set E_1. Off E_2, we have $\psi_k - f \leqslant g - f \leqslant 0$. Thus off the null set $E_1 \cup E_2 \cup F$ the sequence $\{\psi_k - \varphi_n\}$ converges to a non-positive limit, so that the sequence $\{(\psi_k - \varphi_n)^+\}$ converges to zero a.e. Now apply 10.11: this tells us that $\int (\psi_k - \varphi_n)^+ \to 0$, as $n \to \infty$, for each k. But

$$\int \psi_k - \int \varphi_n = \int (\psi_k - \varphi_n) \leqslant \int (\psi_k - \varphi_n)^+.$$

Let $n \to \infty$ to get $\int \psi_k \leqslant \lim \int \varphi_n$, for each k. Finally, letting $k \to \infty$ we obtain $\lim_{k\to\infty} \int \psi_k \leqslant \lim_{n\to\infty} \int \varphi_n$, as required. $\square$

11.9 More about $\mathrm{L}^{\mathrm{inc}}$ functions. Observe that the $\mathrm{L}^{\mathrm{inc}}$ function h considered in 11.5 is not bounded above, but is bounded below. In fact every f in $\mathrm{L}^{\mathrm{inc}}$ is necessarily bounded below on $\mathbb{R} \setminus E$ for some null set E. To see this, take an $\mathrm{L}^{\mathrm{inc}}$-sequence $\{\varphi_n\}$, with $\varphi_n \to f$ off the null set E. As noted above, $\varphi_n(x) \leqslant f(x)$ for all n and for all $x \notin E$. In particular, $\varphi_1(x) \leqslant f(x)$ for $x \notin E$. But φ_1 is bounded below, so there exists a constant $m \in \mathbb{R}$ such that $m \leqslant f(x)$ for $x \notin E$.

Now let $f = -h$ where h is as defined in 11.5. We assert that there do not exist a null set E and a constant M such that $f(x) \geqslant M$ for all $x \notin E$. Suppose it were possible to find such M and E. Fix a natural number $n \geqslant M$. Then $f(x) < M$ on $(0, 1/(n+1)!)$. Therefore $(0, 1/(n+1)!) \subseteq E$, contradicting the fact that E is a null set (remember 10.10). We deduce that we have $h \in \mathrm{L}^{\mathrm{inc}}$ but $-h \notin \mathrm{L}^{\mathrm{inc}}$.

Exercise 11.5(ii) seeks a proof that $k \in \mathrm{L}^{\mathrm{inc}}$, where $k(x) := \mathrm{e}^{-x}\chi_{[0,\infty)}(x)$. Observe that $-k \notin \mathrm{L}^{\mathrm{inc}}$ just because there is **no** step function φ with $\varphi \leqslant -k$. Indeed, no function of compact support lies below $-k$.

These examples are perhaps not totally surprising in view of the inherent asymmetry in the $\mathrm{L}^{\mathrm{inc}}$ definition. Indeed, suppose we have $f \in \mathrm{L}^{\mathrm{inc}}$, with an $\mathrm{L}^{\mathrm{inc}}$ sequence $\{\varphi_n\}$. Now consider $-f$. Clearly $\{-\varphi_n\}$ is a sequence of step functions converging a.e. to $-f$. However $\{-\varphi_n\}$ is a **decreasing** sequence. This should be enough to make us suspicious that we might not be able to prove in general that $f \in \mathrm{L}^{\mathrm{inc}}$ implies $-f \in \mathrm{L}^{\mathrm{inc}}$. Our examples confirm this.

We now assemble those of our Basic Properties that we can prove for $\mathrm{L}^{\mathrm{inc}}$.

11.10 Lemma. Let $f, g \in \mathrm{L}^{\mathrm{inc}}$ with $f \geqslant g$ a.e. Then $\int f \geqslant \int g$.

Proof. The result is just a reiteration of the Consistency Lemma, 11.8. $\square$

11.11 Exercise example. Let $f \in \mathrm{L}^{\mathrm{inc}}$ and, for $d \in \mathbb{R}$, define f_d and f^d as in 2.9. Then f_d and f^d belong to $\mathrm{L}^{\mathrm{inc}}$, and $\int f_d = \int f$ and, if $d \neq 0$, $|d| \int f^d = \int f$. [Hint: you will need 4.7, Exercise 4.2, and 10.7.]

11.12 Lemma.

(a) If $f, g \in \mathrm{L}^{\mathrm{inc}}$ then $f + g \in \mathrm{L}^{\mathrm{inc}}$ and $\int (f+g) = \int f + \int g$.
(b) If $f \in \mathrm{L}^{\mathrm{inc}}$ and $\lambda \in \mathbb{R}$, $\lambda \geqslant 0$, then $\lambda f \in \mathrm{L}^{\mathrm{inc}}$ and $\int (\lambda f) = \lambda \int f$.
(c) If $f, g \in \mathrm{L}^{\mathrm{inc}}$ then $f \vee g, f \wedge g \in \mathrm{L}^{\mathrm{inc}}$. In particular, if $f \in \mathrm{L}^{\mathrm{inc}}$ then $f^+ := f \vee 0 \in \mathrm{L}^{\mathrm{inc}}$.
(d) If $f \in \mathrm{L}^{\mathrm{inc}}$ and $\varphi \in \mathrm{L}^{\mathrm{step}}$ then $f \pm \varphi \in \mathrm{L}^{\mathrm{inc}}$.

Proof. In every case, the obvious argument works. For example, in (a), if $\{\varphi_n\}$, $\{\psi_n\}$ are $\mathrm{L}^{\mathrm{inc}}$ sequences for f, g respectively, then $\{\varphi_n + \psi_n\}$ is an $\mathrm{L}^{\mathrm{inc}}$ sequence for $f + g$ (why?). Also,

$$\int(f+g) = \lim \int(\varphi_n + \psi_n) = \lim(\int \varphi_n + \int \psi_n) = \lim \int \varphi_n + \lim \int \psi_n = \int f + \int g.$$

We leave the remainder of the proof as an exercise. □

Step functions, except the zero function, are not continuous. This is occasionally a nuisance. Exercise 3.6 sought an approximation to a step function φ by a continuous function which differs little from φ. The following technical lemma improves on this. It shows that a step function can be approximated well by step-like functions which are not just continuous but as smooth as we could possibly wish.

11.13 Smooth approximations to step functions. Consider the functions

$$p(x) = \mathrm{e}^{-1/x}\chi_{(0,\infty)}(x), \qquad q(x) = \frac{p(x)}{p(x) + p(1-x)}\chi_{(0,\infty)}(x),$$
$$r(x) = q(2-x)q(x-2),$$

as shown in Fig. 11.3. We claim that all the derivatives $p^{(m)}(0)$ exist and are zero. To prove this it suffices to show for every m that $p^{(m)}(x)/x \to 0$ as $x \to 0+$; for this we use induction, including in the inductive hypothesis the assertion that $p^{(m)}(x)$ takes the form $Q(x)\mathrm{e}^{-1/x}$ for $x > 0$, where Q is a rational function. We can then appeal to 8.16, which implies that $x^k\mathrm{e}^{-1/x} \to 0$ as $x \to 0+$ for any $k \in \mathbb{R}$. We leave the details as an exercise. It follows easily that each of p, q, and r belongs to $\mathrm{C}^\infty(\mathbb{R})$. Also $r = 0$ outside $[-2, 2]$ and $r = 1$ on $[-1, 1]$.

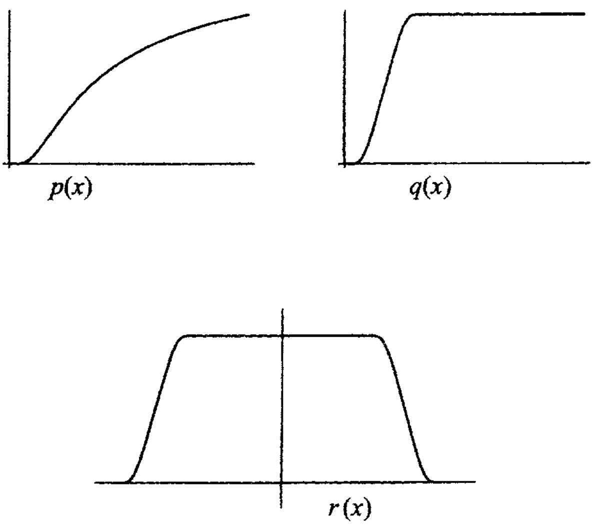


Figure 11.3

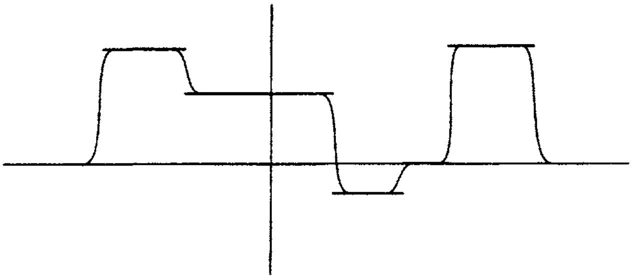


Figure 11.4

Now let $\varphi \in \mathrm{L}^{\mathrm{step}}$ and $\varepsilon > 0$. It follows from above, by taking linear combinations and rescaling, that we can find a C^∞ function g such that $\int |\varphi - g| < \varepsilon$ and $\varphi = g$ except on finitely many intervals of total length not greater than ε. See Fig. 11.4.

11.14 Postscript. We sum up. We have found a class, $\mathrm{L}^{\mathrm{inc}}$, of (real-valued) functions, which contains $\mathrm{L}^{\mathrm{step}}$ and L^{C}. We have defined on $\mathrm{L}^{\mathrm{inc}}$ an integral agreeing with that defined for these subclasses in earlier chapters; this is obvious for $\mathrm{L}^{\mathrm{step}}$ and proved for L^{C} in 9.7 and 9.6. Moreover, $\mathrm{L}^{\mathrm{inc}}$ nearly meets the criteria set out in Chapter 1, failing only to satisfy the requirement that $f \in \mathrm{L}^{\mathrm{inc}}$ implies $-f \in \mathrm{L}^{\mathrm{inc}}$. We want next to enlarge $\mathrm{L}^{\mathrm{inc}}$ to obtain a class of functions closed under subtraction.

Exercises

11.1 Prove that if $f\colon \mathbb{R} \to \mathbb{R}$ is continuous and zero a.e., then $f \equiv 0$ (see 11.4).

11.2 Prove that the following functions belong to $\mathrm{L}^{\mathrm{inc}}$ and evaluate their integrals:

$$\text{(i)} \sum_{k=1}^{\infty} (-1)^k((k-1)!)^{-1}\chi_{(1/(k+1),1/k]}, \qquad \text{(ii)} \sum_{k=1}^{\infty} (-1)^k k^{-1}\chi_{(1/(k+1)!,1/k!]}.$$

11.3 Let f, g have $\mathrm{L}^{\mathrm{inc}}$-sequences $\{\varphi_n\}$, $\{\psi_n\}$, respectively. Why is it not necessarily true that $\{\varphi_n\psi_n\}$ is an $\mathrm{L}^{\mathrm{inc}}$-sequence for fg? Prove that if g is bounded then $fg \in \mathrm{L}^{\mathrm{inc}}$.

11.4 Let I be a subinterval of $\mathbb{R}$. Let $f\colon \mathbb{R} \to \mathbb{R}^+$. Prove that $f \in \mathrm{L}^{\mathrm{inc}}$ implies $f\chi_I \in \mathrm{L}^{\mathrm{inc}}$. For $I = \langle a,b\rangle$ $(-\infty \leqslant a < b \leqslant \infty)$, write $\int_a^b f := \int f\chi_I$. Prove that

$$\int_a^b f = \int_a^c f + \int_c^b f \quad (a < c < b).$$

[See also 12.10.]

11.5 [For masochists only—better methods later!]

(a) By constructing an $\mathrm{L^{inc}}$-sequence prove that $x^{-1/2}\chi_{(0,1]}(x)$ is in $\mathrm{L^{inc}}$.

(b) Prove that $\mathrm{e}^{-x}\chi_{[0,\infty)}(x) \in \mathrm{L^{inc}}$ by constructing an $\mathrm{L^{inc}}$-sequence $\{\varphi_n\}$ in which φ_n is constant on the intervals obtained by subdividing $[0, 2^n]$ into 2^{2n} subintervals of length 2^{-n}.

11.6 Let f be a non-negative continuous function on $\mathbb{R}$ and suppose that there is a constant K independent of n such that $\int_{-n}^{n} f \leqslant K$ for all n. Let $\{\varphi_{n,m}\}$ be an $\mathrm{L^{inc}}$-sequence for f on $[-m, m]$ and let $\psi_n := \varphi_{n,n}$. By showing that $\{\psi_n\}$ is an $\mathrm{L^{inc}}$-sequence for f prove that $f \in \mathrm{L^{inc}}$ and that $\int f = \lim \int_{-n}^{n} f$. [The strategy advocated here covers a range of examples like those in Exercise 11.5. It is superceded by the methods in Chapter 15.]

The two remaining exercises present more exotic examples of $\mathrm{L^{inc}}$ functions. They are more challenging, and are not needed subsequently.

11.7 Define f by

$$f(x) = \begin{cases} 2^{-k} & \text{if } 2^{-k-1} < x \leqslant 2^{-k} \quad (k = 0, 1, 2, \dots), \\ 0 & \text{otherwise.} \end{cases}$$

(a) Prove that $f \in \mathrm{L^{inc}}$ and evaluate $\int f$.

(b) Let

$$F(x) := \int_0^x f(t)\,\mathrm{d}t.$$

Evaluate $F(x)$ for $2^{-n-1} < x \leqslant 2^{-n}$ (recall Exercise 5.7) and deduce that

$$F(x) = xF(x) - \frac{1}{3}\{A(x)\}^2, \text{ where } A(x) = 2^{-(\log x^{-1}/\log 2)}.$$

11.8 Let $\{q_1, q_2, \dots\}$ be an enumeration of the rationals in $[0, 1]$. Define

$$f(x) = \sum_{q_n \leqslant x} 2^{-n}.$$

(a) Prove that f is discontinuous at each rational point in $[0, 1]$ and continuous at each irrational point.

(b) Deduce from 11.2(c) that $f\chi_{[0,1]} \in \mathrm{L^{inc}}$ and show that

$$\int_0^x f(t)\,\mathrm{d}t = 1 - \sum_{q_n \leqslant x} q_n 2^{-n}.$$

12 The class L of integrable functions

We have reached the final stages of the construction of the integral. Most of the hard work has already been done and the properties to be verified are familiar. We finish the construction without presenting further concrete examples. Succeeding chapters abound with examples.

12.1 The class $L_{\mathbb{R}}$. Our observations in 11.9 force us to confront the fact that L^{inc} is not the class we should like to designate as our class $L_{\mathbb{R}}$ of real-valued integrable functions, since we want $f \in L_{\mathbb{R}}$ implies $-f \in L_{\mathbb{R}}$. What is the remedy? Our first idea might be to let $L_{\mathbb{R}} = \{ f : \mathbb{R} \to \mathbb{R} \mid f \in L^{\text{inc}} \text{ or } -f \in L^{\text{inc}} \}$. We want $f, g \in L_{\mathbb{R}}$ implies $f + g \in L_{\mathbb{R}}$. However it is not obvious that if $f \in L^{\text{inc}}$ and $-g \in L^{\text{inc}}$ then $f + g \in L^{\text{inc}}$ or $-(f+g) \in L^{\text{inc}}$. The correct way to extend L^{inc} is to define

$$L_{\mathbb{R}} := \{ f : \mathbb{R} \to \mathbb{R} \mid f = g - h \text{ where } g, h \in L^{\text{inc}} \}.$$

Since $0 \in L^{\text{inc}}$, we have that $f \in L^{\text{inc}}$ implies $f \in L_{\mathbb{R}}$, because we can write $f = f - 0$. Similarly, given $-f \in L^{\text{inc}}$, we can write $f = 0 - (-f)$, so $f \in L_{\mathbb{R}}$.

The functions in $L_{\mathbb{R}}$ are the real-valued functions which we shall call *integrable* or *Lebesgue integrable.* We have written $L_{\mathbb{R}}$ since we wish to reserve the symbol L for the bigger class of complex-valued integrable functions we introduce in 12.11. We must now define $\int f$ for $f \in L_{\mathbb{R}}$. Writing $f = g - h$, where $g, h \in L^{\text{inc}}$, we set

$$\int f := \int g - \int h.$$

Here yet again we are faced with a consistency question. If we decompose f into a difference of L^{inc} functions in another way, then we might get a different value for the integral. Fortunately this is easily resolved.

12.2 Consistency lemma for $L_{\mathbb{R}}$. Assume that $f = g - h = g^* - h^*$, where $g, h, g^*, h^* \in L^{\text{inc}}$. Then

$$\int g - \int h = \int g^* - \int h^*.$$

Proof. We have

$$g + h^* = g^* + h.$$

By 11.12, $\int g + \int h^* = \int g^* + \int h$, from which the result follows at once. □

For functions which are real-valued, we now have the class of integrable functions we seek. The next few results confirm this. The proofs, apart from the first one, require a little cunning.

12.3 Property (T) for $L_{\mathbb{R}}$. Let $f \in L_{\mathbb{R}}$. For $d \in \mathbb{R}$ let f_d be defined by $f_d(x) := f(x+d)$. Then $f_d \in L_{\mathbb{R}}$ and $\int f_d = \int f$. Also, $f^d \in L_{\mathbb{R}}$ (where $f^d(x) = f(dx)$) and $|d| \int f^d = \int f$ for $d \neq 0$.

Proof. Note that the result has already been proved for L^{step} in 4.7 and extended to L^{inc} in 11.11. The extension to $L_{\mathbb{R}}$ is immediate from the definition of the integral on $L_{\mathbb{R}}$. □

12.4 Property (L) for $L_{\mathbb{R}}$. Let $f_1, f_2 \in L_{\mathbb{R}}$ and $\lambda \in \mathbb{R}$. Then $f_1 + \lambda f_2 \in L_{\mathbb{R}}$ and

$$\int (f_1 + \lambda f_2) = \int f_1 + \lambda \int f_2.$$

Proof. Write $f_i = g_i - h_i$, where $g_i, h_i \in L^{\text{inc}}$ $(i = 1, 2)$. Let $\lambda \geqslant 0$. Then $f_1 + \lambda f_2 = (g_1 + \lambda g_2) - (h_1 + \lambda h_2)$, and $g_1 + \lambda g_2, h_1 + \lambda h_2 \in L^{\text{inc}}$, by 11.12. Thus $f_1 + \lambda f_2 \in L_{\mathbb{R}}$ and

$$\begin{aligned} \int (f_1 + \lambda f_2) &:= \int (g_1 + \lambda g_2) - \int (h_1 + \lambda h_2) = \\ &\int g_1 + \lambda \int g_2 - \int h_1 - \lambda \int h_2 = \int f_1 + \lambda \int f_2, \end{aligned}$$

using 11.12 for the middle equality. To complete the proof we need to consider $-f_2$. We have $-f_2 = h_2 - g_2$, and $\int (-f_2) := \int g_2 - \int h_2 = -\int f_2$. □

12.5 Property (P) for $L_{\mathbb{R}}$. Let $f_1, f_2 \in L_{\mathbb{R}}$ with $f_1 \geqslant f_2$ a.e. Then $\int f_1 \geqslant \int f_2$. In particular, $f \in L_{\mathbb{R}}$ and $f \geqslant 0$ imply $\int f \geqslant 0$.

Proof. Write $f_i = g_i - h_i$, where $g_i, h_i \in L^{\text{inc}}$ $(i = 1, 2)$. Then $f_1 \geqslant f_2$ a.e. implies $g_1 + h_2 \geqslant g_2 + h_1$ a.e., and each side belongs to L^{inc}. Hence, by 11.10 and 11.12, we get

$$\int g_1 + \int h_2 = \int (g_1 + h_2) \geqslant \int (g_2 + h_1) = \int g_2 + \int h_1.$$

Therefore $\int f_1 = \int g_1 - \int h_1 \geqslant \int g_2 - \int h_2 = \int f_2$. □

12.6 Property (M) for $L_{\mathbb{R}}$. Let $f \in L_{\mathbb{R}}$. Then $|f| \in L_{\mathbb{R}}$ and $\left|\int f\right| \leqslant \int |f|$.

Proof. Write $f = g - h$ with $g, h \in L^{\text{inc}}$. It is easily verified that

$$|f| = (g \vee h) - (g \wedge h),$$

from which integrability of $|f|$ follows from 11.12 and 12.4. The inequality is harder to derive for $L_{\mathbb{R}}$ than for L^{inc}. To obtain it we need

$$\int g - \int h = \int f \leqslant \int |f| = \int (g \vee h) - (g \wedge h) \quad \text{and}$$

$$\int h - \int g = -\int f \leqslant \int |f| = \int (g \vee h) - (g \wedge h),$$

and by symmetry it suffices to prove the first of these. For this it is enough to check that

$$g + (g \wedge h) \leqslant h + (g \vee h),$$

since the result then follows from 11.12 and 11.10. Take any x. If $g(x) \leqslant h(x)$ then $g(x) + (g \wedge h)(x) = 2g(x) \leqslant 2h(x) = h(x) + (g \vee h)(x)$, while if $g(x) > h(x)$ we have $g(x) + (g \wedge h)(x) = g(x) + h(x) = h(x) + (g \vee h)(x)$, as required. □

12.7 Proposition.

(a) Let $f\colon \mathbb{R} \to \mathbb{R}$. Let $f^+ := f \vee 0$, $f^- := (-f) \vee 0$. Then $f \in \mathrm{L}_{\mathbb{R}}$ if and only if f^+, $f^- \in \mathrm{L}_{\mathbb{R}}$.

(b) Let $f_1, f_2 \in \mathrm{L}_{\mathbb{R}}$. Then $f_1 \wedge f_2$, $f_1 \vee f_2 \in \mathrm{L}_{\mathbb{R}}$.

Proof. Use the preceding result and the formulae

$$f = f^+ - f^-, \qquad |f| = f^+ + f^-,$$
$$2(f_1 \wedge f_2) = (f_1 + f_2) - |f_1 - f_2|, \quad 2(f_1 \vee f_2) = (f_1 + f_2) + |f_1 - f_2|;$$

see 2.7. □

12.8 Remarks. We now know that $f \in \mathrm{L}_{\mathbb{R}}$ implies $|f| \in \mathrm{L}_{\mathbb{R}}$. The converse is not true. Suppose, for example, we break $[0,1]$ into two disjoint sets E and F in a bizarre way. Then the values of

$$f := \chi_E - \chi_F$$

jump wildly between 1 and -1. We shall discuss in Chapter 22 how E and F may be chosen so that f is too wild to be integrable. Of course $|f| = \chi_{[0,1]}$, and this is integrable. These considerations also indicate why we did not define $\mathrm{L}_{\mathbb{R}}$ by $f \in \mathrm{L}_{\mathbb{R}}$ if and only if $|f| \in \mathrm{L}^{\mathrm{inc}}$—at first sight a tempting way to try to extend $\mathrm{L}^{\mathrm{inc}}$.

As a corollary of 12.4 and 12.7, $f_1, \ldots, f_n \in \mathrm{L}_{\mathbb{R}}$ implies that

$$f_1 + \cdots + f_n, \quad f_1 \vee \cdots \vee f_n, \text{ and } f_1 \wedge \cdots \wedge f_n$$

belong to $\mathrm{L}_{\mathbb{R}}$ for any $n \geqslant 2$. We stress that we are dealing here with **finite** sums, **finite** sups, and **finite** infs. It is not automatic that, for a given **infinite** sequence $f_1, f_2, \ldots$ of functions in $\mathrm{L}_{\mathbb{R}}$, $\sum_{k=1}^{\infty} f_k$, $\sup_k f_k$, or $\inf_k f_k$ is integrable. We return to such functions in Chapters 14–16.

If f_1, f_2 both belong to $\mathrm{L}^{\mathrm{step}}$ or both belong to $\mathrm{C}[a,b]$ then their product belongs to the same class and so is integrable. It is not true in general that the product of two integrable functions is integrable. We discuss products in detail in Chapter 21.

12.9 Functions and null sets: a notational convention. Suppose that the functions $f, g\colon \mathbb{R} \to \mathbb{R}$ are equal a.e. Then $f \in \mathrm{L}_{\mathbb{R}}$ if and only if $g \in \mathrm{L}_{\mathbb{R}}$, by the corresponding result for $\mathrm{L}^{\mathrm{inc}}$ proved in 11.10, and $\int f = \int g$. So, as we intended, we don't care what values a given function takes on a null set. We can take this one stage further. Consider for example

$$\frac{\sin x}{x}, \quad \frac{1}{x^2 - 1}, \quad \log(1 - \sin^2(\pi x)), \quad \text{and} \quad \lim_{n\to\infty} \sum_{r=1}^{n} x^r \chi_{[0,1]}(x).$$

As written, none of these is a function from $\mathbb{R}$ to $\mathbb{R}$: each is either undefined, or takes an infinite value, at certain points. The first function is undefined at 0, but becomes a continuous function on $\mathbb{R}$ when assigned its limiting value 1 at 0. The second function is undefined at ± 1, and although continuous on $\mathbb{R} \setminus \{\pm 1\}$, cannot be extended to a continuous function on $\mathbb{R}$. The final function is infinite at 1 and the third one takes value $-\infty$ at every integer point. In every case the 'bad' points form a null set and so should not greatly dismay us. We shall adopt the following convention. Assume that f is defined and takes a value $f(x) \in \mathbb{R}$ at each point x of $\mathbb{R} \setminus E$, where E is null. Then we may define (or redefine) $f(x) = 0$ for $x \in E$. The resulting function f is a bona fide real-valued function on the whole of $\mathbb{R}$, agreeing with the given function except possibly on E. We shall always in our theorems assume that we work with functions $f \colon \mathbb{R} \to \mathbb{R}$, and assume that any particular functions are adjusted in the manner described above to ensure that they fit this requirement.

12.10 Integrals on intervals. Integration can be cut down to any subinterval I of $\mathbb{R}$. It is hardly surprising that $f \in \mathrm{L}_{\mathbb{R}}$ implies $f\chi_I \in \mathrm{L}_{\mathbb{R}}$, but slight care is needed in checking this. Certainly it will be sufficient to prove the claim for $f \in \mathrm{L}^{\mathrm{inc}}$. Let $\{\varphi_n\}$ be an $\mathrm{L}^{\mathrm{inc}}$-sequence for f. The obvious candidate for an $\mathrm{L}^{\mathrm{inc}}$-sequence for $f\chi_I$ is $\{\psi_n\}$, where $\psi_n := \varphi_n \chi_I$. This satisfies the $\mathrm{L}^{\mathrm{inc}}$ conditions, with the possible exception of (iii). We know that $\{\int \varphi_n\}$ is bounded above, by K say. Unfortunately we can't simply say that $\{\int \psi_n\}$ is also bounded above by K—though this is true if $\varphi_n \geqslant 0$ for all n, because then $\int \psi_n \leqslant \int \varphi_n$ for all n. We can reduce to this case by considering $\{\varphi_n - \varphi_1\}$ which provides a non-negative $\mathrm{L}^{\mathrm{inc}}$-sequence for $f - \varphi_1$. Hence $(f - \varphi_1)\chi_I \in \mathrm{L}^{\mathrm{inc}}$. Finally, $f\chi_I = (f - \varphi_1)\chi_I + \varphi_1\chi_I$ and this belongs to $\mathrm{L}_{\mathbb{R}}$ since $\varphi_1\chi_I \in \mathrm{L}^{\mathrm{step}}$.

Given any function f whose domain contains I we say f is integrable on I (and write $f \in \mathrm{L}_{\mathbb{R}}(I)$) just when $f\chi_I$ belongs to $\mathrm{L}_{\mathbb{R}}$. The theorems about $\mathrm{L}_{\mathbb{R}}$ given above have analogues for $\mathrm{L}_{\mathbb{R}}(I)$, obtained by multiplying all functions involved by χ_I. We shall not state such theorems explicitly, but shall use them freely. To avoid a double set of brackets we write $\mathrm{L}_{\mathbb{R}}(\langle a, b\rangle)$ as $\mathrm{L}_{\mathbb{R}}[a, b]$ $(-\infty < a < b < \infty)$; since the various intervals with endpoints a, b differ only by null sets this abuse of notation is justifiable.

We adopt the usual notation

$$\int_a^b f \quad \text{or} \quad \int_I f := \int f\chi_I,$$

where a, b are the endpoints of I $(-\infty \leqslant a \leqslant b \leqslant \infty)$; if $a > b$, then we define $\int_a^b f$ to be $-\int_b^a f$. In Chapter 5 we proved that, for $f \in \mathrm{C}[a, b]$ and $a < c < b$,

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

The same result holds whenever f is such that the integral on the left-hand side exists. This is immediate from 12.4 because $\chi_I = \chi_J + \chi_K$ a.e. whenever I, J, and K are intervals with endpoints a, b, a, c, and c, b, respectively.

Frequently we meet integrals with variable limits such as

$$\int_0^n \mathrm{e}^{-x}\,\mathrm{d}x \quad \text{or} \quad \int_0^y \frac{1}{1 + x^2y^2}\,\mathrm{d}x.$$

When manipulating such integrals—taking the limit as $n \to \infty$ or differentiating with respect to y for example—we recommend using a characteristic function to build the limits into the integrand. This way, the limits require no special attention. For examples of this advice in action, see 16.6 and 20.5.

12.11 Complex-valued functions: the class L. We can extend our integral to complex-valued functions as follows. If $f\colon \mathbb{R} \to \mathbb{C}$ then we can write f in the form $f = \operatorname{Re} f + \mathrm{i} \operatorname{Im} f$, where $\operatorname{Re} f$ and $\operatorname{Im} f$ are real-valued functions. We say f is integrable if $\operatorname{Re} f$ and $\operatorname{Im} f$ belong to $\mathrm{L}_{\mathbb{R}}$, and define

$$\int f := \int \operatorname{Re} f + \mathrm{i} \int \operatorname{Im} f.$$

We write L for the class of complex-valued integrable functions. This class contains the complex analogues of L^{step} and L^{C}: functions $u + \mathrm{i}v$, where $u, v \in \mathrm{L}^{\text{step}}$ or $u, v \in \mathrm{L}^{\mathrm{C}}$. We refer to functions of the former type as complex step functions. We can also, in the obvious way, build $\mathrm{L}(I)$, the complex-valued integrable functions on an interval I. Integrals of complex-valued functions are important in applications of analysis, for example in the theory of Fourier and Laplace transforms. Among the complex L^{C} integrals are the integrals along paths in the complex plane which are so ubiquitous in complex analysis; see **[13]**.

There is no consistency problem with the extension of the integral to complex-valued functions, and properties (L) and (T) extend to it in a routine way. Property (P) does not make sense for functions which are not real-valued. Property (M) does make sense, and is true, but we cannot prove it until later; see 21.6.

12.12 The FTC for complex-valued functions. First consider $f(x) = \mathrm{e}^{\mathrm{i}x}$. Then we have $(\operatorname{Re} f)(x) = \cos x$ and $(\operatorname{Im} f)(x) = \sin x$ and these are integrable on any compact interval. In particular, by definition and the FTC, 5.5,

$$\int_0^\pi \mathrm{e}^{\mathrm{i}x}\,\mathrm{d}x := \int_0^\pi \cos x\,\mathrm{d}x + \mathrm{i}\int_0^\pi \sin x\,\mathrm{d}x = 2\mathrm{i}.$$

Now suppose we wish to evaluate

$$\int_0^\pi \mathrm{e}^{-\alpha x} \cos x\,\mathrm{d}x \quad \text{and} \quad \int_0^\pi \mathrm{e}^{-\alpha x} \sin x\,\mathrm{d}x \qquad (\alpha \in \mathbb{R}).$$

This can be done by integrating by parts twice. However it would be slicker to write

$$\begin{aligned}
\int_0^\pi \mathrm{e}^{-\alpha x} \cos x\,\mathrm{d}x + \mathrm{i}\int_0^\pi \mathrm{e}^{-\alpha x} \sin x\,\mathrm{d}x &= \int_0^\pi \mathrm{e}^{-\alpha x + \mathrm{i}x}\,\mathrm{d}x \qquad \text{(by definition)}\\
&= \left[\frac{1}{\mathrm{i}-\alpha}\mathrm{e}^{-\alpha x + \mathrm{i}x}\right]_0^\pi\\
&= \frac{1}{\alpha - \mathrm{i}}\left(1 + \mathrm{e}^{-\alpha\pi}\right)\\
&= \frac{\alpha + \mathrm{i}}{\alpha^2 + 1}\left(1 + \mathrm{e}^{-\alpha\pi}\right),
\end{aligned}$$

whence, equating real and imaginary parts,

$$\int_0^\pi e^{-\alpha x}\cos x\,dx = \frac{\alpha(1+e^{-\alpha\pi})}{\alpha^2+1} \quad\text{and}\quad \int_0^\pi e^{-\alpha x}\sin x\,dx = \frac{(1+e^{-\alpha\pi})}{\alpha^2+1}.$$

Here we have assumed that

$$(*) \qquad \int_a^b F' = F(b) - F(a)$$

is valid for the function $F(x) = e^{\beta x}$ when β is a complex constant. It is. It can be shown that the equation $(*)$ holds whenever F is holomorphic (or equivalently represented by a convergent complex power series) in a region of the complex plane containing the interval $[a,b]$ on the real axis; cf. **[13]**, in particular section 3.8.

We now give some useful facts on approximations which we shall need en route to important theorems later on. We shall examine more closely what 12.13 tells us about integrable functions when we consider in Chapter 23 what the graphs of integrable functions look like.

12.13 Lemma. Let $\varepsilon > 0$.

(a) Let $f \in L_{\mathbb{R}}$ ($f \in L$). Then there exists a step function (complex step function) φ such that $\int |f - \varphi| < \varepsilon$.

(b) Let $f \in L_{\mathbb{R}}$. Then f may be expressed as $g^* - h^*$ with $g^*, h^* \in L^{\mathrm{inc}}$, $h^* \geqslant 0$ and $0 \leqslant \int h^* < \varepsilon$.

Proof. First assume $f \in L_{\mathbb{R}}$. Write $f = g - h$ where $g, h \in L^{\mathrm{inc}}$ and let $\{\varphi_n\}$, $\{\psi_n\}$ be L^{inc}-sequences for g, h respectively. By definition, $\int \varphi_n \nearrow \int g$ and $\int \psi_n \nearrow \int h$. Hence we can choose N such that, for all $n \geqslant N$,

$$0 \leqslant \int g - \int \varphi_n < \varepsilon/2 \quad\text{and}\quad 0 \leqslant \int h - \int \psi_n < \varepsilon/2.$$

Define $\varphi := \varphi_N - \psi_N$. Then, by the triangle inequality and properties (P) and (L) for $L_{\mathbb{R}}$,

$$\begin{aligned}\int |f - \varphi| &= \int |(g - \varphi_N) - (h - \psi_N)| \\ &\leqslant \int |g - \varphi_N| + |h - \psi_N| \\ &= \int (g - \varphi_N) + (h - \psi_N) \\ &= \int (g - \varphi_N) + \int (h - \psi_N) \\ &< \varepsilon.\end{aligned}$$

For (b), define $g^* := g - \varphi$ and $h^* := h - \varphi$. (This in fact gives $\int h^* < \varepsilon/2$, which certainly suffices.)

In case f is not real-valued, write $f = \operatorname{Re} f + \mathrm{i} \operatorname{Im} f$. There exist, by above, $\psi, \theta \in \mathrm{L}^{\mathrm{step}}$ such that $\int |\operatorname{Re} f - \psi| < \varepsilon/2$ and $\int |\operatorname{Im} f - \theta| < \varepsilon/2$. Let φ be the complex step function $\psi + \mathrm{i}\theta$. Then

$$\int |f - \varphi| = \int |(\operatorname{Re} f - \psi) + \mathrm{i}(\operatorname{Im} f - \theta)| \leqslant \int |\operatorname{Re} f - \psi| + |\operatorname{Im} f - \theta| < \varepsilon.$$

This completes the proof of (a). □

12.14 Stocktaking. We now have our key definitions in place, and the Basic Properties established. We know that functions in $\mathrm{L}^{\mathrm{step}}$ and L^{C} fit into the framework we have set up, and have presented some examples of $\mathrm{L}^{\mathrm{inc}}$ (and hence L) functions outside these classes. We outline the steps in our construction.

Stage 1. For $f = \chi_I$, where I is a bounded interval in $\mathbb{R}$ with endpoints a and b, define $\int f := (b - a)$.

Stage 2. For f a step function, that is, for f a finite linear combination of the functions from Stage 1, define $\int f$ linearly.

Stage 3. For $f \in \mathrm{L}^{\mathrm{inc}}$, that is, for f the limit a.e. of an increasing sequence $\{\varphi_n\}$ of functions from Stage 2 for which $\lim \int \varphi_n$ exists, define $\int f := \lim \int \varphi_n$.

Stage 4. For $f \in \mathrm{L}_{\mathbb{R}}$, that is, for f expressible as the difference of functions from Stage 3, define $\int f$ linearly.

Stage 5. For $f = \operatorname{Re} f + \mathrm{i}\operatorname{Im} f \in \mathrm{L}$, so $\operatorname{Re} f$, $\mathrm{i}\operatorname{Im} f$, $f \in \mathrm{L}_{\mathbb{R}}$, define $\int f := \int \operatorname{Re} f + \mathrm{i} \int \operatorname{Im} f$.

Note that our construction suggests a natural strategy for proving results about functions in $\mathrm{L}_{\mathbb{R}}$ or in L. To establish such a result we may proceed as follows.

(1) Check the result for characteristic functions of compact intervals 'with our bare hands', and extend it to step functions by linearity (or establish it directly for step functions).

(2) Use (1) and a limiting argument to extend the result to $\mathrm{L}^{\mathrm{inc}}$ (the hard step).

(3) Take advantage of linearity again to reach $\mathrm{L}_{\mathbb{R}}$ or L, as appropriate.

The proof of property (T) is of this form, and the proof below of the Riemann-Lebesgue Lemma, 23.4, is another good illustration. Most significantly, this strategy is employed in the proof of the Monotone Convergence Theorem in Chapter 14. Once we have this theorem at our disposal, we rather rarely thereafter need to proceed in stages from step functions.

In the above presentation of our construction, L^{C} is side-lined. It is not necessary to consider the L^{C} integral en route to L. However this elementary integral is a helpful stepping stone for beginners, avoiding the subtleties of our technical theorems. Even if we by-pass the L^{C} integral we shall obviously still require the FTC which we proved in that context as a corollary of the Indefinite Integral Theorem. The latter was obtained in a convoluted way. We conclude this chapter with a slightly strengthened version, relying only on the Basic Properties and so independent of our earlier treatment. The FTC, as previously stated, is a corollary of this.

12.15 Indefinite Integral Theorem II. Let $g \in \mathrm{L}_{\mathbb{R}}$ and define

$$G(x) := \int_{-\infty}^{x} g.$$

Then, for each point c at which g is continuous, $G'(c)$ exists and equals $g(c)$.

Proof. For $h \neq 0$ let I_h be the closed interval with endpoints c and $c+h$.

$$\begin{aligned}
\left|\frac{G(c+h)-G(c)}{h} - g(c)\right| &= \left|\frac{1}{h}\Big(\int_{-\infty}^{c+h} g(y)\,\mathrm{d}y - \int_{-\infty}^{c} g(y)\,\mathrm{d}y\Big) - g(c)\right| \\
&= \left|\Big(\frac{1}{|h|}\int_{I_h} g(y)\,\mathrm{d}y\Big) - g(c)\right| && \text{(see 12.10)} \\
&= \frac{1}{|h|}\left|\int_{I_h} (g(y)-g(c))\,\mathrm{d}y\right| && \text{(by (L))} \\
&\leqslant \frac{1}{|h|}\int_{I_h} |g(y)-g(c)|\,\mathrm{d}y && \text{(by (M))} \\
&\leqslant \frac{1}{|h|}\,|h| \sup\{\,|g(y)-g(c)| \mid y \in I_h\,\} && \text{(by (P))} \\
&\to 0 \quad \text{as } h \to 0 && \text{(by continuity of } g \text{ at } c). \qquad \square
\end{aligned}$$

Exercises

This would be an opportune point at which to review the exercises from Chapter 1, and to try those not yet covered in the text. We set no other exercises here, encouraging the reader not to linger over the construction of the integral, but to proceed with all speed to the succeeding chapters where its virtues are revealed.

13 Non-integrable functions

Our next task is to show the limitations of our theory, by investigating functions which are not integrable.

In order to be integrable a function f must be neither 'too big' nor 'too wild'. In this chapter we explore the first of these imprecise terms, deferring consideration of the second until Chapters 21 and 22. Informally, a non-negative function is 'too big' if the area under its graph is infinite. Because we demand that Property (M) holds, $f \notin \mathrm{L}$ whenever the area under the graph of $|f|$ is infinite. This may happen because the area is 'too big horizontally', as occurs for $\chi_{\mathbb{R}}$, or 'too big vertically', as with x^{-1} on $(0,1]$. To show that a given function cannot be integrable we always argue by contradiction. This allows us to write down $\int |f|$ (or $\int f$ if $f \geqslant 0$), and to apply our Basic Properties. We present a clutch of examples, including the functions mentioned above. If you worked through the exercises for Chapter 1 you should already have discovered the right strategy.

13.1 Examples.

(1) We show $\chi_{\mathbb{R}}$ is not integrable. Note that $\chi_{\mathbb{R}} \notin \mathrm{L}^{\mathrm{step}}$ because $\mathbb{R}$ is not bounded. We truncate, and consider the (integrable!) step functions $g_n := \chi_{[0,n]}$. We have $g_n \leqslant \chi_{\mathbb{R}}$ for each n (see Fig. 13.1). Now suppose for a contradiction that $\chi_{\mathbb{R}}$ were integrable. By Property (P),

$$n = \int g_n \leqslant \int f \quad \text{for all } n.$$

This is impossible because $\int f$ is finite, and we have the required contradiction. (Note that this example shows that it is possible to have $f\chi_I$ integrable for every bounded interval I but f not integrable.)

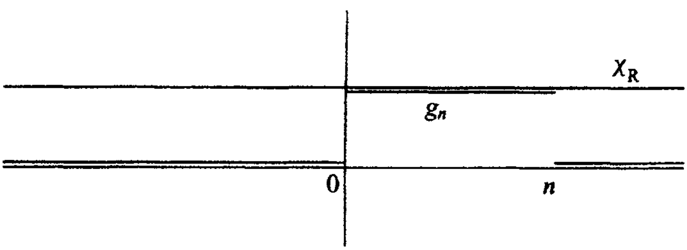


Figure 13.1

(2) We claim $\chi_{\mathbb{R}\setminus\mathbb{Q}}$ is not integrable. Since $\chi_{\mathbb{R}\setminus\mathbb{Q}} = \chi_{\mathbb{R}}$ except on the null set $\mathbb{Q}$, if $\chi_{\mathbb{R}\setminus\mathbb{Q}}$ were integrable, $\chi_{\mathbb{R}}$ would be too.

(3) Let $f(x) := x^{-1}\chi_{(0,1]}(x)$. Then f is not integrable. Despite being continuous, the function f is not in L^{C} because $(0,1]$ is not closed. Take $g_n(x) = x^{-1}\chi_{[1/n,1]}(x)$. Then $g_n \in \mathrm{L}^{\mathrm{C}}$ and $g_n \leqslant f$ for all n. By the FTC and Property (P) we have

$$\log n = \int g_n \leqslant \int f \quad \text{for all } n,$$

which is impossible as $\log n \to \infty$ as $n \to \infty$, by 5.12.

(4) Let $f(x) = \sin x$ on $\mathbb{R}$. Then f is not integrable. Because we do not have $\sin x \geqslant 0$ for all x, we exploit Property (M): if $\sin x$ were integrable then $|\sin x|$ would be too. So suppose for a contradiction that $|\sin x|$ is integrable. Let $g_n(x) := |\sin x|\,\chi_{[0,n\pi]}(x)$; see Fig. 13.2. Then $g_n(x) = \sin(x - k\pi)$ for $x \in [k\pi, (k+1)\pi]$ $(k = 0, \ldots, n-1)$ and $g_n(x) \leqslant |\sin x|$. Hence

$$2n = n\int_0^{\pi} \sin x \,\mathrm{d}x = \int_0^{n\pi} |\sin x| \,\mathrm{d}x \leqslant \int |f| \quad \text{for all } n.$$

We have the required contradiction.

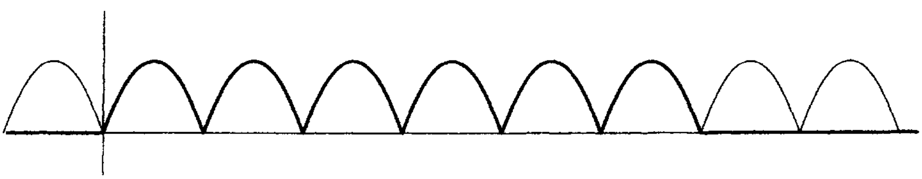


Figure 13.2

With the experience gained from these examples we now present a general strategy for showing a given function is not integrable.

13.2 A sufficient condition for non-integrability. Let $f\colon \mathbb{R} \to \mathbb{R}$. Suppose we can construct a sequence $\{g_n\}$ of functions such that

(i) g_n is integrable,
(ii) $g_n \leqslant |f|$ for all n, and
(iii) $\int g_n \to \infty$ as $n \to \infty$.

Then f is not integrable.

Remarks. In applications, g_n usually belongs to $\mathrm{L}^{\mathrm{step}}$ or L^{C}, but this is not necessary. Usually, but not necessarily, g_n is a truncation of $|f|$, chosen to achieve (i). Usually, but irrelevantly, the sequence $\{g_n\}$ is increasing. It would be possible always to seek $\{g_n\} \subseteq \mathrm{L}^{\mathrm{step}}$. However except in simple cases this leads to fiddly calculations; see Exercise 13.1.

Proof. We assume for a contradiction that $|f|$ is integrable (we work with f itself if $f \geqslant 0$). By (i) and Property (P), $\int g_n \leqslant \int |f|$ for all n, which is incompatible with (iii). We deduce that $|f|$ is not integrable, whence f cannot be, by Property (M). $\square$

Extending 13.1(3), we have 13.3, which we complement with positive results in 15.1. An alternative proof of (b), via step functions, can be extracted from 7.14, provided we prove independently of the Integral Test (7.14) that $\sum k^p$ diverges for $p \geqslant -1$.

13.3 Non-integrability: powers. Let $p \in \mathbb{R}$.
(a) x^p is not integrable on $(0,1]$ if $p \leqslant -1$.
(b) x^p is not integrable on $[1,\infty)$ if $p \geqslant -1$.
(c) x^p is not integrable on $(0,\infty)$ for any p.
(In (a) and (b) the endpoint 1 may be replaced by any $c \in (0,\infty)$.)
Proof. For (a) we apply 13.2 with $g_n(x) = x^p \chi_{[1/n,1]}(x)$. The case $p = -1$ has already been considered. For $p < -1$, we have

$$\int_{1/n}^{1} x^p \, dx = \frac{1}{p+1}(1 - n^{-(p+1)}) \to \infty \quad \text{as } n \to \infty.$$

We treat (b) similarly, with $g_n(x) := x^p \chi_{[1,n]}(x)$. Part (c) follows from (a) and (b); see 12.10. $\square$

13.4 Exercise example. Prove that the following functions are not integrable:

(i) $(x-1)^{-2}$ on $[1,\infty)$, (ii) $1/(x \log x)$ on $[2,\infty)$,
(iii) $\chi_{(-\infty,0]} - \chi_{[0,\infty)}$, (iv) $\cos^3 x$.

The manipulations in the next two sections look rather formidable, but involve only results from elementary calculus.

13.5 Non-integrable functions which can be integrated. We have already seen how the periodicity of the function $\sin x$ prevents it from being integrable (13.1(4)). We now consider a more subtle example: let

$$f(x) = \frac{\sin x}{x} \text{ for } x \in (0,\infty), \quad f(0) = 1.$$

The graphs of f and $|f|$ are shown in Fig. 13.3. Because f is continuous on $\mathbb{R}^+$, the restriction of f to every interval $[0,X]$ $(0 < X < \infty)$ is integrable. We shall show that

(i) f is not integrable (because $|f|$ isn't), but nevertheless

(ii) $\lim_{X\to\infty} \int_0^X f(x)\,dx$ exists.

What this says, informally, is that the area under the graph of $|f|$ is infinite (each of the humps makes a positive contribution and the sum of these contributions does not converge), whereas, because of the alternation of sign of f, (ii) does hold.

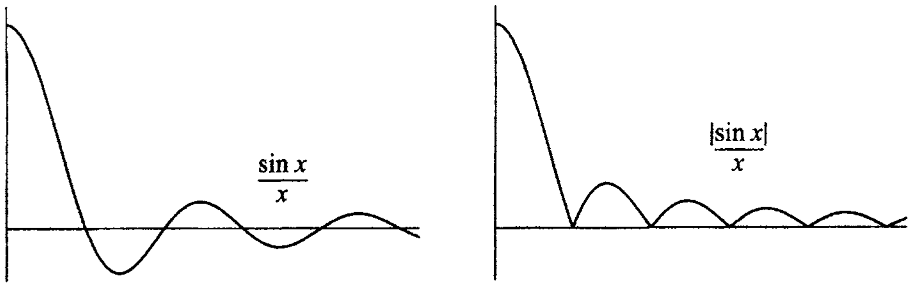


Figure 13.3

One strategy for proving (i) was given in Exercise 1.7. A simple proof for (ii) can be found in Exercise 16.5. Alternatively, it is an easy exercise in complex analysis to show that the limit in (ii) exists and equals $\pi/2$; see [**13**], 8.5.

Here we present a method which gives (i) and (ii) together. Define

$$I_k := \int_{k\pi}^{(k+1)\pi} \frac{\sin x}{x}\,\mathrm{d}x \qquad \text{and} \qquad J_k := \int_{k\pi}^{(k+1)\pi} \left|\frac{\sin x}{x}\right| \mathrm{d}x.$$

We use the change of variable $y = x - k\pi$ to obtain

$$I_k = \int_0^\pi (-1)^k \frac{\sin y}{y + k\pi}\,\mathrm{d}y \quad \text{and} \quad J_k = \int_0^\pi \frac{\sin y}{y+k\pi}\,\mathrm{d}y.$$

Note that $k\pi \leqslant y + k\pi \leqslant (k+1)\pi$ for $y \in [0, \pi]$. By 5.2,

$$J_k \geqslant \frac{1}{(k+1)\pi} \int_0^\pi \sin y\,\mathrm{d}y = \frac{2}{(k+1)\pi},$$

as $\int_0^\pi \sin y\,\mathrm{d}y = 2$. Hence $\sum J_k$ diverges (to infinity) by the Comparison Test. But if $|f|$ were integrable we would have

$$\sum_{k=1}^{n} J_k \leqslant \int |f| \quad \text{for all } n.$$

Hence (i) holds.

For (ii) we first show that $\sum I_k$ converges. The series has terms of alternating sign, so we apply Leibnitz' test. We have $|I_{k+1}| \leqslant |I_k|$ and $|I_k| \leqslant 2/(k\pi)$ (check the details) and hence $\sum I_k$ converges. Therefore

$$\lim_{n\to\infty} \int_0^{n\pi} f(x)\,\mathrm{d}x = K := \sum_{k=1}^{\infty} I_k,$$

We have claimed a little more than this in (ii), since there X is any positive real number. The difference is that $\int_0^{n\pi} f(x)\,\mathrm{d}x$ represents the sum of the areas of the first n humps of f, while $\int_0^X f(x)\,dx$ represents the area under the graph between 0 and X, which will in general consist of a certain number n of complete humps plus part of a further hump (between $n\pi$ and X). To finish the proof, assume that n is such that $n\pi \leqslant X < (n+1)\pi$. Then

$$\int_0^X f(x)\,dx = \int_0^{n\pi} f(x)\,dx + \int_{n\pi}^X f(x)\,dx.$$

As $X \to \infty$, $n \to \infty$ too, so the first term on the right-hand side tends to K. The modulus of the second term is not more than $|I_{n+1}|$, the area of the $(n+1)$th hump, and we have already shown this tends to 0. Hence (ii) holds.

We could 'define' the integral of f on $(0,\infty)$ to be the limit in (ii). There is nothing contradictory about this. It just means that we have moved outside the scope of the Lebesgue theory. The existence of functions such as f, which 'have integrals' but are not Lebesgue integrable is the price we pay for working with a theory in which Property (M) holds. We contend that it is a very small price.

13.6 Exercise example. Let $f(x) = x^{-1}\cos x$ on $[\pi/2,\infty)$. Prove that f is not integrable by considering the integral of the modulus over the interval $[\pi/2,(2n+1)\pi/2)]$, broken down into subintervals

$$[(2k-1)\pi/2,(2k+1)\pi/2] \quad (k = 1,\ldots,n)$$

on which f is of constant sign. Show further that $\lim_{X\to\infty}\int_0^X f(x)\,dx$ exists.

13.7 A non-integrable derivative. The FTC, 5.5, tells us that $\int_a^b f' = f(b) - f(a)$ so long as f' is continuous on $[a,b]$. It is imperative that the interval involved here is compact [though in 16.3 we do obtain extensions of the FTC to intervals which are not compact, under additional assumptions on f]. Consider

$$f(x) = x^2\sin(x^{-2})\chi_{(0,1]}(x).$$

Certainly f itself is integrable, since by defining $f(0) = 0$ we obtain a function in $C[0,1]$. We claim that f' is not integrable. On $(0,1]$,

$$f'(x) = 2x\sin(x^{-2}) - 2x^{-1}\cos(x^{-2}).$$

The first term on the right-hand side tends to 0 as $x \to 0$, so equals a.e. a function in L^C and thus is integrable. Therefore f' is integrable if and only if g is, where $g(x) := x^{-1}\cos(x^{-2})$ on $(0,1]$. Assume for a contradiction that g is integrable, whence $|g|$ would be too. The idea is to change the variable in $\int |g|$, writing $u = x^{-2}$ and thereby transforming to an integral $\int_1^\infty |u^{-1}\cos u|\,du$, which 13.6 tells us does not exist as a Lebesgue integral. Unfortunately the moduli get in the way, preventing us using Theorem 6.6 directly. However we can argue more circumspectly as follows. For any integer k,

$$\begin{aligned}
I_k &:= \int_{\sqrt{2/(2k+1)\pi}}^{\sqrt{2/(2k-1)\pi}} \left|x\cos(x^{-2})\right| \mathrm{d}x \\
&= (-1)^k \int_{\sqrt{2/(2k+1)\pi}}^{\sqrt{2/(2k-1)\pi}} x\cos(x^{-2})\,\mathrm{d}x \\
&= \frac{(-1)^k}{2}\int_{\frac{(2k-1)\pi}{2}}^{\frac{(2k+1)\pi}{2}} \frac{\cos u}{\sqrt{u}}\,\mathrm{d}u && \text{(by 6.6, putting } u = x^{-2}\text{)} \\
&= \frac{1}{2}\int_0^{\pi} \frac{\sin y}{\sqrt{y+(2k-1)\pi/2}}\,\mathrm{d}y && \text{(by 6.6, putting } y = u - \tfrac{1}{2}(2k-1)\text{)} \\
&\geqslant \frac{\pi}{\sqrt{(2k+1)\pi/2}}.
\end{aligned}$$

Hence $\sum I_k$ diverges by comparison with $\sum 1/\sqrt{k}$. But $\int |g| \geqslant \sum_{k=1}^{n} I_k$, so that $|g|$ cannot be integrable.

Exercises

13.1 Prove, using suitable $\mathrm{L^C}$ truncations, that the following are not integrable.

$$\text{(i) } (x^2-1)^{-1}\chi_{(-1,1)}(x), \qquad \text{(ii) } (1-x^2)^{-1}\chi_{[0,1]}(x),$$
$$\text{(iii) } x^{-1}\mathrm{e}^{-x}\chi_{(0,1]}(x), \qquad \text{(iv) } \frac{1}{x(1+x^2)}.$$

For comparison, find for one of these an alternative argument using step functions.

13.2 By making a suitable substitution in the integral $\int_0^a \tan x\,\mathrm{d}x$, where $0 < a < \pi/2$, and arguing by contradiction prove that $\tan x \notin \mathrm{L}[0,\pi/2]$.

13.3 Prove that x^{-2} is not integrable on $[-1,1]$. [Cf. Exercise 5.8.]

13.4 (a) Prove that $\sum_{k=1}^{\infty} k\chi_{(1/(k+1),1/k]}$ is not integrable.

(b) Prove that $\sum_{k=1}^{\infty} \sqrt{k}\chi_{(1/(k+1),1/k]}$ belongs to $\mathrm{L^{inc}}$.

(c) Deduce that the product of two $\mathrm{L^{inc}}$ functions need not be integrable. [Recall Exercise 11.3.]

13.5 Let f be a function which is continuous, not identically zero, and periodic of period $T > 0$.

(a) Prove that f cannot be integrable.

(b) Let $g(x) := x^{-1}f(x)$. Prove that g cannot be integrable.

13.6 Let $f(x) = \sin(x^2)$.

(a) Sketch the graph of f.

(b) Prove that f is not integrable.

(c) Prove that $\lim_{X\to\infty} \int_0^X f(x)\,\mathrm{d}x$ exists.

14 Convergence Theorems: MCT and DCT

The convergence theorems form the backbone of the Lebesgue theory. Both the Monotone Convergence Theorem (MCT) and the Dominated Convergence Theorem (DCT) have the same conclusion:

$$\lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n$$

—valid under different assumptions on the sequence $\{f_n\}$ in the two cases.

The proof of the MCT is technical and you may wish to skip over it at a first reading to the applications in the next chapter. The DCT is derived from the MCT. Its proof is a good theoretical illustration of the use of the MCT. Before embarking on the theory we indicate why the convergence theorems are needed.

14.1 Integration and limiting processes. We already know that finite processes and integrals interact well. Suppose that $u_1, u_2, \ldots$ are real-valued and integrable. Property (L) implies that, for any $n = 1, 2, \ldots$,

$$\int (u_1 + \cdots + u_n) = \int u_1 + \cdots + \int u_n,$$

that is, $\int \sum_{k=1}^{n} u_k = \sum_{k=1}^{n} \int u_k$. Also, $u_1 \vee \cdots \vee u_n$ and $u_1 \wedge \cdots \wedge u_n$ are integrable. However, as soon as we consider infinite sums and infinite sups and infs then limits are involved:

$$\int \sum_{k=1}^{\infty} u_k = \int \lim_{n\to\infty} \sum_{k=1}^{n} u_k \quad \text{and} \quad \int \sup_{k\geqslant 1} u_k = \int \lim_{n\to\infty} (u_1 \vee \cdots \vee u_n).$$

Truncation processes also bring in limits: for $f\colon \mathbb{R} \to \mathbb{R}$, $f = \lim_{n\to\infty} f\chi_{[-n,n]}$ and $f = \lim_{n\to\infty}(f \wedge n)$, where $(f \wedge n)(x) := \min\{f(x), n\}$. Is it true that if f is integrable then

$$\int f = \lim_{n\to\infty} \int f\chi_{[-n,n]} = \lim_{n\to\infty} \int_{-n}^{n} f \quad \text{and} \quad \int f = \lim_{n\to\infty} \int (f \wedge n)\,?$$

(The first of these is **not** the definition of $\int f$; recall 13.5.) By giving sufficient conditions for $\lim \int f_n = \int \lim f_n$, the MCT and DCT allow us to handle the many mathematical processes come under the umbrella of limits: besides those mentioned above, integration itself and differentiation have limits in their definitions.

14.2 Guilty until proved innocent. Here is some bad news:

- $\lim_{n\to\infty} \int n^2\chi_{(0,1/n)} = \infty \neq 0 = \int \lim_{n\to\infty} n^2\chi_{(0,1/n)}$ (8.2(4));

- $\displaystyle\int_0^1 \sum_{k=1}^{\infty} (\mathrm{e}^{-kx} - 2\mathrm{e}^{-2kx})\,\mathrm{d}x \neq \sum_{k=1}^{\infty} \int_0^1 (\mathrm{e}^{-kx} - 2\mathrm{e}^{-2kx})\,\mathrm{d}x$ (Exercise 17.3);
- $\displaystyle\frac{\mathrm{d}}{\mathrm{d}t}\int f(x,t)\,\mathrm{d}x \neq \int \frac{\partial}{\partial t} f(x,t)\,\mathrm{d}x$ sometimes (see Exercise 20.3);
- $\displaystyle\int_0^1 \left\{ \int_0^1 (x-y)(x+y)^{-3}\mathrm{d}x \right\} \mathrm{d}y \neq \int_0^1 \left\{ \int_0^1 (x-y)(x+y)^{-3}\mathrm{d}y \right\} \mathrm{d}x \quad (26.3).$

These conclusions are based on direct calculations. In every one of these examples some limiting process is involved, and what is happening is that the limit and the integral do not commute. Moral: beware tacitly assuming that limits and integrals can be taken in either order and avoid surreptitiously sneaking a limit past an integral.

The following cautionary tale may help to reinforce the point. A traveller to Ruritania must present to the border-guard at the frontier a passport stamped with a valid Ruritanian visa. Visas are not issued automatically and only day-trippers are exempt from the visa requirement. The authorities deal with Illegal entrants in a highly unpleasant manner, and ignorance of the regulations is not accepted as grounds for leniency.

Suppose we are given a sequence $\{f_n\}$ of functions, and wish to write

$$\lim \int f_n = \int \lim f_n,$$

that is, we wish to take the limit past the integral sign. Think of the limit as a traveller and the integral sign as a Ruritanian border-guard. Either the MCT or the DCT can serve as a passport, and a check that the particular sequence $\{f_n\}$ meets the conditions for one of these theorems acts as its visa. Only when a valid check has been presented may the limit pass the integral sign. We shall later supply passport theorems for other limiting processes (infinite sums, integrals, differentiation, ...). Visas are not granted indiscriminately. Each of our 'bad news' examples above is an instance of refusal. Finite processes qualify as day-trippers and may pass unchecked.

We shall prove the MCT by showing successively that the theorem is true for a sequence of functions in $\mathrm{L}^{\mathrm{step}}$, in $\mathrm{L}^{\mathrm{inc}}$, and in $\mathrm{L}_{\mathbb{R}}$. The argument takes off from Technical Theorem I, 9.9, which we can now recast as follows.

14.3 Technical Theorem I (restated). Let $\{\psi_n\}$ be an increasing sequence of step functions such that $\int \psi_n \leqslant K$ for some finite constant K. Then there exists a real-valued function $g \in \mathrm{L}^{\mathrm{inc}}$ such that $g(x) := \lim \psi_n(x)$ for almost all x, and $\int g := \lim \int \psi_n$.

14.4 The Monotone Convergence Theorem (MCT). Let $\{f_n\}$ be a sequence of functions $f_n : \mathbb{R} \to \mathbb{R}$ such that

(M1) f_n is integrable for each n,

(M2) $f_n \leqslant f_{n+1}$ for all n,

(M3) there exists a finite constant K independent of n such that $\int f_n \leqslant K$ for all n.

Then there exists $g\colon \mathbb{R} \to \mathbb{R}$ such that $f_n \xrightarrow{\text{a.e.}} g$, g is integrable, and

$$\int g = \int \lim f_n = \lim \int f_n.$$

(As in the $\mathrm{L}^{\mathrm{inc}}$ definition we could replace (M3) by

(M3′) the sequence $\{\int f_n\}$ converges.

In practice (M3) is preferable.)

Proof. Stage 1. $\{f_n\} \subseteq \mathrm{L}^{\mathrm{step}}$.

Technical Theorem I, as in 14.3, is exactly the MCT for a sequence of step functions.

Stage 2. $\{f_n\} \subseteq \mathrm{L}^{\mathrm{inc}}$.

We must show that $\lim f_n$ is finite a.e. and that $\int \lim f_n$ exists and equals $\lim \int f_n$. Let $\{\varphi_{nm}\}_{m \geqslant 1}$ be an $\mathrm{L}^{\mathrm{inc}}$-sequence for f_n, converging to f_n off the null set E_n. Define, for each n,

$$\psi_n := \max\{\, \varphi_{ij} \mid 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant n \,\}.$$

We first prove that $\{\psi_n\}$ satisfies the conditions for Technical Theorem I, 14.3, from which it will follow that there exists $g\colon \mathbb{R} \to \mathbb{R}$ such that $g \in \mathrm{L}^{\mathrm{inc}}$ and $g = \lim \psi_n$ except on some null set, E_0 say. Certainly $\psi_n \in \mathrm{L}^{\mathrm{step}}$ and $\psi_n \leqslant \psi_{n+1}$. Also, for any k,

$$\varphi_{km} \leqslant f_k \quad \text{for all } m,$$

whence, taking the maximum over $k, m \leqslant n$, we have

$$\psi_n \leqslant f_n \quad \text{for all } n.$$

Thus, by (M3),

$$\int \psi_n \leqslant \int f_n \leqslant K \quad \text{for all } n.$$

Hence $\{\psi_n\}$ does indeed satisfy the conditions for 14.3.

To complete Stage 2 we must show that $g = \lim f_n$ a.e., and that $\lim \int \varphi_n = \lim \int f_n$. The obvious candidate for the exceptional set is $E := \bigcup_{n \geqslant 0} E_n$ (which is null by 10.6). Let $x \notin E$. For any n,

$$\varphi_{nm}(x) \leqslant \psi_m(x) \leqslant g(x) \quad \text{for all } m \geqslant n,$$

so, keeping n fixed and letting $m \to \infty$, we have

$$f_n(x) \leqslant g(x).$$

Also, from above, $\psi_n(x) \leqslant f_n(x)$. Hence, by a sandwiching argument, $g(x) = \lim \psi_n(x) = \lim f_n(x)$. In addition, $\psi_n \leqslant f_n \leqslant g$ a.e. implies $\int \psi_n \leqslant \int f_n \leqslant \int g$, so again by sandwiching, $\lim \int f_n = \int \lim f_n$.

Stage 3. $\{f_n\} \subseteq \mathrm{L}_{\mathbf{R}}$.

We can write $f_n = s_n - t_n$, where $s_n, t_n \in \mathrm{L}^{\mathrm{inc}}$. Our initial idea is to apply Stage 2 to $\{s_n\}$ and $\{t_n\}$. However this may not be possible: these sequences may not be monotonic and the sequences $\{\int s_n\}$, $\{\int t_n\}$ may not be bounded.

Fortunately Proposition 12.13(b) gives us some control over the decomposition of an integrable function as the difference of $\mathrm{L}^{\mathrm{inc}}$ functions. Write $f_1 = s_1 - t_1$, $s_1, t_1 \in \mathrm{L}^{\mathrm{inc}}$, in any manner. To facilitate the presentation of the proof we now express f_n (for $n \geqslant 2$) in the form

$$f_n = f_1 + (f_2 - f_1) + \cdots + (f_n - f_{n-1})$$

(cf. the Telescoping Lemma, 2.19). Use 12.13(b) to choose $p_k, q_k \in \mathrm{L}^{\mathrm{inc}}$ such that

$$f_k - f_{k-1} = p_k - q_k, \quad q_k \geqslant 0 \quad \text{and} \quad \int q_k < 2^{-k+1} \quad (k \geqslant 2).$$

Define, for $n \geqslant 2$,

$$s_n := s_1 + p_2 + \cdots + p_n \quad \text{and} \quad t_n := t_1 + q_2 + \cdots + q_n.$$

Certainly s_n and t_n belong to $\mathrm{L}^{\mathrm{inc}}$. We claim that $\{s_n\}$ and $\{t_n\}$ satisfy the MCT conditions. By construction,

$$t_n - t_{n-1} = q_n \geqslant 0 \quad (\text{for } n \geqslant 2) \quad \text{and}$$
$$\int t_n \leqslant \int t_1 + 2^{-1} + \cdots + 2^{-n+1} \leqslant K' := \int t_1 + 1.$$

Also $s_n = f_n + t_n$, whence $\{s_n\}$ is increasing and $\int s_n \leqslant K + K'$ for all n. This proves our claim.

Now we can apply Stage 2 to $\{s_n\}$ and $\{t_n\}$. We conclude that $s_n \xrightarrow{\text{a.e.}} s$ and $t_n \xrightarrow{\text{a.e.}} t$, where s and t belong to $\mathrm{L}^{\mathrm{inc}}$, and $\int s = \lim \int s_n$ and $\int t = \lim \int t_n$. Finally $f_n \xrightarrow{\text{a.e.}} g := s - t$ and

$$\int g = \int s - \int t = \lim \int s_n - \lim \int t_n = \lim \left(\int s_n - \int t_n \right) = \lim \int f_n. \qquad \square$$

We often want to use the MCT to obtain information about a function which can be realized as the limit a.e. of a monotonic sequence of integrable functions. For this we need a corollary of the MCT.

14.5 Monotone Convergence Theorem (Corollary). Let $f: \mathbb{R} \to \mathbb{R}$ be given and suppose that there exists a sequence $\{f_n\}$ of functions $f_n: \mathbb{R} \to \mathbb{R}$ satisfying (M1)–(M3) and

(M4) $f_n \xrightarrow{\text{a.e.}} f$.

Then f is integrable and

$$\int f = \int \lim f_n = \lim \int f_n.$$

Proof. By the Monotone Convergence Theorem, (M1)–(M3) imply that there exists a real-valued function g such that $f_n \to g$ off some null set E_1 and g is integrable with $\lim \int f_n = \int g$. Since $f_n \to f$ off some null set E_2, uniqueness of limits implies that $f = g$ off the null set $E_1 \cup E_2$. The result follows from 12.9. $\square$

14.6 Remarks. Henceforth we shall refer to both 14.4 and 14.5 as the MCT, bringing in (M4) when we have a pre-existing limit.

We can clearly formulate an analogous theorem (the **downward MCT**) for decreasing sequences, replacing the universal upper bound K in (M3) by a universal lower bound.

The MCT, in its various forms, may be used in the following ways.

- To justify interchange of limit and integral: $\lim \int f_n = \int \lim f_n$.
- To establish that a given function f is integrable, by constructing a sequence $\{f_n\}$ satisfying (M1)–(M4) (see Chapter 15).
- To show that a monotonic sequence of functions must have a finite limit a.e. Our next result, which is of major theoretical importance, illustrates this nicely.

14.7 Theorem. Let f be integrable and $\int |f| = 0$. Then $f = 0$ a.e.

Proof. We apply the MCT to $\{f_n\}$, where $f_n := n\,|f|$. Certainly f_n is integrable, $f_n \leqslant f_{n+1}$ and $\int f_n = 0$ for all n. Hence by the MCT, there exists an integrable function $g\colon \mathbb{R} \to \mathbb{R}$ such that $\lim f_n = g$ a.e. Note that

$$\lim_{n\to\infty} f_n(x) = \begin{cases} \infty & \text{if } f(x) \neq 0, \\ 0 & \text{if } f(x) = 0. \end{cases}$$

Since $g(x) \in \mathbb{R}$ for every x, we must have $\{\, x \in \mathbb{R} \mid f(x) \neq 0 \,\}$ is null. $\square$

The MCT has one severe disadvantage: it applies only to monotonic sequences. We would like a similar theorem guaranteeing $\lim \int f_n = \int \lim f_n$ for a sequence $\{f_n\}$ which is not monotonic. The trick is to convert an arbitrary sequence into a monotonic one.

14.8 Monotonicity engineered. Let $\{\alpha_n\}$ be a real sequence. Define

$$\beta_n := \sup_{k\geqslant n} \alpha_k.$$

Since $A \subseteq B$ implies $\sup A \leqslant \sup B$ for non-empty subsets A, B of $\mathbb{R}$, we have

$$\beta_n \geqslant \beta_{n+1} \quad \text{for all } n.$$

By the Monotonic Sequence Theorem there exists $\beta \in \mathbb{R} \cup \{-\infty\}$ such that $\beta_n \searrow \beta := \inf \beta_n$. The limit β is denoted $\limsup \alpha_n$. So we have

$$\limsup \alpha_n := \inf \beta_n = \lim_{n\to\infty} \sup_{k\geqslant n} \alpha_k.$$

In a similar manner we may define

$$\liminf \alpha_n := \sup \gamma_n = \lim_{n\to\infty} \inf_{k\geqslant n} \alpha_k,$$

where $\gamma_n := \inf_{k\geqslant n} \alpha_k$ is such that $\gamma_n \nearrow \gamma \in \mathbb{R} \cup \{\infty\}$.

We thus have associated with our arbitrary sequence $\{\alpha_n\}$ two monotonic sequences, one decreasing, the other increasing. These monotonic sequences necessarily have limits, possibly $-\infty$, ∞. Further, we have defined a natural generalization of the ordinary limit: if the sequence $\{\alpha_n\}$ converges to $\alpha \in \mathbb{R}$, then

$$\alpha = \limsup \alpha_n = \liminf \alpha_n.$$

Conversely, if $\limsup \alpha_n = \liminf \alpha_n$ (necessarily a real number), then $\{\alpha_n\}$ converges to this value. To prove the first of these assertions, fix $\varepsilon > 0$ and choose N such that $k \geqslant N$ implies $\alpha - \varepsilon < \alpha_k < \alpha + \varepsilon$. Take $n \geqslant N$. Then

$$\alpha - \varepsilon < \beta_n \leqslant \alpha + \varepsilon,$$

from which we deduce that $\beta_n \to \alpha$. Thus $\limsup \alpha_n = \alpha$, and $\liminf \alpha_n = \alpha$ similarly. In the other direction, assume $\alpha := \limsup \alpha_n = \liminf \alpha_n$. Then, given $\varepsilon > 0$, we can find N such that $n \geqslant N$ implies

$$\alpha - \varepsilon < \inf_{k \geqslant n} \alpha_k \leqslant \alpha_n \leqslant \sup_{k \geqslant n} \alpha_k \leqslant \alpha + \varepsilon,$$

which shows that $\alpha_n \to \alpha$.

All this extends pointwise to functions in the usual way.

14.9 The Dominated Convergence Theorem (DCT). Let $\{f_n\}$ be a sequence of real-valued functions on $\mathbb{R}$ such that

(D1) f_n is integrable for each n;

(D2) there exists $f\colon \mathbb{R} \to \mathbb{R}$ such that $f_n \to f$ a.e.;

(D3) there exists a function G which is integrable and independent of n such that $|f_n| \leqslant G$ for all n.

Then f is integrable and

$$\int f = \int \lim f_n = \lim \int f_n.$$

(The function G in (D3) is known as a *dominating function.* If the functions f_n are continuous then we do not need to check condition (D1) for each individual f_n, since this follows for all n at once from (D3), by the Comparison Theorem as stated in 15.6. [Later, when we have the definitive Comparison Theorem, we shall see that (D1) can be replaced by

(D1′) f_n is measurable for each n.

In 21.7 we extend the DCT to cover the case that the functions f_n belong to L rather than to $\mathrm{L}_{\mathbb{R}}$.])

Proof. The proof consists of two applications of the MCT, one nested inside the other.

MCT application 1. Let $p_n := \inf_{k \geqslant n} f_k$.

(M1) We claim that p_n is integrable. We have

$$p_n = \inf\{f_n, f_{n+1}, \dots\}.$$

Writing p_n this way, we recognize it as being obtained as a limit, so that integrability of the functions f_k does not immediately imply that p_n is integrable. More precisely we have

$$p_n = \lim_{m\to\infty} p_{nm}, \quad \text{where } p_{nm} := f_n \wedge \cdots \wedge f_{n+m}.$$

MCT application 2. With n fixed, we apply the downward MCT to $\{p_{nm}\}_{m\geqslant 1}$:—

(M1) Since each f_k is integrable, p_{nm} is integrable for each m, by 12.7.

(M2) We have $p_{nm} \geqslant p_{n(m+1)}$ for any m (more functions, smaller inf).

(M3) For any m, we have $p_{nm} \geqslant -G$, so that $\int p_{nm} \geqslant -\int G$, which is a constant independent of m.

(M4) $p_{nm} \to p_n$ as $n \to \infty$.

Thus the downward MCT implies that p_n is integrable.

(M2) By construction, $p_n \leqslant p_{n+1}$ (fewer functions, bigger inf).

(M3) Since $p_n \leqslant f_n$ we have $p_n \leqslant G$ for all n, whence $\int p_n \leqslant K := \int G$, independent of n.

(M4) Since $f_n \to f$ a.e., $p_n \to f$ a.e. too (see 14.8).

Thus, by the (upward) MCT, f is integrable and

$$\int f = \int \lim p_n = \lim \int p_n.$$

In the same way,

$$\int f = \int \lim q_n = \lim \int q_n,$$

where $q_n := \sup_{k\geqslant n} f_k$. We complete the proof by a sandwiching argument. We have $p_n \leqslant f_n \leqslant q_n$ for each n. Thus $\int p_n \leqslant \int f_n \leqslant \int q_n$. As $n \to \infty$ the two outer terms tend to $\int f$. Hence $\int f_n \to \int f$ too. □

14.10 The DCT on bounded intervals. Suppose that $\{f_n\}$ is a sequence of integrable functions on a bounded interval I and that $f_n \to f$ a.e. on I. Suppose further that the functions f_n are *uniformly bounded*, that is,

$$(\forall n)\ |f_n| \leqslant K \text{ on } I \qquad \text{for some constant } K.$$

Then $G := K\chi_I$ is integrable on I (because I is *bounded*). Therefore we may take G as the dominating function in the DCT. Hence the DCT applies to $\{f_n\}$. This special case of the DCT is known as the **Bounded Convergence Theorem**. It applies in particular when $\{f_n\}$ is uniformly convergent on I (Exercise 14.4), and so extends the elementary result proved in 8.9.

Exercises

The exercises for this chapter are mostly theoretical. You may wish to defer them until after you have studied Chapters 15 and 16.

14.1 (a) Show that the conclusion of the MCT remains valid if condition (M1) is weakened to the assumption that, for each n, $f_n \leqslant f_{n+1}$ a.e.

(b) Show that the conclusion of the DCT remains valid if condition (D3) is weakened to the assumption that, for each n, $|f_n| \leqslant G$ a.e.

14.2 (a) Let $f_n = n^{-1}\chi_{(0,n)}$. Prove that $f_n \xrightarrow{\text{u}} 0$ on $\mathbb{R}$ and that $\lim \int f_n \neq \int \lim f_n$. Why does this contradict neither the MCT nor the DCT?

(b) Let $f_n(x) = n^{-1}(1 - n^{-1}|x|)\chi_{[-n,n]}(x)$. Prove that $f_n \xrightarrow{\text{u}} 0$ on $\mathbb{R}$ and that $\lim \int f_n \neq \int \lim f_n$. Why does this contradict neither the MCT nor the DCT?

14.3 Sketch the graph of the function f defined on $(0,1]$ by

$$f(x) = (k+1)((k+1)x - 1) \quad \text{if } x \in [1/(k+1), 1/k) \qquad (k = 1, 2, \dots).$$

Apply the MCT to a suitable sequence to show that f is integrable.

14.4 Let $\{f_n\}$ be a sequence of integrable functions on a bounded interval I such that $f_n \xrightarrow{\text{u}} f$ on I. Use the Bounded Convergence Theorem to show that f is integrable on I and $\lim \int_I f_n = \int_I f$.

14.5 (a) Assume that $f\colon \mathbb{R} \to \mathbb{R}$ is integrable.

(i) Prove that $f \wedge n$ is integrable for each fixed n, and

(ii) prove that $\lim_{n\to\infty} \int (f \wedge n) = \int f$.

[Hint for (i): express $n\chi_{\mathbb{R}}$ as the limit of a suitable sequence of step functions.]

(b) Assume that $f\colon \mathbb{R} \to \mathbb{R}$ is such that $f \wedge n$ is integrable for each n. Prove that f is integrable if and only if $\{\int (f \wedge n)\}$ is bounded.

14.6 Let $\{f_n\}$ be a sequence of non-negative integrable functions such that $f_n \xrightarrow{\text{a.e.}} f$, where f is integrable.

(a) Let $p_n := \inf_{k \geqslant n} f_k$. Show that $\int f = \lim \int p_n$.

(b) By writing $f - f_n$ as $f - p_n + p_n - f_n$ prove that $\lim \int |f - f_n| = 0$.

14.7 Let $f \in \mathrm{C}[0,1]$. Prove that

$$\int_0^1 f = \lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} f\left(\frac{r}{n}\right).$$

Give an example to show that this may not be true if it is merely assumed that f is integrable on $[0,1]$. [Cf. 9.1 and 9.2.]

14.8 Calculate $\limsup \alpha_n$ and $\liminf \alpha_n$ for the following sequences $\{\alpha_n\}$:

$$\text{(i) } \alpha_n = \sin(\pi n/2), \quad \text{(ii) } \alpha_n = \begin{cases} n & \text{if } n \text{ is prime,} \\ 0 & \text{otherwise,} \end{cases} \quad \text{(iii) } \alpha_n = (-1)^n n.$$

15 Recognizing integrable functions I

Our catalogue of integrable functions so far is limited. We have

- every step function $\varphi = \sum_{r=1}^{n} a_r \chi_{I_r}$ is integrable, and, by definition,

$$\int \varphi = \sum_{r=1}^{n} a_r \ell(I_r);$$

- every $f \in \mathrm{C}[a, b]$ is integrable, and, if we can write $f = F'$, then, by the FTC (see 11.2 and 6.1),

$$\int_a^b f = F(b) - F(a);$$

- every f which is bounded and continuous a.e. on a bounded interval $\langle a, b \rangle$ is integrable (see 11.2).

We have deliberately avoided giving more than a few token examples of integrable functions outside the above classes until the convergence theorems allowed us to so elegantly and efficiently. This chapter deals with simple examples. More complicated examples are given in Chapter 18.

15.1 Standard integrable functions: powers. Let $p \in \mathbb{R}$.

(a) If $p > -1$ then x^p is integrable on $(0, 1]$ and $\int_0^1 x^p \, dx = 1/(p + 1)$.

(b) If $p < -1$ then x^p is integrable on $[1, \infty)$ and $\int_1^\infty x^p \, dx = -1/(p + 1)$.

Proof. For (a) we apply the MCT to $\{f_n\}$, where $f_n(x) := x^p \chi_{[1/n, 1]}(x)$. We check the conditions.

(M1) f_n is integrable since $f_n \in \mathrm{C}[1/n, 1]$.

(M2) $f_n \leqslant f_{n+1}$ because $x^p \geqslant 0$. (Except on $[1/(n+1), 1/n)$, the functions are equal, and for $x \in [1/(n + 1), 1/n)$ we have $f_n(x) = 0 < x^p = f_{n+1}(x)$; see Fig. 15.1.)

(M3) By the FTC, $\int f_n = \int_{1/n}^1 x^p \, dx = (1 - n^{-(p+1)})/(p+1) \leqslant K := 1/(p+1)$.

(M4) As $n \to \infty$, $f_n(x) \to x^p$ for any $x \in (0, 1]$ (because $f_n(x) = x^p$ if $n \geqslant 1/x$).

The argument for (b) is essentially the same, using $f_n(x) := x^p \chi_{[1,n]}(x)$—do it! □

Combining this result with 13.3 (whose proof does not need the MCT) we have the following theorem.

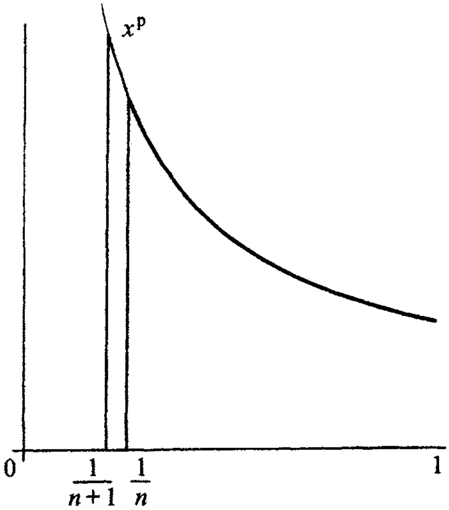


Figure 15.1

15.2 Theorem. Let $p \in \mathbb{R}$.
(a) x^p is integrable on $(0, 1]$ if and only if $p > -1$.
(b) x^p is integrable on $[1, \infty)$ if and only if $p < -1$.

15.3 Standard integrable functions: exponentials.
(a) e^{-x} is integrable on $[0, \infty)$, with $\int_0^\infty e^{-x}\,dx = 1$.
(b) $e^{-|x|}$ is integrable on $\mathbb{R}$, with $\int_{-\infty}^\infty e^{-|x|}\,dx = 2$.
(c) e^{-x^2} is integrable on $\mathbb{R}$. [The value of the integral, $\sqrt{\pi}$, is established by a variety of methods in 20.10, 26.12, and 27.6.]

Proof. We leave (a) and (b) as exercise examples (use $f_n(x) = e^{-|x|}\chi_{[-n,n]}$ in (b)), and turn to (c). We apply the MCT to $\{f_n\}$, where $f_n(x) = e^{-x^2}\chi_{[-n,n]}(x)$; see Fig. 15.2, which shows the graph of e^{-x^2}. (M1), (M2), and (M4) follow in the same way as in the preceding examples. For (M3) we note that we cannot recognize e^{-x^2} as a derivative and so cannot evaluate $\int f_n$ by the FTC. Instead we argue as follows, exploiting the fact that $e^{-x^2} \leqslant e^{-x}$ for $x \geqslant 1$:

$$\int f_n = \int_{-n}^{n} e^{-x^2}\,dx = 2\int_0^n e^{-x^2}\,dx \leqslant 2\int_0^1 e^{-x^2}\,dx + 2\int_1^n e^{-x}\,dx \leqslant 2\int_0^1 e^{-x^2}\,dx + 2.$$

Hence the MCT applies with $K := 2\int_0^1 e^{-x^2}\,dx + 2$ in (M3). $\square$

15.4 Exercise example.
(a) Prove that $(1 + x^2)^{-1}$ is integrable on $\mathbb{R}$ and that $\int_{-\infty}^\infty (1 + x^2)^{-1}\,dx = \pi$.

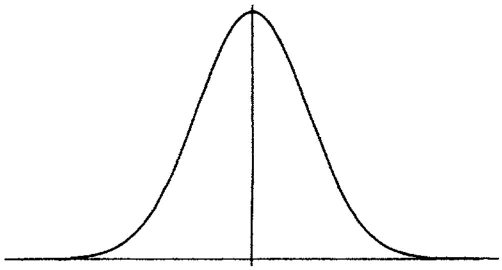


Figure 15.2

(b) Prove that xe^{-x} is integrable on $[0, \infty)$ and that $\int_0^\infty xe^{-x}\,dx = 1$.

The general strategy for using the MCT to establish integrability should now be clear. To prove the statement below, just apply the MCT to $\{f_n\}$ where $f_n := f\chi_{I_n}$. Note that (ii) makes $\{f_n\}$ increasing.

15.5 Integrability by the MCT. Let $\{I_n\}$ be a sequence of intervals with $I_n \subseteq I_{n+1}$ for all n. Let $f\colon I \to \mathbb{R}$, where $I := \bigcup_{n=1}^\infty I_n$. Assume

(i) $f\chi_{I_n}$ is integrable for each n,

(ii) $f \geqslant 0$ on I, and

(iii) there exists K independent of n such that $\int_{I_n} f \leqslant K$ for all n (equivalently, $\{\int_{I_n} f\}$ converges).

Then f is integrable on I and $\int_I f = \lim \int_{I_n} f$.

We shall not want to resort directly to the MCT every time we wish to show a function is integrable, any more than we should want to have to show a series of positive terms convergent by exhibiting an upper bound for its partial sums. We shall instead compare a given function with our standard functions. Later we shall obtain the definitive Comparison Theorem (21.3). However the preliminary version below is adequate for most practical purposes.

15.6 Comparison Theorem (simple form). Let I be an interval and assume that, on I, f is continuous and there exists an integrable function g such that $|f| \leqslant g$. Then f is integrable on I.

Proof. We define $f_n := f\chi_{I_n}$, where $\{I_n\}$ is a sequence of compact intervals such that $I = \bigcup I_n$, and apply the DCT to $\{f_n\}$.

(D1) f_n is continuous on the compact interval I_n and so integrable by 11.2.

(D2) Clearly $f_n \to f$ on I.

(D3) $|f_n| \leqslant g$ on I by construction.

Hence $f = \lim f_n$ is integrable on I, as required. □

Success in using the Comparison Theorem depends on knowing which inequalities between familiar functions are valid (and which are not). Useful inequalities are collected for reference in Appendix II.

15.7 Examples.

(1) Consider $f(x) = x^{-2}\sin x$ on $[1,\infty)$. Then f is continuous, and $|f| \leqslant g$, where $g(x) := x^{-2}$ is integrable on $[1,\infty)$. So f is integrable by the Comparison Theorem. Compare this with our earlier result that $x^{-1}\sin x \notin$ L on $[1,\infty)$ (see 13.5, Exercise 16.9, and Fig. 15.3): $x^{-2}\sin x$ decays fast enough to be integrable but $x^{-1}\sin x$ does not.

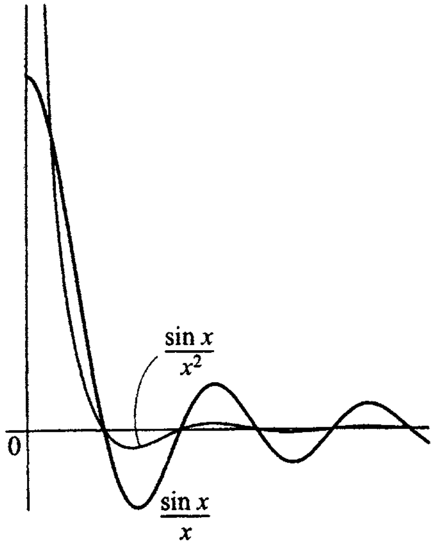


Figure 15.3

(2) Consider $f(x) = x^2(1+x^4)^{-1}$ on $(0,\infty)$. We can say that $|f(x)| \leqslant x^{-2}$ for all $x > 0$. This gives a good bound if x is large, and a very bad bound if x is close to 0. Hence we argue as follows. Because f is continuous on $[0,1]$ it is integrable on $[0,1]$ and so on $(0,1]$. Also, by the Comparison Theorem with $g(x) = x^{-2}$, f is integrable on $[1,\infty)$. Hence f is integrable on $(0,\infty)$ by 12.10.

(3) Consider $f(x) = 1/(\sqrt{x}(1+x^2))$ on $(0,\infty)$. Here we have $|f(x)| \leqslant x^{-1/2}$ on $(0,1]$ and $|f(x)| \leqslant x^{-3/2}$ on $[1,\infty)$. Hence f is integrable on each of $(0,1]$ and $[1,\infty)$ by 15.1. Therefore f is integrable on $(0,\infty)$.

(4) We show that $\log x$ is integrable on $(0,1]$. Note that Theorem 11.2(c) is not applicable because $\log x$ is unbounded. However, $\log x$ is continuous, and hence integrable, on $[\delta,1]$ for any δ with $0<\delta<1$. Further, $x^p \log x \to 0$ as $x \to 0+$ for any $p>0$. Pick $p = 1/2$. Then there exists $\delta > 0$ such that

$$|\log x| \leqslant x^{-1/2} \quad \text{if } 0 < x < \delta.$$

Hence $\log x$ is integrable on $(0, \delta]$ by 15.1 and the Comparison Theorem.

15.8 Exercise example. Use the Comparison Theorem to prove that the following are integrable:

$$\text{(i) } (1+x^6)^{-1/2}, \qquad \text{(ii) } \frac{\cos x}{\sqrt{x}(1+x)} \text{ on } [0,\infty),$$
$$\text{(iii) } e^{-x}\log x \text{ on } [1,\infty), \quad \text{(iv) } e^{-x}\log x \text{ on } (0,\infty).$$

[Hint for (iii): remember $\log x \leqslant x$ for $x \geqslant 1$.]

The preceding examples were contrived to illustrate use of the Comparison Theorem. The next examples are of some regularly occurring functions.

15.9 Examples.

(1) Let $f(x) = e^{-ax}\sin x$ on $[0,\infty)$, where $a > 0$. Certainly f is continuous on $[0,\infty)$. Also, $|f(x)| \leqslant e^{-ax}$. Since e^{-ax} is integrable on $[0,\infty)$ by 15.3, the Comparison Theorem tells us that the given function f is integrable.

(2) Generalizing 15.4(b), let $f(x) = x^p e^{-ax}$ on $[0,\infty)$ ($p \geqslant 0$, $a > 0$). We could check integrability of f directly from the MCT applied to $\{f_n\}$ where $f_n := f\chi_{[0,n]}$, but it is easy to get bogged down in detailed calculations. It is preferable to proceed as follows. Observe that, for any m,

$$|x^p e^{-ax}| = \frac{x^p}{1 + ax + (ax)^2/2! + \cdots + (ax)^m/m! + \ldots} \leqslant m!a^{-m}x^{p-m}.$$

Choosing $m > p+1$, the function on the right-hand side is integrable on $[1,\infty)$ by 15.1. By the Comparison Theorem, f is integrable on $[1,\infty)$. Since f is continuous on $[0,1]$ it is integrable on this interval too, so f is integrable on $[0,\infty)$.

15.10 Exercise example.

(a) Prove that $|x|^p e^{-|x|}$ and $|x|^p e^{-x^2}$ are integrable for any $p \geqslant 0$.

(b) Prove that $e^{-x^2}\cosh x$ is integrable.

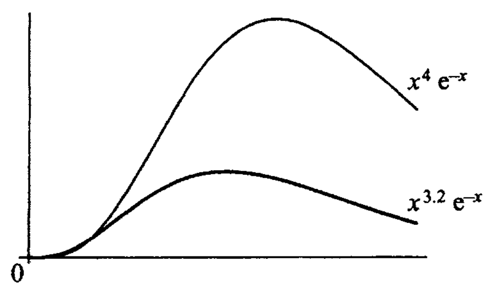


Figure 15.4

From the above examples, we derive a result worthy to be called a theorem.

15.11 Theorem. Let $a > 0$ and let $q(x) = a_0 + a_1 x + \cdots + a_n x^n$ ($n \geqslant 0$, $a_0, \ldots, a_n \in \mathbb{R}$) be any polynomial. Then
(a) $q(x)\mathrm{e}^{-ax}$ is integrable on $[0, \infty)$;
(b) $q(x)\mathrm{e}^{-a|x|}$ and $q(x)\mathrm{e}^{-ax^2}$ are integrable on $\mathbb{R}$.

The reason Theorem 15.11 works, of course, is that a negative exponential e^{-ax} ($a > 0$) decays very rapidly as $x \to \infty$, even when multiplied by any power x^p; see Fig. 15.4. We assert that in integration theory negative exponentials are a Good Thing.

Exercises

15.1 Timely reminders. Give counterexamples to the following common misconceptions. [Hint: some require a function which has not appeared already.]
   (a) An integrable function is bounded.
   (b) A continuous function on a bounded interval is integrable.
   (c) $f\chi_I \leqslant f$ ($f\colon \mathbb{R} \to \mathbb{R}$, I a subinterval of $\mathbb{R}$).
   (d) An integrable function is continuous a.e. [hint: exhibit a function which is zero a.e. but continuous nowhere].
   (e) The product of two integrable functions is integrable.
   (f) f integrable implies $f(x) \to 0$ as $x \to \infty$.
   (g) If f is continuous on $\mathbb{R}$ and $\lim_{n\to\infty} \int_{-n}^{n} f$ exists, then f is integrable.

15.2 For each of the choices of I below, specify a subset A of $\mathbb{R}$ so as to make the following statement valid: $|x|^p$ is integrable on the specified interval if and only if $p \in A$.

(i) $(0, 5]$, (ii) $[1/2, \infty)$, (iii) $[-1, 2]$, (iv) $(-\infty, \infty)$.

15.3 For each of the following sequences $\{f_n\}$ calculate $\lim \int f_n$ and $\int \lim f_n$:

$$\text{(i) } f_n(x) = n^2 x \mathrm{e}^{-nx} \chi_{[0,\infty)}(x), \qquad \text{(ii) } f_n(x) = n^2 \mathrm{e}^{-n^2 x^2} \chi_{[0,1)}(x).$$

In (ii) you may assume that $\int_0^\infty \mathrm{e}^{-y^2}\,\mathrm{d}y = \sqrt{\pi}/2$. Comment on your answers.

15.4 Use the Comparison Theorem to prove that the following are integrable:

$$\text{(i) } \frac{\sin^2 x}{x^2} \text{ on } \mathbb{R}, \qquad \text{(ii) } x\mathrm{e}^{-x}\cos x \text{ on } [0, \infty),$$
$$\text{(iii) } \sin^2(1/x) \text{ on } [10^{-23}, \infty), \qquad \text{(iv) } \mathrm{e}^{-|x|-x^{-2}} \text{ on } \mathbb{R}.$$

15.5 (a) Prove that $\log(\sin x)$ and $\log(\cos x)$ are integrable on $(0, \pi/2)$. [Hint: for $\log(\sin x)$ compare with $\log x$.]

(b) Prove that

$$\begin{aligned}\int_0^{\pi/2} \log(\cos x)\,\mathrm{d}x &= \int_0^{\pi/2} \log(\sin x)\,\mathrm{d}x \\ &= \int_{\pi/2}^{\pi} \log(\sin x)\,\mathrm{d}x = \frac{1}{2}\int_0^{\pi} \log(\sin x)\,\mathrm{d}x.\end{aligned}$$

By writing $x = 2y$, find $\int_0^{\pi/2} \log(\cos x)\,\mathrm{d}x$.

15.6 Prove that $\mathrm{e}^{-x}\cos x \log(x+\alpha)$ is integrable on $(0,\infty)$ for any $\alpha > 0$.

15.7 Here is a construction of a continuous integrable function f such that $f(x) \not\to 0$ as $x \to \infty$. We call g a 'spike function' if there exists $a \in \mathbb{R}$ and $\delta > 0$ such that $g = 0$ outside $(a-\delta, a+\delta)$ and g is continuous (so $g \in \mathrm{L}^{\mathrm{C}}$), and g is a linear function on $[a-\delta, a]$ and on $[a, a+\delta]$.

(a) For $k = 1, 2, \ldots,$ construct a spike function g_k zero outside the interval $[k-2^{-k}, k+2^{-k}]$ such that $\int g_k = 2^{-k}$.

(b) Define $f := \sum_{k=1}^{\infty} g_k$. Sketch the graph of f and apply the MCT to show that f is integrable.

16 Techniques of integration II

In Chapter 6 we showed how to evaluate a wide range of integrals $\int_a^b f$, where f is continuous and $[a,b]$ is closed and bounded. We now extend our methods to the evaluation of integrals of functions on intervals of other types.

Let $f \in \mathrm{L}(\mathbb{R}^+)$. In the Lebesgue theory it is not true **by definition** that

$$\int_0^\infty f = \lim_{n\to\infty} \int_0^n f$$

(though this is the definition in certain other theories of integration). For us results such as this are consequences of the DCT.

16.1 Proposition. Let $f \in \mathrm{L}_{\mathbb{R}}$ [more generally, $f \in \mathrm{L}$]. Then

(a) $\int_{-\infty}^n f \to \int f$ as $n \to \infty$;

(b) $\int_n^\infty f \to 0$ as $n \to \infty$;

(c) $\int_{-\infty}^Y f \to \int f$ as $Y \to \infty$;

(d) $\int_{-X}^Y f \to \int f$ as $X, Y \to \infty$.

Proof. For (a), we apply the DCT to $\{f\chi_{(-\infty,n]}\}$, noting that (D1) and (D2) are immediate, while $|f|$ serves as the dominating function G required for (D3). Since $\int_n^\infty f = \int_{-\infty}^\infty f - \int_{-\infty}^n f$, (b) follows from (a).

For (c), choose $n \in \mathbb{N}$ such that $n \leqslant Y < n+1$. Then, using (b),

$$\left|\int_n^Y f\right| \leqslant \int_n^Y |f| \leqslant \int_n^\infty |f| \to 0 \text{ as } n \to \infty.$$

We leave (d) as an exercise. $\square$

16.2 Exercise example. Let $I = \langle a, b\rangle$ where $-\infty \leqslant a < b \leqslant \infty$ and let $\{I_n\}$ be a sequence of compact intervals such that $I_n \subseteq I_{n+1}$ and $I = \bigcup I_n$ (for example, if $I = (-\infty, \infty)$ we might have $I_n = [-n, n]$). Use the DCT to prove that if $f \in \mathrm{L}(I)$ then $\int_I f = \lim_{n\to\infty} \int_{I_n} f$.

The result in 16.2 has worthwhile practical consequences, foreshadowed in 15.9. Remember that the FTC and its corollaries apply to suitable functions on a **compact interval**. We can now extend these to intervals which are unbounded or not closed. To illustrate how to proceed we consider the FTC itself.

16.3 Extended FTC. Let $I = \langle a,b\rangle$ $(-\infty \leqslant a < b \leqslant \infty)$ be a non-compact interval, and assume that $F\colon I \to \mathbb{R}$ is continuously differentiable and such that $F' \in \mathrm{L}(I)$. Then

$$\int_I F' = F(b-) - F(a+),$$

provided the limits on the right-hand side exist.

Proof. Take a sequence $\{[a_n, b_n]\}$ of compact intervals such that $I = \bigcup [a_n, b_n]$, $a_n \searrow a$ and $b_n \nearrow b$. Then

$$\begin{aligned} \int_I F' &= \lim_{n\to\infty} \int_{I_n} F' && \text{(by 16.2)} \\ &= \lim_{n\to\infty} F(b_n) - \lim_{n\to\infty} F(a_n) && \text{(by FTC).} \quad \square \end{aligned}$$

16.4 Exercise example.

(a) Prove the following **extended integration by parts** criterion: if $I = \langle a,b\rangle$ $(-\infty \leqslant a < b \leqslant \infty)$ and, on I, u and v have continuous derivatives and are such that uv' and vu' are integrable on I, then

$$\int_I uv' = (uv)(b-) - (uv)(a+) - \int_I vu',$$

provided the limits on the right-hand side exist.

(b) Prove the following criterion for **extended substitution**: Let $I = \langle a,b\rangle$ $(-\infty \leqslant a < b \leqslant \infty)$ and $J = \langle c,d\rangle$ $(-\infty \leqslant c < d \leqslant \infty)$, and assume that g maps J onto I and has a continuous positive derivative, that $f \in \mathrm{L}(I)$ and f is continuous, and that $(f\circ g)g' \in \mathrm{L}(J)$; then

$$\int_a^b f(t)\,\mathrm{d}t = \int_c^d f(g(x))g'(x)\,\mathrm{d}x.$$

[Hints: Choose a sequence $\{J_n\}$ of compact intervals such that $J = \bigcup J_n$, appeal to 6.6 and apply 16.2 twice.]

We do not recommend that you memorize the conditions for extended integration by parts, etc.—it is better to revert to the DCT in specific cases. Later on we shall use the FTC, integration by parts, or substitution, on arbitrary intervals without explicitly going through the truncation arguments except where circumspection is called for.

It is important to realise that the results above allow us to evaluate integrals once we know the functions involved are integrable. They cannot be used to **prove** integrability; see Exercise 16.9.

16.5 Examples.

(1) Let $f(x) = \mathrm{e}^{-ax}\sin x$ on $[0,\infty)$, where $a > 0$. We proved in 15.9 that f is integrable. We can now calculate $\int f$. We use integration by parts, proceeding as in 16.4:

$$\begin{aligned}\int_0^\infty e^{-ax}\sin x\,dx &= \lim_{n\to\infty}\int_0^n e^{-ax}\sin x\,dx \qquad\qquad \text{(by the DCT)}\\ &= \left[-e^{-ax}\cos x\right]_0^\infty - \lim_{n\to\infty}\left(\int_0^n ae^{-ax}\cos x\,dx\right)\\ &= 1+\lim_{n\to\infty}\left(\left[-e^{-ax}\cos x\right]_0^n - \int_0^n a^2e^{-ax}\sin x\,dx\right),\end{aligned}$$

whence

$$\int_0^\infty e^{-ax}\sin x\,dx = \frac{1}{1+a^2}.$$

In a similar way we may prove that

$$\int_0^\infty e^{-ax}\cos x\,dx = \frac{a}{1+a^2}.$$

As indicated in 12.12, we can alternatively treat the above integrals together: we may write

$$\int_0^\infty e^{ix-ax}\,dx = \left[\frac{e^{ix-ax}}{i-a}\right]_0^\infty = \frac{1}{a-i},$$

and then equate real and imaginary parts.

(2) Consider $f(x) = x^p e^{-ax}$ on $[0,\infty)$ ($p = 1,2,\dots,\ a>0$). We can use extended integration by parts to prove that

$$\int_0^\infty x^p e^{-ax}\,dx = \frac{p}{a}\int_0^\infty x^{p-1}e^{-ax}\,dx.$$

Hence by induction, using 8.16, and 15.3 for the base case $p=0$, we obtain

$$\int_0^\infty x^p e^{-ax}\,dx = \frac{p!}{a^{p+1}}.$$

16.6 Example. We claim that

$$\lim_{n\to\infty}\int_0^n x^r\left(1-\frac{x}{n}\right)^n dx = r!$$

for any positive integer r. First note that it is advisable to build the variable limit of integration into the integrand—then we can forget about it! So define

$$f_n(x) := x^r\left(1-\frac{x}{n}\right)^n\chi_{[0,n]}(x),$$

so that $\lim\int f_n$ is the left-hand side of the given expression. We evaluate it by proving that $\lim\int f_n = \int\lim f_n$. In this case both the MCT and the DCT are available to us. We opt for the DCT, since it is not particularly easy to show that $\{f_n\}$ is increasing. We check the DCT conditions. (D1) holds because $f_n \in \mathrm{L}^{\mathrm{C}}$ and (D2) is immediate from the standard limit $\lim_{n\to\infty}(1-x/n)^n = e^{-x}$ (see

Appendix II). For (D3) it will be sufficient to show that $|f_n(x)| \leqslant x^{-r}\mathrm{e}^{-x}$ a.e. on $(0,\infty)$, by Example 15.3. For $0 \leqslant t < 1$ we have

$$(1-t)^{-1} = 1 + t + \cdots + t^k + \ldots \geqslant 1 + \frac{t}{1!} + \cdots + \frac{t^k}{k!} + \cdots = \mathrm{e}^t.$$

Hence for $0 \leqslant x < n$ we have $(1 - x/n)^{-1} \geqslant \mathrm{e}^{x/n}$. Taking the nth power and rearranging we obtain $(1 - x/n)^n \leqslant \mathrm{e}^{-x}$ on $[0,n)$.

16.7 Exercise example. Evaluate $\displaystyle\lim_{n\to\infty} \int_0^n \left(1 + \frac{x}{n}\right)^n \mathrm{e}^{-2x}\,\mathrm{d}x$.

16.8 Improper integrals. Given a function f such that

(i) $f\chi_{[X,Y]}$ is integrable whenever $-\infty < X < Y < \infty$, and

(ii) $J := \lim_{X,Y\to\infty} \int_X^Y f$ exists,

then f is said to have an *improper integral*, namely J. We have seen examples in Chapter 13 of functions whose improper integrals exist but which are not (Lebesgue) integrable, notably $x^{-1}\sin x\chi_{[0,\infty)}(x)$. The DCT shows that if f is Lebesgue integrable then the improper integral J exists and equals the Lebesgue integral $\int f$, as does the *principal value integral* $\lim_{X\to\infty}\int_{-X}^X f$ (see 16.1(d)). This observation is useful because one of the traditional methods of evaluating integrals, by contour integration in the complex plane, calculates the improper integral or the principal value integral; see Chapters 8 and 9 of **[13]**.

In 13.7 we exhibited a derivative which failed to be integrable. This apparent oddity falls within the penumbra of improper integrals.

Exercises

16.1 Prove that $1/\sqrt{x(1-x)}$ is integrable on $(0,1)$ and by justifying the substitution $x = \sin^2 y$, prove that

$$\int_0^1 \frac{1}{\sqrt{x(1-x)}}\,\mathrm{d}x = \pi.$$

16.2 Prove that

$$\int_{-\infty}^{\infty} \mathrm{e}^{-|x|-\mathrm{i}yx}\,\mathrm{d}x = 2(1+y^2)^{-1}.$$

16.3 Prove that, for $t > -1$ and $y > 0$,

$$\lim_{n\to\infty} n^t \int_0^1 x^{t-1}(\mathrm{e}^{-xy}(1-x))^n\,\mathrm{d}x = (y+1)^{-t}\Gamma(t),$$

where

$$\Gamma(t) := \int_0^\infty x^{t-1}\mathrm{e}^{-x}\,\mathrm{d}x \quad (t > 0).$$

[Here Γ is the *Gamma function*. By 16.5(2), $\Gamma(n) = n!$ for $n \in \mathbb{N}$, so Γ extends the factorial function.]

16.4 (a) Prove that $(1+x^2)^{-1}\log x$ is integrable on $(0,\infty)$.

(b) Show, by extended substitution, that

$$\int_0^1 \frac{\log x}{1+x^2}\,dx = -\int_1^\infty \frac{\log x}{1+x^2}\,dx$$

and deduce that

$$\int_0^\infty \frac{\log x}{1+x^2}\,dx = 0.$$

(c) Deduce that $\displaystyle\int_0^\infty \frac{\log x}{a^2+x^2}\,dx = \frac{\pi}{2a}\log a$ for every $a>0$.

16.5 Prove by extended integration by parts, with appropriate justification, that

$$\int_0^1 (x\log x)^n\,dx = (-1)^n\frac{n!}{(n+1)^{n+1}} \quad (n=1,2,\dots).$$

[Each function $x^r(\log x)^s$ $(r,s=1,2,\dots)$ is continuous on $[0,1]$ when defined to take the value 0 at 0. Why cannot the integration by parts result from Chapter 6 be used to prove the result above?]

16.6 (a) By comparison with y^s where $q>s>-1$ prove that $y^q(\log y)^p$ is integrable on $(0,1]$ for $p=1,2,\dots,$ $q>-1$.

(b) By justifying the substitution $y=e^{-x}$ and using the equation

$$\int_0^\infty x^p e^{-ax}\,dx = \frac{p!}{a^{p+1}} \quad (a>0, p=1,2,\dots)$$

(recall Example 16.5(2)), prove that

$$\int_0^1 y^q(\log y)^p\,dy = (-1)^q\frac{p!}{(q+1)^{p+1}}.$$

16.7 Prove by induction, giving a brief justification, that

$$\int_0^\infty x^{2n}e^{-x^2}\,dx = \frac{1.3.5.\dots.(2n-1)}{2^n}\int_0^\infty e^{-x^2}\,dx \quad (n=1,2,\dots).$$

16.8 Let $\gamma_n := \sum_{k=1}^n 1/k - \log n$ (so $\gamma_n \to \gamma$, Euler's constant; see 7.15). Prove that

$$\gamma_n = \int_0^n \frac{1}{x}\left(1-\left(1-\frac{x}{n}\right)^n\right)dx - \int_1^n \frac{1}{x}\,dx.$$

Hence prove that

$$\gamma = \int_0^1 \frac{1}{x}(1-e^{-x})\,dx - \int_1^\infty \frac{e^{-x}}{x}\,dx.$$

16.9 Let $f(x)=x^{-1}\sin x$ on $[\pi,\infty)$. By integrating by parts over $[\pi,X]$, prove that $\lim_{X\to\infty}\int_\pi^X f(x)\,dx$ exists. [Cf. 13.5 and 15.9. The integration by parts argument here does not constitute a proof that $f\in \mathrm{L}$—we know $f\notin \mathrm{L}$.]

17 Sums and integrals

In Chapter 8 we considered term-by-term integration of series in the context of continuous functions on compact intervals. It is important to realize that the elementary arguments applicable there do not extend to functions on unbounded sets. In the general setting the convergence theorems can be used instead. The examples and exercises in this chapter make use of well-known series expansions; most of these are recalled in Appendix II.

17.1 Interchange of $\sum$ and $\int$. Linearity tells us that, for finitely many integrable functions, $u_1, \ldots, u_n$,

$$\int (u_1 + \cdots + u_n) = \int u_1 + \cdots + \int u_n.$$

Remember that an infinite sum is the limit of its partial sums. Let $\{u_k\}$ be a sequence of integrable functions and let $f_n := \sum_{k=1}^{n} u_k$. Note that f_n is integrable. Then

$$\begin{aligned}
\sum_{k=1}^{\infty} \int u_k &= \lim_{n\to\infty} \sum_{k=1}^{n} \int u_k && \text{(by definition)}\\
&= \lim_{n\to\infty} \int \sum_{k=1}^{n} u_k && \text{(by (L) (the sum is finite))}\\
&= \lim_{n\to\infty} \int f_n && \text{(by definition)}\\
&\stackrel{*}{=} \int \lim_{n\to\infty} f_n && \text{(if lim and } \textstyle\int \text{ can be interchanged)}\\
&= \int \sum_{k=1}^{\infty} u_k && \text{(by definition).}
\end{aligned}$$

Thus the problem of interchanging $\sum$ and $\int$ reduces to justifying $(*)$. We issue a repeat warning that this is not always possible: see Exercise 17.3 for an example. However the interchange is legitimate whenever one of the convergence theorems applies to the sequence $\{f_n\}$ of partial sums.

In order to apply the MCT to the partial sum sequence we need $f_n \leqslant f_{n+1}$ for all n. Since $f_{n+1} - f_n = u_{n+1}$ we see that $u_k \geqslant 0$ for all k gives us $\{f_n\}$ increasing.

17.2 MCT applied to series of non-negative terms. Let $\{u_k\}$ be a sequence of real-valued functions such that

(i) u_k is integrable for all k,

(ii) $u_k \geqslant 0$ for all k, and

(iii) $\sum_{k=1}^{\infty} \int u_k$ converges.

Then $\sum_{k=1}^{\infty} u_k$ converges a.e. and

$$\sum_{k=1}^{\infty} \int u_k = \int \sum_{k=1}^{\infty} u_k.$$

(Note: given (i) and (ii), condition (iii) is necessary as well as sufficient for the conclusion of the theorem to hold.)

Proof. Most of the work has already been done, but for reference we run through the MCT check list for $\{f_n\}$, where $f_n := \sum_{k=1}^{n} u_k$.

(M1) f_n is a finite sum of integrable functions, so integrable.

(M2) $f_n \leqslant f_{n+1}$ for all n because each $u_k \geqslant 0$.

(M3) $\int f_n = \sum_{k=1}^{n} \int u_k \leqslant \sum_{k=1}^{\infty} \int u_k := K$ (by (L) and (P)). Certainly K is independent of n. By hypothesis K is finite. □

17.3 Example. We prove that, for $0 \leqslant t \leqslant 1$,

$$\int_0^1 \frac{1-x}{1-tx^3}\,\mathrm{d}x = \sum_{k=0}^{\infty} \frac{t^k}{(3k+1)(3k+2)}.$$

For $|tx^3| < 1$ we have, expanding $(1-tx^3)^{-1}$ as a geometric series,

$$\frac{1-x}{1-tx^3} = \sum_{k=0}^{\infty} t^k(1-x)x^{3k}.$$

For fixed $t \in [0,1]$ define $f_n\colon [0,1] \to \mathbb{R}$ by

$$f_n(x) = \sum_{k=0}^{n} t^k(1-x)x^{3k}.$$

We apply the MCT to $\{f_n\}$, so illustrating 17.2 rather than encouraging commitment of its conditions to memory.

(M1) f_n is integrable because $f_n \in C[0,1]$.

(M2) $f_n \leqslant f_{n+1}$ for all n because $t^k(1-x)x^{3k} \geqslant 0$ for all $x \in [0,1]$.

(M3) By the FTC we have

$$\begin{aligned}\int f_n = \sum_{k=0}^{n} \int_0^1 t^k(1-x)x^{3k}\,\mathrm{d}x &= \sum_{k=0}^{n} \frac{t^k}{(3k+1)(3k+2)} \\ &\leqslant K := \sum_{k=0}^{\infty} \frac{t^k}{(3k+1)(3k+2)},\end{aligned}$$

where K is finite because $\sum t^k/((3k+1)(3k+2))$ converges by comparison with $\sum 1/k^2$.

(M4) $f_n(x) \to \dfrac{1-x}{1-tx^3}$ a.e. on $[0,1]$ (the single exceptional point being $x=1$).
This proves the stated result.

We now deduce the value of $\sum_{k=0}^{\infty} 1/((3k+1)(3k+2))$. Putting $t=1$ above and remembering 6.4 we get

$$\sum_{k=0}^{\infty} \frac{1}{(3k+1)(3k+2)} = \int_0^1 \frac{1}{1+x+x^2}\,\mathrm{d}x = \left[\frac{2}{\sqrt{3}}\tan^{-1}\frac{(2x+1)}{\sqrt{3}}\right]_0^1 = \frac{1}{3\sqrt{3}}.$$

Remark. As we are working with a compact interval here, we might be tempted to think that the elementary techniques of 8.10 could be used. For each $t \neq 1$ this is indeed the case—it is an easy exercise that the series converges uniformly. However, for $t=1$ (the value we are ultimately interested in) convergence is not uniform and the elementary theory cannot help us.

17.4 Exercise example. Prove that if $p>0$ then

$$\int_0^{\infty} \frac{x}{\mathrm{e}^{px}(1-\mathrm{e}^{-x})}\,\mathrm{d}x = \int_0^{\infty} \sum_{k=0}^{\infty} x\mathrm{e}^{-px-kx}\,\mathrm{d}x = \sum_{k=0}^{\infty} \frac{1}{(k+p)^2}.$$

We now turn to series which do not have non-negative terms. A real or complex series $\sum a_k$ converges if $\sum |a_k|$ does (that is, if $\sum a_k$ converges absolutely). The next theorem is in the same spirit. It is useful for justifying interchange of $\sum$ and $\int$ in concrete examples. We show in Chapter 28 that it also has theoretical importance.

17.5 Theorem. Let $\{u_k\}_{k\geqslant 1}$ be a sequence of real- or complex-valued integrable functions.

(a) Assume that $\sum_{k=1}^{\infty} \int |u_k|$ converges. Then $\sum_{k=1}^{\infty} |u_k|$ converges a.e. and is integrable.

(b) Assume that $\sum_{k=1}^{\infty} |u_k|$ is integrable (and so is necessarily finite a.e.). Then $\sum_{k=1}^{\infty} u_k$ is integrable and

$$\int \sum_{k=1}^{\infty} u_k = \sum_{k=1}^{\infty} \int u_k.$$

(c) Under the same hypothesis as in (b),

$$\lim_{n\to\infty} \int \left| f - \sum_{k=1}^{n} u_k \right| = 0.$$

Proof. For (a) we apply the MCT to $\{g_n\}$ where $g_n := \sum_{k=1}^{n} |u_k|$, noting that, for all n, $\int g_n \leqslant K := \sum_{k=1}^{\infty} \int |u_k|$, which is finite by hypothesis.

Consider (b). We apply the DCT to $\{f_n\}$ where $f_n := \sum_{k=1}^{n} u_k$. Certainly each f_n is integrable. By the triangle inequality

$$|f_n| \leqslant \sum_{k=1}^{n} |u_k| \leqslant \sum_{k=1}^{\infty} |u_k|$$

and $\sum_{k=1}^{\infty} |u_k|$ provides our dominating function. We know that this function is finite a.e. Because absolute convergence implies convergence, $\sum_{k=1}^{\infty} u_k$ converges a.e., so there exists a finite-valued function f such that $f_n \to f$ a.e. Hence the DCT conditions are satisfied and the first part of (b) follows as in 17.1.

For (c) we use the infinite version of the triangle inequality (Exercise 2.9) and then apply (b) to the tail of the series to get

$$\int \Big| f - \sum_{k=1}^{n} u_k \Big| \leqslant \int \sum_{k=n+1}^{\infty} |u_k| = \sum_{k=n+1}^{\infty} \int |u_k| \to 0 \text{ as } n \to \infty, \quad \square$$

Remarks. If the functions u_k are real-valued we have an alternative strategy for the proof of 17.5: to apply the MCT to the sequences $\{\sum_{k=1}^{n}(u_k)^+\}$ and $\{\sum_{k=1}^{n}(u_k)^-\}$. We leave the details as an exercise.

We may sum up the conclusion of Theorem 17.5 by saying that

$$\sum \int u_k = \int \sum u_k$$

provided each u_k is integrable and either
(a) $\sum |u_k|$ is integrable, or
(b) $\sum \int |u_k|$ converges.
The proof shows that (b) implies (a); in practice we usually check (b).

17.6 Example. We prove that if $\alpha \in \mathbb{R}$ then

$$\int_0^{\infty} \frac{\sin \alpha x}{e^x - 1}\, dx = \sum_{k=1}^{\infty} \frac{\alpha}{\alpha^2 + k^2}.$$

We must expand the integrand on the left-hand side as a series $\sum u_k$ and then justify interchange of sum and integral. Evidently, each u_k must be integrable, with an integral we can evaluate. This leads us to expand the denominator $(e^x - 1)^{-1}$ binomially, and to leave the numerator as it is. So write

$$\frac{\sin \alpha x}{e^x - 1} = \sum_{k=1}^{\infty} e^{-kx} \sin \alpha x;$$

this expansion is valid when $e^{-x} < 1$, that is, when $x > 0$. Then, for any n, $e^{-kx} \sin \alpha x$ is integrable by comparison with $|\alpha|\, x e^{-kx}$ (see 15.8 or 15.9) and

$$\sum_{k=1}^{n} \int_0^{\infty} \left| e^{-kx} \sin \alpha x \right| dx \leqslant \sum_{k=1}^{n} \int_0^{\infty} |\alpha|\, x e^{-kx}\, dx = \sum_{k=1}^{\infty} \frac{|\alpha|}{k^2}$$

(by 16.5(2)). The series $\sum |\alpha| / k^2$ converges. Applying Theorem 17.5 we obtain

$$\int_0^{\infty} \frac{\sin \alpha x}{e^x - 1}\, dx = \sum_{k=1}^{\infty} \int_0^{\infty} e^{-kx} \sin \alpha x\, dx = \sum_{k=1}^{\infty} \frac{\alpha}{\alpha^2 + k^2}.$$

Remark. In the above we used the inequality $|\sin \alpha x| \leqslant |\alpha x|$ to show by comparison that $\sum_{k=1}^{\infty}\{\int_0^{\infty}|e^{-kx}\sin \alpha x|\,dx\}$ converges. Our first instinct would have been to use the cruder estimate $|\sin \alpha x| \leqslant 1$, but this would have given us the upper bound $\sum_{k=1}^{\infty} 1/k$, which is useless because $\sum 1/k$ diverges.

17.7 Exercise example. Assuming that $\int_0^{\infty} e^{-x^2}\,dx = \sqrt{\pi}/2$ prove that

$$\int_0^{\infty} e^{-x^2}\cos 2xt\,dx = \frac{\sqrt{\pi}}{2}e^{-t^2} \quad (t \in \mathbb{R}).$$

[Hint: remember the concluding remark in Chapter 15! You will need the result of Exercise 16.7. An alternative method, using differentiation, is given in 20.5.]

17.8 Summary. Theorem 17.5 subsumes (in a heavy-handed way) the results on term-by-term integration of series in Chapter 8. However, where they can be applied the elementary techniques of Chapter 8 should be used. For series of positive terms, Theorem 17.2 is the definitive result. Inapplicability of Theorem 17.5 does not always mean that term-by-term integration is invalid; see Exercise 17.5.

Exercises

17.1 Prove that

$$\int_0^1 \sum_{k=1}^{\infty} (e^{-kx} - 2e^{-2kx})\,dx \neq \sum_{k=1}^{\infty} \int_0^1 (e^{-kx} - 2e^{-2kx})\,dx.$$

[Hint: show the left-hand side is > 0 and the right-hand side is < 0.]

17.2 Prove, with the aid of Exercise 16.5, that

$$\int_0^1 x^{-x}\,dx = \sum_{k=1}^{\infty} k^{-k}.$$

17.3 Let $t > 1$. Prove that $x^{t-1}e^{-x}/(1-e^{-x})$ is integrable on $[0,\infty)$ and

$$\int_0^{\infty} \frac{x^{t-1}e^{-x}}{1-e^{-x}}\,dx = \Gamma(t)\sum_{n=1}^{\infty} n^{-t},$$

where $\Gamma(t) = \int_0^{\infty} x^{t-1}e^{-x}\,dx$.

17.4 Prove that, for any positive real numbers r, s,

$$\int_0^1 \frac{x^{r-1}}{1+x^s}\,dx = \sum_{k=0}^{\infty} \frac{(-1)^k}{r+ks}.$$

[Hint: apply the MCT to $\{f_n\}$ where $f_n(x) := \sum_{k=0}^{2n-1}(-1)^k x^{r+ks-1}$.] Deduce

(i) $\dfrac{\pi}{4} = 1 - \dfrac{1}{3} + \dfrac{1}{5} - \dfrac{1}{7} + \ldots,$

(ii) $\log 2 = 1 - \dfrac{1}{2} + \dfrac{1}{3} - \dfrac{1}{4} + \ldots.$

17.5 Let $u_k(x) := (-1)^k x^k$. Show that it is not possible to prove by any of the following methods that $\sum_{k=0}^{\infty}\int_0^1 u_k = \int_0^1 \sum_{k=0}^{\infty} u_k$:

(a) applying the MCT for series (17.2) to $\sum u_k$;

(b) applying the MCT for series to $\sum (u_k)^+$ and $\sum (u_k)^-$;

(c) applying Theorem 17.5 to $\sum u_k$;

(d) applying 8.10 (uniform convergence).

17.6 Prove that

$$\int_0^1 \frac{1+t}{1+xt^3}\,\mathrm{d}t = \sum_{r=0}^{\infty}(-1)^r\left(\frac{1}{3r+1} + \frac{1}{3r+2}\right)x^r \qquad (-1 \leqslant t \leqslant 1).$$

Deduce the value of

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

17.7 Let $0 < a < 1$. By expanding $1/(1-\mathrm{e}^{-x})$ as a series of powers of e^{-x}, prove that $(\mathrm{e}^{-ax} - \mathrm{e}^{(a-1)x})/(1-\mathrm{e}^{-x})$ is integrable and that

$$\int_0^{\infty} \frac{\mathrm{e}^{-ax} - \mathrm{e}^{(a-1)x}}{1-\mathrm{e}^{-x}}\,\mathrm{d}x = \frac{1}{a} + \sum_{r=1}^{\infty}\left(\frac{1}{a-r} + \frac{1}{a+r}\right).$$

17.8 (a) Assuming that $\int_0^{\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}/2$, prove that, for $\alpha \in \mathbb{R}$,

$$\int_0^{\infty} \mathrm{e}^{-x^2}\sin^2(\alpha x)\,\mathrm{d}x = \frac{\sqrt{\pi}}{4}\left(1 - \mathrm{e}^{-\alpha^2}\right).$$

(b) Prove that, for $\alpha \in \mathbb{R}$,

$$\int_0^{\infty} \frac{\sin^2(\alpha x)}{\mathrm{e}^{x^2}+1}\,\mathrm{d}x = \frac{\sqrt{\pi}}{2}\sum_{k=1}^{\infty}(-1)^{k-1}\frac{1-\mathrm{e}^{-\alpha^2/k}}{\sqrt{k}}.$$

18 Recognizing integrable functions II

This is an opportune stage at which to reinforce what we have discovered about integrable functions in practice, and to show how to confirm integrability or otherwise in an economical way. We have used the following facts in particular examples earlier (see 15.11 and 15.7(4)):
(a) $x^p e^{-x} \leqslant 1$ if x is sufficiently large;
(b) for any $p > 0$, $x^{-p} \log x \leqslant 1$ if x is sufficiently large;
(c) for any $p > 0$, there exists $\delta > 0$ such that $x^p \log x \leqslant 1$ for $0 < x < \delta$.
They follow from the limits derived in 8.16 and 5.12 by the Local Boundedness Lemma, 2.21. We now recall a useful piece of standard notation which will allow us to use such bounds more systematically, focusing on the order of magnitude of a function.

18.1 The O notation. We say U is a punctured neighbourhood of $a \in \mathbb{R}$ if $U = (a - \delta, a + \delta) \smallsetminus \{a\}$ for some $\delta > 0$. Likewise, a punctured neighbourhood of ∞ is a set of the form (X, ∞) for some $X \in \mathbb{R}$, and similarly $(-\infty, X)$ is a punctured neighbourhood of $-\infty$. Given real- or complex-valued functions f and g and $a \in \mathbb{R} \cup \{-\infty, \infty\}$, we write

$$f(x) = \mathrm{O}(g(x)) \text{ as } x \to a$$

if there exists a constant K and a punctured neighbourhood U of a such that $|f(x)| \leqslant K\,|g(x)|$ for $x \in U$. From the Local Boundedness Lemma we see that $f(x)/g(x) \to \ell$ as $x \to a$ (ℓ finite) implies $f(x) = \mathrm{O}(g(x))$ as $x \to a$.

The following easy lemma lets us localize the investigation of integrability.

18.2 Lemma. Let I be an interval which is the disjoint union of finitely many intervals $I_1, \ldots, I_k$. Then $f \in \mathrm{L}(I)$ if and only if $f \in \mathrm{L}(I_j)$ for $j = 1, \ldots, k$.

We may amalgamate earlier results to obtain the following integrability criterion. It should be treated as guidance on how to proceed in examples rather than a result to quote verbatim.

18.3 A localized comparison test for integrability. Let f be continuous except at the points of a finite set S, possibly empty. Assume that for each $a \in S \cup \{-\infty, \infty\}$ there exists a punctured neighbourhood U_a of a and a function g_a integrable on U_a such that $f(x) = \mathrm{O}(g_a(x))$ as $x \to a$. Then f is integrable.

Proof. We may break $\mathbb{R}$ into a finite union of disjoint intervals on each of which f is integrable, and then invoke 18.2. For integrability on an interval around a point of S we use the Comparison Theorem, and for a compact interval containing no point of S we note that f is integrable there because it is continuous. $\square$

18.4 Examples.

(1) A reminder that negative exponentials are a Good Thing. Let $f(x) = x^{100}e^{x^2-x^4}$ on $\mathbb{R}$. Since f is continuous on $\mathbb{R}$ it is integrable on any compact interval. Hence only its behaviour as $x \to \pm\infty$ is at issue. For large $|x|$ the negative exponential e^{-x^4} dominates both the polynomial x^{100} and the positive exponential e^{x^2}. More explicitly,

$$x^2 - x^4 \leqslant -\frac{1}{2}x^4 \quad \text{and} \quad |x^{100}|\,e^{-x^4/4} \leqslant 1$$

if $|x|$ is sufficiently large. Hence

$$x^{100}e^{x^2-x^4} \leqslant x^{100}e^{-x^4/2} = O(e^{-x^4/4}) \quad \text{as } |x| \to \infty.$$

Since $e^{-x^4/4}$ is integrable on $\mathbb{R}$, we conclude from 15.6 that f is integrable. Note how we split e^{-x^4} into two parts—$e^{-3x^4/4}$ to tame $x^{100}e^{x^2}$ and $e^{-x^4/4}$ to supply an integrable function for the Comparison Theorem.

(2) (Compare with 15.7(4)) Let $f(x) = x^{-1/2}\log x$ on $(0,1]$. We claim that f is integrable. Note that f is continuous, and hence integrable, on any interval $[\delta, 1]$ where $0 < \delta < 1$. Also,

$$x^{1/4}\log x \to 0 \quad \text{as } x \to 0+.$$

Hence $x^{-1/2}\log x = \mathbf{O}(x^{-3/4})$ as $x \to 0+$. Since $x^{-3/4}$ is integrable on $(0,1]$ we conclude that f is integrable.

(3) Let $f(x) = x\sin x(1+x^2)^{-1}$ on $\mathbb{R}$. We claim that f is not integrable. Note that $f(x)$ behaves like $g(x) := (x\sin x)/x^2$ for large $|x|$, and that g is not integrable, by 13.5. Since $1 + x^2 \leqslant 2x^2$ for large $|x|$ we have

$$\frac{|x\sin x|}{1+x^2} \geqslant \left|\frac{\sin x}{2x}\right|.$$

Thus, arguing by contradiction, we see that f cannot be integrable.

18.5 Exercise example.

(a) Prove that the following are integrable:

$$\text{(i) } \frac{\sin x}{x(1+x^2)}, \qquad \text{(ii) } x^{-3/2}e^x \sin x \quad \text{on } (0,1].$$

(b) Prove that the following are not integrable:

$$\text{(i) } \frac{x^2\cos x}{\sqrt{1+x^3}} \quad \text{on } [0,\infty), \qquad \text{(ii) } x(1-\sin x)^{-1} \quad \text{on } (0,\pi).$$

18.6 Exercise example. Let $f(x) = x^p \log x$ on $[1, \infty)$ where $p < -1$. Choose $\alpha > 0$ such that $p < -1 - 2\alpha < -1$. Show that there exists X such that $|f(x)| \leqslant x^{p+\alpha}$ for $x \geqslant X$. Deduce that f is integrable on $[1, \infty)$.

18.7 An example for the intrepid. Take $f(x) := -\mathrm{e}^{-x} \log(\sin^2 x)$ on $(0, \infty)$. Note that $\log(\sin^2 x)$ is non-positive and periodic of period 2π, with $\log(\sin^2 x) \to -\infty$ as $x \to a_k := k\pi$ and $\log(\sin^2 x) = 0$ at $b_k := (2k+1)\pi/2$ ($k = -1, 0, 1, \dots$). For each k, $-\log(\sin^2 x)$ is decreasing on $(a_k, b_k]$ and increasing on $[b_k, a_{k+1})$. The graph of f is indicated in Fig. 18.1. Near a_k, put $y = x - a_k$. Then $\sin^2 x = \sin^2 y = \mathrm{O}(y^2)$ as $y \to 0$, and $\log(\sin^2 y) = \mathrm{O}(\log y^2) = \mathrm{O}(y^{-1/2})$. Since $\mathrm{e}^{-(y+a_k)} y^{-1/2}$ is integrable near $y = 0$ (why?) we deduce that f is integrable in a neighbourhood of each individual 'bad point' a_k ($k \geqslant 0$). This encourages us to hope that f may be integrable. However the set of bad points is infinite, and the combined contribution to $\int |f|$ of integrals $\int_{I_k} |f|$, where I_k is a small interval round a_k, might not be finite. Fix δ with $0 < \delta < \pi/2$ (to be chosen later), let $I_k := (a_k - \delta, a_k + \delta)$ and $J_k = [a_k + \delta, a_{k+1} - \delta]$. We shall show that

$$f\chi_{\bigcup_{k \geqslant 0} I_k} = \sum_{k=0}^{\infty} f\chi_{I_k} \quad \text{and} \quad f\chi_{\bigcup_{k \geqslant 1} J_k} = \sum_{k=1}^{\infty} f\chi_{J_k}$$

are integrable. This will be sufficient to show f is integrable on $(0, \infty)$.

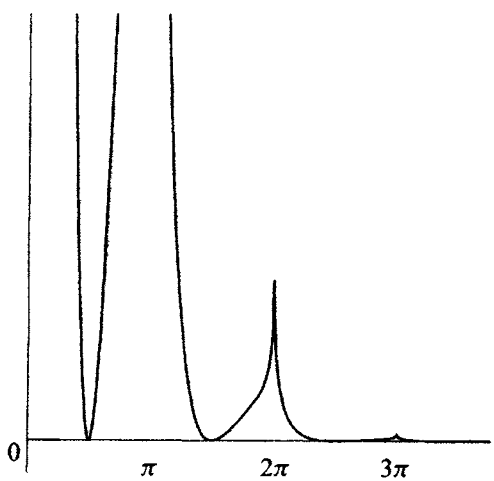


Figure 18.1

For $k \geqslant 0$ and for δ so small that $|\log(\sin^2 y)| \leqslant |y|^{-1/2}$ for $y \in (-\delta, \delta)$,

$$\begin{aligned}\int_{I_k} \left|e^{-x} \log(\sin^2 x)\right| dx &\leqslant e^{-a_{k-1}} \int_{-\delta}^{\delta} \left|\log(\sin^2 y)\right| dy \\ &\leqslant e^{-a_{k-1}} \int_{-\delta}^{\delta} |y|^{-1/2}\, dy, \\ &= 4\sqrt{\delta} e^{-a_{k-1}}.\end{aligned}$$

The geometric series $\sum e^{-a_k}$ converges. We may then apply the MCT to the partial sums of the series $\sum_{k=0}^{\infty} f\chi_{I_k}$ to show that it is integrable.

The intermediate intervals J_k are easier to handle. We have

$$\left|\log(\sin^2 x)\right| \leqslant \left|\log(\sin^2 \delta)\right| \quad \text{on } J_k.$$

Because e^{-x} is integrable on $(0, \infty)$ it is easy to see that the MCT implies that $\sum_{k=1}^{\infty} f\chi_{J_k}$ is integrable.

18.8 Summary. Our strategy for investigating a given function f is as follows.

(1) Find the set S of points of discontinuity of f.

(2) Investigate the behaviour of f near each point of $S \cup \{\pm\infty\}$ separately, identifying the dominant term and relating it to standard comparison functions (exponentials, etc.).

(3) Apply the Comparison Theorem in a neighbourhood of each point of S, arguing by contradiction in cases of non-integrability.

(4) Consider the cumulative effect of the 'bad points' if S is infinite. (If S has limit points in $\mathbb{R}$, see if 11.2 applies, at least locally.)

Exercises

18.1 Find for which real values of p and q the following functions are integrable and for which are they non-integrable, and justify your answers:

$$\text{(i) } x^p \sin(qx), \qquad \text{(ii) } x^p e^{-qx} \chi_{[0,\infty)}(x), \qquad \text{(iii) } x^p \log(x^q) \chi_{(0,1]}(x).$$

18.2 For each of the following functions either prove it is integrable or that it is not integrable.

$$\text{(i) } x^{-1.0001} e^{-x/1000} \log x \chi_{[0,\infty)}(x), \qquad \text{(ii) } \frac{x^2}{1 + x^2 + x^4},$$

$$\text{(iii) } \frac{\cos x}{e^x - 1}, \qquad \text{(iv) } \frac{x^3 \cos x}{1 + x^4}.$$

18.3 (a) Prove that $\dfrac{\sin x}{e^x - 1}$ is integrable on $[0, \infty)$.

(b) Prove that $\dfrac{e^{-ax} - e^{(a-1)x}}{1 - e^{-x}}$ is integrable on $\mathbb{R}$ $(0 < a < 1)$.

18.4 By comparing with

$$\chi_{[-\delta,\delta]}(y)(1 + b^2k^2y^2)^{-1} \quad \text{where } bk\delta \geqslant 1$$

near $x = k\pi$, or otherwise, prove that $x(1 + x^2 \sin^2 x)^{-1}$ is integrable on $[0, \infty)$.

19 The Continuous DCT

We frequently encounter functions defined by integrals, and wish to establish their properties. Typically we would have a function F of the form

$$F(t) := \int f(x,t)\,\mathrm{d}x \quad (t \in J),$$

where J is some open interval in $\mathbb{R}$, and $f(x,t)$ is a real- or complex-valued integrable function of x for each fixed $t \in J$. Questions we might wish to ask are the following.

- Does $\lim_{t\to a} F(t)$ exist? If so, is its value $\int \lim_{t\to a} f(x,t)\,\mathrm{d}x$?
- Is F continuous at a given point $a \in J$? That is, is it the case that $\lim_{h\to 0} F(a+h) = F(a)$?
- Is F differentiable, with its derivative given by 'differentiation under the integral sign'? That is, do we have

  $$F'(t) := \int \frac{\partial}{\partial t}\mathrm{f(x,t)}\,\mathrm{d}x \quad (t \in J)?$$

  For example, is it true that

  $$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^\infty \frac{\mathrm{e}^{-x}\sin xt}{x}\,\mathrm{d}x = \int_0^\infty \frac{\partial}{\partial t}\left(\frac{\mathrm{e}^{-x}\sin xt}{x}\right)\mathrm{d}x = \int_0^\infty \mathrm{e}^{-x}\cos xt\,\mathrm{d}x?$$

  Beware of assuming that this is obviously true—to get it, the order of differentiation with respect to t and integration with respect to x have been swopped. See Exercise 20.3 for an example where this is not admissible. Remember our Ruritanian tale in Chapter 14!

There is a common form to all three problems above. Take

in the first, $f_h(x) = f(x, a+h)$,

in the second, $f_h(x) = f(x,t)$ (t any fixed point of J), and

in the third, $f_h(x) = \dfrac{f(x,t+h) - f(x,t)}{h}$ (t any fixed point of J).

In each case we would like to prove that

$$\lim_{h\to 0}\int f_h \quad \text{exists and equals} \quad \int \lim_{h\to 0} f_h.$$

This is a parallel problem to that which the convergence theorems are designed to solve, the difference being that here we have a limit involving a continuous variable, namely h, while before we had a discrete variable, n. Thus we need convergence theorems with a continuous variable.

It is a well-known result from elementary analysis that a function $F\colon \mathbb{R} \to \mathbb{R}$ is continuous at a point a if and only if $\lim_{n\to\infty} F(y_n) = F(a)$ for every sequence $\{y_n\}$ converging to a. We use the same technical trick here for switching backwards and forwards between discrete and continuous limits.

19.1 Technical lemma. Assume that $F\colon \mathbb{R} \to \mathbb{R}$ and that $a, \alpha \in \mathbb{R}$. Then the following are equivalent:

(a) $F(y) \to \alpha$ as $y \to a$;

(b) $F(y_n) \to \alpha$ as $n \to \infty$ for every sequence $\{y_n\}$ such that $a \neq y_n$ and $y_n \to a$ as $n \to \infty$.

Proof. This is a standard piece of ε-ology. The easier implication is (a) $\Longrightarrow$ (b)—going from the general to the particular. Assume (a). Then, given $\varepsilon > 0$, we may find $\delta > 0$ such that

$$(*) \qquad 0 < |y - a| < \delta \implies |F(y) - \alpha| < \varepsilon.$$

For any sequence $\{y_n\}$ as in (b) we may find N (depending on δ) such that

$$n \geqslant N \implies |y_n - a| < \delta.$$

Combining these two statements gives

$$n \geqslant N \implies |F(y_n) - \alpha| < \varepsilon,$$

which tells us that $F(y_n) \to \alpha$ as $n \to \infty$. Thus (b) holds.

Conversely, we prove the contrapositive: we assume that (a) fails and construct a 'bad' sequence $\{y_n\}$ for which (b) fails. The failure of (a) means that there is some ε for which no δ works in $(*)$. Fix such an ε. Then for each n the value $1/n$ will not serve for δ, so there is a point, y_n say, such that

$$|F(y_n) - \alpha| \geqslant \varepsilon \quad \text{and} \quad 0 < |y_n - a| < 1/n.$$

Clearly $\{y_n\}$ is the required sequence violating (b). $\square$

[**Note**. Remember that the limit definition demands that $|F(y) - \alpha| < \varepsilon$ for $0 < |y - a| < \delta$, ensures that we consider only the value of $F(y)$ as y **approaches** a: $F(a)$ need not be defined, and, even if it is, we do not wish to demand that $F(a) = \alpha$. This explains the restriction '$y_n \neq a$'.]

We present the most frequently used form of Continuous DCT. Variants, for example for a limit as $h \to \infty$, are easily formulated.

19.2 The Continuous DCT. Let $\{f_h\}$ be a family of real-valued [or complex-valued] functions on $\mathbb{R}$, where $0 < |h| < H$ (H constant). Assume that

(C1) f_h is integrable for each h,

(C2) there exists a function f such that $f_h \to f$ a.e. as $h \to 0$,

(C3) there exists an integrable function G such that $|f_h| \leqslant G$ for all h, where G is independent of h.

Then f is integrable and

$$\lim_{h \to 0} \int f_h = \int \lim_{h \to 0} f_h = \int f.$$

[If (C3) holds, then (C1) is satisfied whenever each f_h is continuous (by 15.6 [or merely measurable (by 21.3)].]

Proof. Let $\{h_n\}$ be any sequence of non-zero real numbers with $h_n \to 0$. Define $g_n := f_{h_n}$. Then $\{g_n\}$ satisfies the DCT conditions, with $g_n \xrightarrow{\text{a.e.}} f$ (by the easy half of 19.1) and with dominating function G. Hence, by the DCT, $f \in \mathrm{L}$ and

$$\int f = \lim_{n\to\infty} \int g_n = \lim_{n\to\infty} \int f_{h_n}.$$

Now apply 19.1 to F where $F(h) := \int f_h$. □

19.3 Example. We prove that

$$\lim_{h\to 0+} \int_0^\infty h\mathrm{e}^{-x} \cos x \log\left(x + \frac{1}{h}\right) \mathrm{d}x = 0.$$

We let $f_h(x) := h\mathrm{e}^{-x} \cos x \log\left(x + h^{-1}\right)$ $(x \geqslant 0,\ h > 0)$. The function f_h is integrable (f_h is continuous on $[0, \infty)$ and near ∞ the negative exponential is dominant; Exercise 15.6 seeks a detailed verification from the Comparison Theorem). Also, as $h \to 0+$, $h\log(x+h^{-1}) = h\log(xh+1) - h\log h \to 0$ for any $x \geqslant 0$. So (C1) and (C2) hold. Finally, for (C3) first note that we lose nothing by assuming $0 < h < 1$. Then

$$|f_h(x)| \leqslant h\mathrm{e}^{-x} \log\left(x + \frac{1}{\mathrm{h}}\right) \leqslant h(x + \frac{1}{\mathrm{h}})\mathrm{e}^{-x} \leqslant (x+1)\mathrm{e}^{-x}.$$

Since $(x + 1)\mathrm{e}^{-x}$ is integrable and independent of h, it serves as the required dominating function. We remark that the cosine term is nothing more than a distraction. The trick of restricting to values of h close to the limiting value 0 is very helpful, and is used repeatedly in applications of the Continuous DCT.

19.4 Example: continuity of the indefinite integral. We claim that, for $f \in \mathrm{L}$,

$$F(t) := \int_{-\infty}^{t} f(x)\,\mathrm{d}x$$

defines a continuous function F. For fixed t, we wish to show that $F(t + h) \to F(t)$ as $h \to 0$. We apply the Continuous DCT to $\{f_h\}$, where $f_h := f\chi_{(-\infty,t+h]}$, for fixed t. Certainly each f_h is integrable, and $|f_h| \leqslant G := |f|$, which gives the dominating function we need. Finally, unless x is between t and $t + h$,

$$f_h(x) - f_0(x) == f(x)(\chi_{(-\infty,t+h]}(x) - \chi_{(-\infty,t]}(x)) = 0,$$

so that $f_h(x) \xrightarrow{\text{a.e.}} f(x)\chi_{(-\infty,t]}(x)$, the exceptional null set containing (at most) the single point t. Thus by the Continuous DCT,

$$\lim_{h\to 0} F(t + h) = \lim_{h\to 0} \int f_h = \int \lim_{h\to 0} f_h = F(t).$$

19.5 Remarks. We do not explicitly treat integrals with variable limits of integration, since we can incorporate these into the integrand, as in 19.4—but don't forget to do this!

In a few cases—principally where we have continuous functions on compact intervals—we can avoid the Continuous DCT and use elementary estimation instead. See Exercise 19.1.

Note that when we want to prove continuity of a function

$$F(t) := \int f(x,t)\,\mathrm{d}x$$

on some interval J, then we do it for each point t individually. When we apply the Continuous DCT to $\{f_h\}$ where $f_h(x) := f(x, t+h)$, we are regarding t as fixed. Thus the dominating function G is allowed to depend on t—but it **must not depend on** h.

Exercises

19.1 Prove that the Bessel function J_0, given by

$$J_0(t) := \frac{2}{\pi}\int_0^{\pi/2} \cos(t\cos\theta)\,\mathrm{d}\theta,$$

is continuous on $\mathbb{R}$

(i) by an elementary estimation of $J_0(t+h) - J_0(t)$,

(ii) by the Continuous DCT.

19.2 Prove that, for any positive real number λ, the function f_λ given by $f_\lambda(x) = \mathrm{e}^{-\lambda \sinh^2 x}\sin x$ is integrable and that

$$\int_{-\infty}^{\infty} f_\lambda(x)\,\mathrm{d}x \to 0 \quad \text{as } \lambda \to \infty.$$

19.3 Let f be a continuous function on $[0,\infty)$ such that $f(\infty) := \lim_{x\to\infty} f(x)$ exists. Prove that for any $\lambda > 0$, $f(x)\mathrm{e}^{-\lambda x}$ is integrable on $[0,\infty)$ and

$$\lim_{\lambda\to\infty} \lambda\int_0^{\infty} f(x)\mathrm{e}^{-\lambda x}\,\mathrm{d}x = f(0) \quad \text{and} \quad \lim_{\lambda\to 0+} \lambda\int_0^{\infty} f(x)\mathrm{e}^{-\lambda x}\,\mathrm{d}x = f(\infty).$$

19.4 Recall 7.14, 7.15. Let $N \in \mathbb{N}$.

(a) Prove that

$$\lim_{p\to 0+}\int_1^N x^{-(p+1)}\,\mathrm{d}x = \log N.$$

(b) Prove that

$$0 < \sum_{k=N}^{\infty} k^{-(p+1)} - \int_N^{\infty} x^{-(p+1)}\,\mathrm{d}x < N^{-(p+1)} \quad (p > -1).$$

(c) Deduce that

$$\lim_{p\to 0+}\Big(\sum_{k=1}^{\infty} k^{-(p+1)} - \frac{1}{p}\Big) = \gamma.$$

20 Differentiation of integrals

In the previous chapter we presented the Continuous DCT and gave some simple applications. Here we focus on its use in establishing differentiability of functions defined by integrals, as foreshadowed in Chapter 19. The rudiments of the method are covered in 20.2–20.6; other sections may be studied only as needed.

20.1 Partial derivatives. We need some notation concerning partial derivatives. The expressions

$$\frac{\partial}{\partial t} f(x,t) \quad \text{and} \quad \left.\frac{\partial}{\partial t} f(x,t)\right|_{t=u}$$

are understood to mean, respectively, the derivative with respect to t of the function $f(x,t)$ evaluated with x held fixed, and this function evaluated at $t=u$. We shall also write the latter more compactly as $f_2(x,u)$, the subscript indicating that differentiation is with respect to the second variable. This allows us to avoid any confusion between fixed and variable arguments.

20.2 Differentiability by the Continuous DCT. Given $F(t) := \int f(x,t)\,\mathrm{d}x$, defined for t lying in an open interval J in $\mathbb{R}$, we should like to equate

$$\frac{\mathrm{d}}{\mathrm{d}t}\int f(x,t)\,\mathrm{d}x = \lim_{h\to 0}\frac{F(t+h)-F(t)}{h} = \lim_{h\to 0}\int \frac{f(x,t+h)-f(x,t)}{h}\,\mathrm{d}x$$

and

$$\int \frac{\partial}{\partial t} f(x,t)\,\mathrm{d}x = \int \lim_{h\to 0}\frac{f(x,t+h)-f(x,t)}{h}\,\mathrm{d}x.$$

To ensure these expressions make sense we assume that, for each fixed $t \in J$,

(i) $f(x,t)$ is integrable as a function of x, and

(ii) $f_2(x,t)$ exists for almost all x.

Because J is open, $t+h \in J$ for $|h|$ sufficiently small. Recalling our remarks in Chapter 19 we aim to apply the Continuous DCT to $\{f_h\}$, where

$$f_h(x) := \frac{f(x,t+h)-f(x,t)}{h} \quad (t \in J,\ t \text{ fixed}).$$

Our assumptions (i) and (ii) imply that f_h is defined and is integrable for $h \neq 0$, $|h|$ suitably small, and that $\lim_{h\to 0} f_h(x)$ exists a.e. and equals $f_2(x,t)$. Our discussion below focuses on the key question: whether (C3) holds, that is, whether there exists G such that $|f_h(x)| \leqslant G(x)$, where G is integrable and independent of h, but may depend on the fixed point t. We concentrate on illustrative examples, but do present in 20.4 a simple sufficient (but not necessary!) condition for G to exist.

20.3 Example. We shall evaluate

$$F(t) := \int_0^\infty \frac{e^{-x}\sin xt}{x}\,dx.$$

The obstacle to recognizing the integrand as a derivative with a view to applying the FTC is the factor x in the denominator. This disappears if we differentiate the integrand with respect to the parameter t. So we want to justify writing

$$F'(t) = \int_0^\infty \frac{\partial}{\partial t}\left(\frac{e^{-x}\sin xt}{x}\right)dx = \int_0^\infty e^{-x}\cos xt\,dx.$$

Fix t and define

$$f_h(x) := \frac{\sin x(t+h) - \sin xt}{xh}e^{-x} \quad (h \neq 0).$$

Then $f_h(x) \to e^{-x}\cos xt$ as $h \to 0$. We put forward two possible strategies for getting a dominating function. First, we may use the formula

$$\sin A - \sin B = 2\sin\left(\tfrac{A-B}{2}\right)\cos\left(\tfrac{A+B}{2}\right)$$

to get

$$|f_h(x)| = \left|e^{-x}\cos(x(2t+h)/2)\frac{\sin(xh/2)}{xh/2}\right| \leqslant \left|e^{-x}\right|.$$

To obtain the inequality recall that $|\sin u| \leqslant |u|$ for all $u \in \mathbb{R}$. Since f_h is continuous on $(0,\infty)$ and dominated by an integrable function, it is integrable by the Comparison Theorem. For our dominating function we may take $G(x) = e^{-x}\chi_{[0,\infty)}(x)$, which is certainly integrable and independent of h. Alternatively, apply the Mean Value Theorem (2.26(2)) to the function $y \mapsto \sin xy$ (with x fixed) on the interval with endpoints $t, t+h$. This gives

$$\sin x(t+h) - \sin xt = xh\cos x(t+\theta h),$$

where $|\theta| < 1$ (and θ depends on x, t, and h). Hence we see once again that each f_h is integrable by the Comparison Theorem and that e^{-x} provides a dominating function. Hence, by the Continuous DCT,

$$F'(t) = \int_0^\infty e^{-x}\cos xt\,dx = \frac{1}{1+t^2}$$

(cf. 16.5). Thus, by the FTC, $F(t) = F(0) + \tan^{-1} t = \tan^{-1} t$.

20.4 A sufficient (but not necessary) condition for 'differentiation under the integral sign'. Let f be a real-valued function defined on $\mathbb{R} \times J$, where J is an open interval in $\mathbb{R}$, and assume that, for each fixed $t \in J$,

(i) $x \mapsto f(x,t)$ is integrable,
(ii) $f_2(x,t)$ exists for almost all x,
(iii) there exists an integrable function G, independent of y, such that

$$\left|\frac{\partial}{\partial y} f(x,y)\right| \leqslant G(x) \quad \text{for almost all } x \text{ and for all } y \in J.$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{\infty} f(x,t)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{\partial}{\partial t} f(x,t)\,\mathrm{d}x \quad (t \in J).$$

Proof. Fix $t \in J$ and let $f_h(x) := (f(x,t+h) - f(x,t))/h$ for h such that $h \neq 0$ and $t+h \in J$. As in 20.2, (i) implies that each f_h is integrable (as a function of x) and (ii) tells us that, for almost all x, $\lim_{h\to 0} f_h(x)$ exists and equals $f_2(x,t)$. By the MVT and (iii), there exists θ such that

$$|f_h(x)| = |f_2(x, t+\theta h)| \leqslant G(x)$$

(here $t+\theta h$ is some point between t and $t+h$, and so is necessarily in J). Thus the Continuous DCT applies to $\{f_h\}$, with dominating function G. □

Remark. When using 20.4 be very careful to check that your candidate dominating function G really is integrable and is independent of t as t varies over J.

20.5 Exercise example. For $t \in \mathbb{R}$, let $F(t) = \int_0^\infty \mathrm{e}^{-x^2}\cos 2xt\,\mathrm{d}x$. By applying 20.4, find a differential equation for $F(t)$, and hence evaluate $F(t)$. [You may assume the fact that $\int_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}/2$ (see 20.10).]

20.6 A warning example. We know that

$$F(t) := \int_0^\infty \frac{1}{t+x^2}\,\mathrm{d}x = \left[\frac{1}{\sqrt{t}}\tan^{-1}\left(\frac{x}{\sqrt{t}}\right)\right]_0^\infty = \frac{\pi}{2\sqrt{t}} \quad (t > 0).$$

Provided we may differentiate under the integral sign with respect to t we get

$$\int_0^\infty \frac{-1}{(t+x^2)^2}\,\mathrm{d}x = -\frac{\pi}{2t^{3/2}}.$$

For 20.4 to apply on $J := (0,\infty)$ we need an integrable bound on $(y+x^2)^{-2}$, independent of $y \in J = (0,\infty)$. The only possible choice is x^{-4}, which is **not** integrable on $[0,\infty)$. So we can't use 20.4 on $(0,\infty)$. However if we take $J_c := (c,\infty)$ where c is a constant > 0, then we can apply 20.4 on J_c, with $G(x) := (c+x^2)^{-2}$. We conclude that $F(t)$ is differentiable at each point t of (c,∞). Since this is true for every $c > 0$, $F'(t)$ is obtained by differentiation under the integral sign for each $t > 0$. The reason this ruse works is that the derivative at a point t is calculated locally, from values of the function close to t.

20.7 Looking for a dominating function. Consider finding G such that

$$|f_h(x)| = \left|\frac{f(x,t+h) - f(x,t)}{h}\right| \leqslant G(x),$$

where t is fixed, and G is required to be integrable and independent of h (but allowed to depend on the fixed point t).

- Whenever we use the Continuous DCT we are concerned with limiting behaviour as $h \to 0$. Therefore nothing is lost by restricting to small values of h, for example $|h| < 1$ or $|h| < T$, where T depends on the point t at which we wish to differentiate; see 20.10 for an illustration.
- Examining the proof of 20.4 we see that it relies on finding a blanket bound $G(x)$ on $f_h(x)$—the same bound for **every** point $t \in J$, and that this is just what we cannot do in 20.6. By looking at one point t at a time and focusing on the behaviour of $F(t)$ near this point, we escape from this strait jacket. In 20.6 we achieved the same end in a slightly different way, by applying 20.4 on truncated intervals; below we show how to proceed instead 'one point at a time'. Either way, we exploit the fact that differentiation is a local affair.
- Whether working globally as in 20.4 or 'one point at a time', the MVT is often useful for estimating the quotient $(f(x,t+h) - f(x,t))/h$. Remember that if this quotient is written as

  $$f_2(x, t+\theta h), \quad \text{where } |\theta| < 1,$$

  then θ will in general depend on x, t, and h. Therefore θ may not appear in G. Remember that the MVT is not available to us when f is complex-valued, since it is inherently a theorem about real-valued functions: it comes from Rolle's Theorem, and the proof of this uses the order structure of $\mathbb{R}$.
- When we do our estimation we must not forget the factor h in the denominator of f_h. Not only must we cancel out the explosive behaviour of $1/h$ as $h \to 0$ but we must get an integrable bound on $(f(x,t+h)-f(x,t))/h$ which does not involve h. This may be tricky: a good armoury of inequalities is recommended!

20.8 Example 20.6 revisited. Let $f(x,t) := (t+x^2)^{-1}$ $(x \geqslant 0,\ t > 0)$ and define f_h from f as in 20.6. Fix $t > 0$. Then, by the MVT,

$$\begin{aligned} |f_h(x)| = \left|\frac{f(x,t+h) - f(x,t)}{h}\right| &= \left|\frac{-1}{((t+\theta h)+x^2)^2}\right| && \text{(where } 0 < \theta < 1) \\ &\leqslant G(x) := \frac{1}{((t/2)+x^2)^2} && \text{when } 0 < |h| < t/2. \end{aligned}$$

For the bound on the denominator, note that $t/2 < t + \theta h$ (remember that h could be negative). Here G is integrable and independent of h. Hence (C3) holds for $\{f_h\}_{0<|h|<t/2}$. The other two conditions certainly hold, so differentiation under the integral sign is permissible at each $t > 0$.

20.9 Example: the derivative of the Fourier transform. Assume $f, g \in \mathrm{L}$, where $g(x) := xf(x)$. Let

$$\hat{f}(y) := \int_{-\infty}^{\infty} f(x)\mathrm{e}^{-\mathrm{i}yx}\,\mathrm{d}x$$

be the Fourier transform of f; we prove in 23.2(a) the plausible fact that $f(x)\mathrm{e}^{-\mathrm{i}yx}$ is integrable for any $y \in \mathbb{R}$. Then $\hat{f}$ is differentiable and $(\hat{f})'(y) = -\mathrm{i}\hat{g}(y)$.

Proof. We apply the Continuous DCT for complex-valued functions with

$$f_h(x) := f(x)\frac{\mathrm{e}^{-\mathrm{i}(y+h)x} - \mathrm{e}^{-\mathrm{i}yx}}{h} = f(x)\mathrm{e}^{-\mathrm{i}yx}\frac{\mathrm{e}^{-\mathrm{i}hx} - 1}{h}.$$

Since f_h is complex-valued we cannot use the MVT to find a dominating function. Instead we proceed as follows (remember that $|\mathrm{e}^{\mathrm{i}\alpha}| = 1$ when α is real):

$$\begin{aligned} |f_h(x)| &= \left| f(x)\,\mathrm{e}^{-\mathrm{i}yx}\mathrm{e}^{\mathrm{i}hx/2}\left(\frac{\mathrm{e}^{-\mathrm{i}hx/2} - \mathrm{e}^{\mathrm{i}hx/2}}{h}\right)\right| \\ &= 2\,|f(x)| \left|\frac{\sin hx/2}{h}\right| = |xf(x)| \left|\frac{\sin hx/2}{hx/2}\right| \leqslant |g(x)|, \end{aligned}$$

the last line coming from the fact that $|\sin u| \leqslant |u|$ for all $u \in \mathbb{R}$. Therefore the postulate that $xf(x) \in \mathrm{L}$ is just what we need for (C3). Conditions (C1) and (C2) certainly hold, $\lim f_h(x)$ being $-\mathrm{i}xf(x)$.

We already know that e^{-x^2} is integrable. We alluded in Chapters 1 and 15 to the famous result that $\int_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}/2$. This cannot be proved by the methods of elementary calculus—substitution, integration by parts, etc. We can now present the first of three derivations; for the others see 26.12 and 27.6.

20.10 Example. We prove—at last—that $\int_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}/2$. Let

$$F(t) = \int_0^\infty \frac{\mathrm{e}^{-(1+x^2)t^2}}{2(1+x^2)}\,\mathrm{d}x.$$

For fixed $t > 0$, we apply the Continuous DCT to $\{f_h\}$ where

$$f_h(x) := \frac{\mathrm{e}^{-(1+x^2)(t+h)^2} - \mathrm{e}^{-(1+x^2)t^2}}{2h(1+x^2)} \qquad (0 < |h| < \delta)$$

and δ is chosen later. The function $(1+x^2)^{-1}$ is integrable (see 15.4). By the Comparison Theorem each f_h is integrable. By the chain rule for differentiation,

$$f_h(x) \to -t\mathrm{e}^{-(1+x^2)t^2} \quad \text{as } h \to 0 \quad (t \text{ fixed}).$$

By the MVT, there exists θ $(0 < \theta < 1$, θ depending on x, t, and $h)$ such that

$$f_h(x) = -(t+\theta h)\mathrm{e}^{-(1+x^2)(t+\theta h)^2}.$$

Restrict attention to h satisfying $0 < |h| < \delta := t/2$; this is a convenient rather than a necessary choice. For such h we have $t/2 < t + \theta h < 3t/2$. Hence

$$|f_h(x)| \leqslant G(x) := \frac{3t}{2} e^{-(1+x^2)t^2/4},$$

and G serves as the required dominating function; note the way we have used our upper bound on $t + \theta h$ to control the first factor in f_h and the lower bound to control the negative exponential term. We conclude that, for $0 < t < \infty$,

$$F'(t) = -\int_0^\infty -te^{-(1+x^2)t^2}\,dx = -e^{-t^2}\int_0^\infty te^{-x^2t^2}\,dx = -Ce^{-t^2},$$

where the constant C is $\int_0^\infty e^{-v^2}\,dv$. For the last step here we use the substitution $v = xt$ (t fixed).

We may now use the (extended) FTC (16.3):

$$C\int_0^\infty e^{-x^2}\,dx = \lim_{Y\to\infty} F(Y) - \lim_{X\to 0+} F(X).$$

The left-hand side is $\left(\int_0^\infty e^{-x^2}\,dx\right)^2$ so to finish the proof it is enough to show that

$$F(t) \to 0 \text{ as } t \to \infty \quad\text{and}\quad F(t) \to \frac{\pi}{4} \text{ as } t \to 0+.$$

We could prove these assertions by the Continuous DCT. Alternatively we can simply observe that, by Property (P),

$$0 \leqslant F(t) \leqslant e^{-t^2}\int_0^\infty \frac{1}{2(1+x^2)}\,dx = \frac{\pi}{4}e^{-t^2} \to 0 \text{ as } t \to \infty, \quad\text{and}$$

$$0 \leqslant \frac{\pi}{4} - F(t) = \int_0^\infty \frac{1}{2(1+x^2)}(1 - e^{t^2})\,dx = \frac{\pi}{4}(1 - e^{t^2}) \to 0 \text{ as } t \to 0+.$$

Exercises

20.1 Let $t > 0$. Evaluate $\int_0^\infty (a^2 + x^2)^{-(n+1)}\,dx$ ($n = 1, 2, \dots$) by writing $t = a^2$ and differentiating with respect to t. [This extends 20.8.]

20.2 Evaluate $\int_0^1 x^n(-\log x)^m\,dx$ (m, n non-negative integers) by replacing m by a variable t and differentiating with respect to t.

20.3 Let $F(t) = \int_1^\infty f(x,t)\,dx$ where

$$f(x,t) = \begin{cases} t^3e^{-t^2x} & \text{if } t > 0, \\ 0 & \text{if } t \leqslant 0. \end{cases}$$

(a) Prove that

$$F'(t) = \int_1^\infty \frac{\partial}{\partial t} f(x,t)\,dx \quad (t > 0),$$

(b) Prove that

$$F'(0) = -1 \quad \text{and} \quad \int_1^\infty f_2(x,0)\,\mathrm{d}x = 0.$$

[So differentiation under the integral sign is not valid in this case.]

20.4 Prove that for any $\alpha \in \mathbb{R}$,

$$\int_0^\infty \mathrm{e}^{-x^2}\sin^2(\alpha x)\,\mathrm{d}x = \frac{\sqrt{\pi}}{4}\left(1 - \mathrm{e}^{-\alpha^2}\right).$$

20.5 Let $F(t) := \int_{-\infty}^{\infty} \mathrm{e}^{x^2 - t^2 x^{-2}}\,\mathrm{d}x$. Prove that F is differentiable on $\mathbb{R}$, with $F'(t) = 2F(t)$ and deduce that $F(t) = \sqrt{\pi}\mathrm{e}^{-2|t|}$.

20.6 Prove that the function F defined by

$$F(t) = \int_0^\infty \frac{\mathrm{e}^{-xt}}{1+x^3}\,\mathrm{d}x \quad (t > 0)$$

satisfies the differential equation $F(t) - F'''(t) = t^{-1}$.

20.7 For $a, b > 0$, let $F(a,b) := \int_0^{\pi/2} \log(a^2\cos^2 x + b^2 \sin^2 x)\,\mathrm{d}x$. For fixed b, let $g(t) = F(t,b)$. Show that

$$g'(t) = \int_0^{\pi/2} \frac{2t\sin^2 x}{t^2\sin^2 x + b^2\cos^2 x}\,\mathrm{d}x \quad (t > 0).$$

[Hint: when applying the Continuous DCT, take $0 < |h| < a/2$ and use the MVT.] Deduce, using 6.10, that

$$F(a,b) = \pi \log\left(\frac{a+b}{2}\right) \quad (a, b > 0).$$

20.8 (a) Let $\alpha \in \mathbb{R}$ and let $q(x)$ be a polynomial.

(i) Prove that e^{-x^4} is integrable.

(ii) Prove that $\mathrm{e}^{-x^4+\alpha x}$ is integrable.

(iii) Prove that $q(x)\mathrm{e}^{-x^4+\alpha x}$ is integrable.

(b) Prove by induction that the kth derivative of F exists for all $k = 1, 2, \ldots$, where

$$F(t) = \int_0^\infty \mathrm{e}^{-x^4+xt}\,\mathrm{d}x.$$

20.9 [The derivative of the Laplace transform, for those with a little knowledge of analysis in the complex plane.] Let $f\colon \mathbb{R}^+ \to \mathbb{C}$ be such that its Laplace transform

$$\overline{f}(p) := \int_0^\infty f(t)\mathrm{e}^{-pt}\,\mathrm{d}t$$

exists for $\operatorname{Re} p > c$, where $c \geqslant 0$.

(a) Fix p with $a := \operatorname{Re} p > c$. By expanding $e^{-ht} - 1$ as a series, prove that, for $|h|$ sufficiently small,

$$\left| \frac{e^{-(p+h)t} - e^{-pt}}{h} \right| \leqslant t e^{|h|t} e^{-at} \leqslant e^{-bt},$$

where $b > c$ and b depends on p, but not on h or t.

(b) By justifying differentiation under the integral sign prove that $(\overline{f})'(p)$ exists for $\operatorname{Re} p > c$ and

$$\overline{f}'(p) = -\int_0^\infty t f(t) e^{-pt}\, dt.$$

21 Measurable functions

Elementary real analysis concentrates on functions which are, at a minimum, continuous. Integrability and continuity do not mesh well: there are continuous functions which are 'too big' to be integrable and integrable functions which are not continuous at any point (examples?). The appropriate concept of minimal good behaviour in integration theory is measurability.

21.1 Introducing measurable functions. Assume that $f \in \mathrm{L}_{\mathbb{R}}$. Then we may write $f = g - h$, where $g, h \in \mathrm{L}^{\mathrm{inc}}$, with $\mathrm{L}^{\mathrm{inc}}$-sequences $\{\varphi_n\}$, $\{\psi_n\}$ converging to g, h off null sets E_1, E_2, respectively. Thus

$$f = \lim \varphi_n - \lim \psi_n = \lim \theta_n$$

off the null set $E := E_1 \cup E_2$, and $\theta_n := \varphi_n - \psi_n \in \mathrm{L}^{\mathrm{step}}$. The condition for integrability of f thus has two constituents:

(i) f is the limit a.e. of a sequence $\{\theta_n\}$ of step functions, and

(ii) the step function θ_n must be decomposable as $\varphi_n - \psi_n$ so that $\{\varphi_n\}$ and $\{\psi_n\}$ are $\mathrm{L}^{\mathrm{inc}}$-sequences, thereby ensuring in particular that $\lim \int \theta_n$ exists.

A function f which satisfies (i) is said to be *measurable*; as usual, we may permit f to fail to have a well- defined real value on some null set. The historical origins of the term 'measurable' will emerge as we proceed. Certainly every integrable function is measurable, but not conversely. For example, $\chi_{\mathbb{R}}$ is the limit a.e. of the sequence of step functions $\{\chi_{[-n,n]}\}$, but is not integrable, as we showed in 13.1. Loosely, we may regard measurability as a regularity condition on f, preventing it from oscillating too erratically. For integrability we require additionally that $|f|$ should not be too large. This informal distinction is made more precise by the following lemma. For another result along the same lines see Exercise 21.2.

21.2 Truncation Lemma. Let f be a real-valued function on $\mathbb{R}$. Then f is measurable if and only if for each k the truncated function

$$f^{\square_k} := ((f \wedge k) \vee (-k))\chi_{[-k,k]}$$

is integrable. (See Fig. 21.1. Note that, for each x, we have $f^{\square_k}(x) = f(x)$ for all k sufficiently large.)

Proof. Assume that f is measurable, and $f = \lim \theta_n$ a.e., where $\theta_n \in \mathrm{L}^{\mathrm{step}}$. Fix k and let $g_n := (\theta_n)^{\square_k}$. Then g_n is a step function, and $g_n \to f^{\square_k}$ a.e. as $n \to \infty$. Further, $|g_n| \leqslant G := k\chi_{[-k,k]}$ for all n, and G is a step function, and so integrable. Hence, by the DCT, $f^{\square_k}$ is integrable.

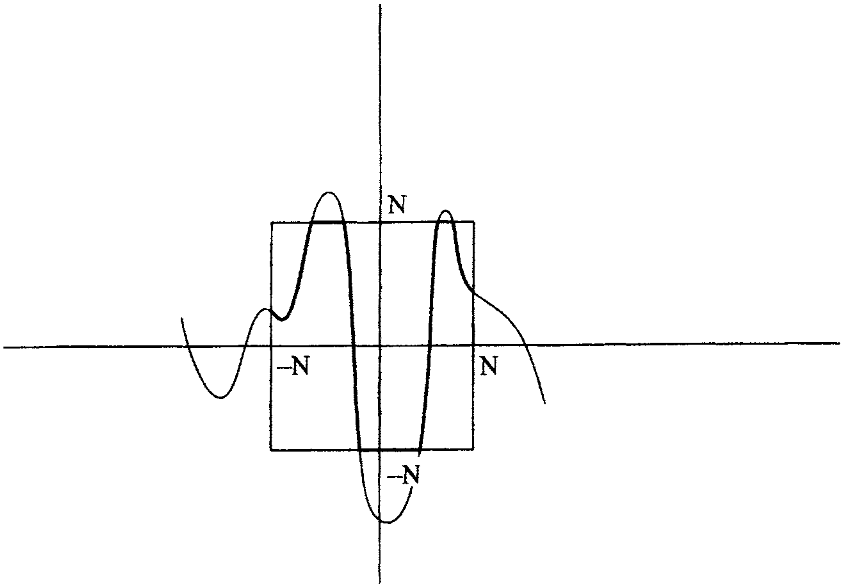


Figure 21.1

Conversely, assume that each $f^{\square_k}$ is integrable. For each k we may find a sequence of step functions $\{\theta_{n,k}\}$ converging to $f^{\square_k}$ as $n \to \infty$ off a null set E_k. Let $E := \bigcup_k E_k$; this is a null set by 10.6. Let $\eta_n := \theta_{n,n}$. We claim that $\eta_n \to f$ off E. Fix $x \notin E$ and take n so large that $x \in [-n, n]$ and $-n \leqslant f(x) \leqslant n$. Then $f(x) = f^{\square_n}(x) = \lim_{m\to\infty} \theta_{m,n}(x)$. Given $\varepsilon > 0$, pick N so that $|f(x) - \theta_{m,n}(x)| < \varepsilon$ for all $m \geqslant N$, where N depends on n. In particular, $|f(x) - \eta_p(x)| = |f(x) - \theta_{p,p}(x)| < \varepsilon$ if p is sufficiently large. Hence f is measurable. $\square$

We can now present our definitive comparison theorem for real-valued functions; for the complex case see 21.7.

21.3 Comparison Theorem. Assume that f is real-valued and measurable, that g is integrable, and that $|f| \leqslant g$. Then f is integrable.

In particular, if f is measurable and $|f|$ is integrable then f is integrable.

Proof. Take a sequence $\{\theta_n\}$ in $\mathrm{L}^{\mathrm{step}}$ converging to f a.e. Then $(\theta_n \wedge g) \vee (-g)$ is integrable, and converges a.e. to $(f \wedge g) \vee (-g)$, that is, to f (because $-g \leqslant f \leqslant g$). Also $|(\theta_n \wedge g) \vee (-g)| \leqslant g$. Thus f is integrable by the DCT. $\square$

21.4 Corollary. A bounded measurable function on a bounded interval I is integrable on I.

Proof. Use the fact that a constant function on I is integrable. $\square$

21.5 The class, M, of real-valued measurable functions.

(a) Let $f, g \in \mathrm{M}$ and $\lambda \in \mathbb{R}$. Then

(i) $f+g$, fg, and λf belong to M;

(ii) $1/f \in \mathrm{M}$ if $f(x) \neq 0$ for almost all x and $\sqrt{f} \in \mathrm{M}$ if $f \geqslant 0$ a.e.;

(iii) $|f|$, $f \vee g$, and $f \wedge g$ belong to M.

(b) If $\{f_n\}$ is a sequence in M such that $f_n \to f$ a.e. then $f \in \mathrm{M}$.

(c) Let $f \in \mathrm{M}$. Then $\operatorname{sgn} f \in \mathrm{M}$ (where $(\operatorname{sgn} f)(x) = 1$, -1, or 0 according as $f(x) > 0$, $f(x) < 0$, or $f(x) = 0$).

Proof. Suppose $\{\varphi_n\}$ and $\{\psi_n\}$ are sequences in $\mathrm{L}^{\mathrm{step}}$ converging a.e. to f and g respectively. Then, for example, $f(x) + g(x) = \lim(\varphi_n + \psi_n)(x)$ a.e. Hence $f+g \in \mathrm{M}$. The other assertions in (a) are proved similarly. To avoid handling them individually we may observe that if $p\colon \mathbb{R} \to \mathbb{R}$ and $q\colon \mathbb{R}^2 \to \mathbb{R}$ are continuous and $f, g \in \mathrm{M}$, then $p \circ f$ and $q \circ (f \sqcap g)$ (where $(f \sqcap g)(x) := (f(x), g(x))$) are measurable. This is proved using the approximating sequences $\{p \circ \varphi_n\}$ and $\{q \circ (\varphi_n \sqcap \psi_n)\}$ (to check that these lie in $\mathrm{L}^{\mathrm{step}}$ see Exercise 3.8). The notation used here may be unfamiliar: to de-mystify it, note that, for example, $(f+g)^2$ could be expressed as $q \circ (f \sqcap g)$, with $q(u,v) := (u+v)^2$ so that $q((f(x), g(x))) = (f(x) + g(x))^2$.

Part (b) is less elementary. To prove it we let $q(x) = \mathrm{e}^{-|x|}$ (to serve as a damping function; any other strictly positive integrable function would do). Let

$$g_n := q\frac{f_n}{1+|f_n|} \quad (n = 1, 2, \dots).$$

Then $g_n \in \mathrm{M}$ and $|g_n| \leqslant q$ so, by the Comparison Theorem, g_n is integrable. By the DCT, $g := \lim g_n = qf/(1+|f|)$ is integrable too, and hence $g \in \mathrm{M}$. But $f = g/(q - |g|)$, with the denominator never zero. Therefore $f \in \mathrm{M}$ by (a).

We now prove (c). Notice that $\operatorname{sgn} f = h - g$, where

$$g(x) = \begin{cases} 1 & \text{if } f(x) \leqslant 0, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad h(x) = \begin{cases} 1 & \text{if } f(x) \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

Define $g_n := n\left((f \vee n^{-1}) - (f \vee 0)\right)$ $(n = 1, 2, \dots)$. Then $g_n \in \mathrm{M}$, and

$$g_n(x) = \begin{cases} 1 & \text{if } f(x) \leqslant 1/n, \\ 0 & \text{otherwise} \end{cases}$$

(to check this consider separately the cases $f(x) \leqslant 0$, $f(x) > 0$). Hence $g = \lim g_n$. By (b), $g \in \mathrm{M}$. Similarly $h \in \mathrm{M}$, so $\operatorname{sgn} f \in \mathrm{M}$. □

21.6 Complex-valued functions again. In 12.11, we declared $f\colon \mathbb{R} \to \mathbb{C}$ to be integrable if and only if its real and imaginary parts $\operatorname{Re} f$ and $\operatorname{Im} f$ belong to $\mathrm{L}_{\mathbb{R}}$, and in that case defined $\int f := \int \operatorname{Re} f + \mathrm{i} \int \operatorname{Im} f$. Measurability extends in the expected way: a complex-valued function is measurable if and only if its real and imaginary parts are measurable. Clearly the complex-valued measurable functions are those functions which are a.e. limits of complex step functions.

Much of the theory of $L_{\mathbb{R}}$ carries over without difficulty to L: a general rule of thumb is 'if it makes sense, it works'. Remember that order properties and complex numbers do not mix well, as is emphasized in [**13**], so that, for example, there is no complex-valued version of the MCT. The Comparison Theorem and the DCT do extend.

21.7 Comparison Theorem and DCT: complex versions. Let f be a complex-valued measurable function such that $|f| \leqslant g$, where g is integrable. Then $\operatorname{Re} f$ and $\operatorname{Im} f$ are real-valued measurable functions dominated by g. Hence, by 21.3, $\operatorname{Re} f, \operatorname{Im} f \in L_{\mathbb{R}}$, so $f \in L$.

Now consider the DCT. Assume $\{f_n\}$ is a sequence of complex-valued measurable functions such that $|f_n| \leqslant G \in L_{\mathbb{R}}$ and $f_n \xrightarrow{\text{a.e.}} f$. From above, the DCT for $L_{\mathbb{R}}$ applies to $\{\operatorname{Re} f_n\}$ and $\{\operatorname{Im} f_n\}$, from which it follows easily that $f \in L$ and $\lim \int f_n = \int f$.

One further result warrants explicit mention.

21.8 Lemma. Let $f : \mathbb{R} \to \mathbb{C}$. Then
(a) f measurable implies $|f|$ measurable.
(b) $f \in L$ implies $|f| \in L_{\mathbb{R}}$ and $\left|\int f\right| \leqslant \int |f|$.

Proof. Assume $f \in L$, so $\operatorname{Re} f, \operatorname{Im} f \in L_{\mathbb{R}}$. Since $|f| = ((\operatorname{Re} f)^2 + (\operatorname{Im} f)^2))^{1/2}$, (a) follows from 21.5. Further,

$$|f| \leqslant |\operatorname{Re} f| + |\operatorname{Im} f|,$$

whence $|f|$ is integrable by the Comparison Theorem. Finally we must prove the inequality. We can write $\left|\int f\right| = \alpha \int f$ where $\alpha \in \mathbb{C}$, $|\alpha| = 1$. Then

$$\left|\int f\right| = \alpha \int f = \int \alpha f := \int \operatorname{Re}(\alpha f) + i \int \operatorname{Im}(\alpha f) = \left|\int \operatorname{Re}(\alpha f)\right|,$$

since the left-hand side is real and non-negative. By the Basic Properties of $L_{\mathbb{R}}$,

$$\left|\int \operatorname{Re}(\alpha f)\right| \leqslant \int |\operatorname{Re}(\alpha f)| \leqslant \int |\alpha f| = \int |f|. \qquad \square$$

21.9 Examples of measurable functions. Certainly every integrable function is measurable. Now let $f : \mathbb{R} \to \mathbb{R}$ be continuous a.e. We claim that $f \in M$. To see this, consider the truncated functions $f^{\square_k}$. On $[-k, k]$, the function $f^{\square_k}$ is continuous at each continuity point of f, and is bounded. Hence, by 11.2(c), $f^{\square_k}$ is integrable. Thus f is measurable, by 21.2. By looking at real and imaginary parts we see that a complex-valued function is measurable if it is continuous a.e. Almost all functions encountered in applications are of this sort, so we have found a very adequate means of checking that individual functions are measurable.

21.10 Non-measurable functions. The natural question arises: how about an example of a non-measurable function? We cannot supply one: non-measurable functions do not arise 'in nature'. This rather surprising statement needs qualifying. It turns out that the existence of non-measurable functions is tightly bound up with the fine structure of the real line and its set-theoretic underpinnings; see 22.8. It may be helpful to compare non-measurable functions to yetis. Some people believe that yetis exist, others believe that they do not. Certainly yetis are not to be found outside highly inaccessible terrain, and no creature ever seen has been confirmed to be a yeti. Nonetheless, until you have evidence to the contrary, it is prudent to assume that an unfamiliar animal you meet might be a dangerous yeti. Likewise, you should give a reason why any function to which you apply a theorem on measurable functions is indeed measurable.

21.11 Non-negative measurable functions. A series of positive terms fails to converge only if its sequence of partial sums diverges to ∞. Similarly, a non-negative measurable function f can only fail to be integrable because the increasing sequence $\{\int f^{\square k}\}$ diverges to ∞ as $k \to \infty$. For any non-negative measurable function f we may adopt the convention that $\int f$ denotes the usual integral if f is integrable, and ∞ otherwise. This allows us to write down $\int f$ without knowing f to be integrable, and saves pedantry later; see the remarks in 26.3. With this convention we may combine the MCT and 13.2 as follows. Let $\{f_n\}$ be an increasing sequence of real-valued integrable functions and let

$$f(x) := \begin{cases} \lim f_n(x) & \text{if } \{f_n(x)\} \text{ converges to a finite limit,} \\ 0 & \text{otherwise.} \end{cases}$$

Then f is measurable (by 21.5(b)) and f is integrable, with $\int f = \lim \int f_n$, if and only if $\lim \int f_n$ is finite.

The next result is in the same spirit, but with weaker hypotheses on $\{f_n\}$.

21.12 Fatou's Lemma. Let $\{f_n\}$ be a sequence of non-negative measurable functions and assume $f_n(x) \to f(x) \in \mathbb{R}$ for almost all x. Then

$$\int f \leqslant \liminf \int f_n.$$

Further, f is integrable if and only if the right-hand side is finite.

Proof. Certainly f is measurable. Let $g_k := f^{\square k}$, so that each g_k is integrable by the Truncation Lemma. Further, $g_k \nearrow f$ as $k \to \infty$, and 21.11 implies that f is integrable if and only if $\lim \int g_k$ is finite.

With k fixed, we apply the DCT to $\{u_{n,k}\}_{n \geqslant 1}$, where $u_{n,k} := g_k \wedge f_n$, which is measurable. Since $0 \leqslant u_{n,k} \leqslant g_k$, we can take g_k as the dominating function. The DCT gives

$$\int g_k = \int \lim_{n \to \infty} u_{n,k} = \lim \int u_{n,k} \leqslant \liminf \int f_n.$$

Now take the supremum over k:

$$\int f = \lim \int g_k \leqslant \liminf f_n. \qquad \square$$

Exercises

21.1 (a) Let $f, g\colon \mathbb{R} \to \mathbb{R}$ and suppose that $f = g$ a.e. Prove that $f \in \mathrm{M}$ if and only if $g \in \mathrm{M}$.

(b) Let $f \in \mathrm{M}$. Prove that f_d and f^d belong to M for $d \in \mathbb{R}$.

(c) Let $f\colon \mathbb{R} \to \mathbb{R}$ be such that $f'(x)$ exists a.e. Use 21.5(b) to prove that $f' \in \mathrm{M}$.

21.2 (a) Let $f\colon \mathbb{R} \to \mathbb{R}^+$. Prove that $f \in \mathrm{M}$ if and only if $f \wedge g$ is integrable for every integrable function g.

(b) Let $f\colon \mathbb{R} \to \mathbb{R}$. Prove that $f \in \mathrm{M}$ if and only if $(f \wedge g) \vee (-g)$ is integrable for every integrable function g.

21.3 Let f be a continuous function and g a bounded measurable function. Prove that the composite function $f \circ g$ is measurable. [It is not sufficient to assume merely that f is measurable.]

21.4 [A technical trick, needed below.] Let g be a bounded measurable function, with $|g| \leqslant M$ on $\mathbb{R}$. Prove that there exists a sequence $\{\psi_n\}$ of step functions with $\psi_n \to g$ a.e. and $|\psi_n| \leqslant M$ for all n.

21.5 Let f be a continuous function on $\mathbb{R}$ which is convex (see Exercise 5.11 for the definition). Let g be a bounded integrable function on $\mathbb{R}$. Prove that

$$f\Big(\int_0^1 g(x)\,\mathrm{d}x\Big) \leqslant \int_0^1 f(g(x))\,\mathrm{d}x.$$

[Do it first for g a step function, and then use the two preceding exercises.]

21.6 Prove that the following are equivalent for $f \in \mathrm{L}_{\mathbb{R}}$:

(i) $f = 0$ a.e.;
(ii) $\int_{-\infty}^{u} f = 0$ for all $u \in \mathbb{R}$;
(iii) $\int f\varphi = 0$ for all $\varphi \in \mathrm{L}^{\mathrm{step}}$;
(iv) $\int fg = 0$ for all bounded measurable functions g;
(v) $\int |f| = 0$.

[Recall Theorem 14.7 and Exercise 21.4.]

21.7 Prove that a function $f\colon \mathbb{R} \to \mathbb{R}$ is measurable if and only if it is the limit a.e. of a sequence of continuous functions.

22 Measurable sets

This chapter investigates the notion of the length, or measure, of a subset S of $\mathbb{R}$. Many texts on integration take the notion of measure as fundamental, and develop the theory of integrals from a theory of measure. We do quite the opposite, using what we have already learnt about intervals and integrals to give a meaningful concept of length for as wide a class of sets as possible.

22.1 Definition. Let S be a subset of $\mathbb{R}$. Then S is (*Lebesgue*) *measurable* if χ_S is a measurable function, and is *non-measurable* otherwise. If S is measurable we define the *measure*, $m(S)$, of S by

$$m(S) := \begin{cases} \int \chi_S & \text{if } \chi_S \in \mathrm{L}, \\ \infty & \text{otherwise.} \end{cases}$$

In this context we use ∞ as a pseudo real number as explained in 21.11, adopting the conventions $x < \infty$ and $x + \infty = \infty + x = \infty + \infty$ for all $x \in \mathbb{R}$.

Measure captures the intuitive idea of length insofar as we have defined it.

22.2 Length vs. measure.

(a) Assume that S is the disjoint union of finitely many intervals $I_i := \langle a_i, b_i \rangle$ ($-\infty < a_i < b_i < \infty$, $i = 1, \dots, k$). Then

$$m(S) = \ell(S) := \sum_{i=1}^{k} (b_i - a_i).$$

(b) Let $S \subseteq \mathbb{R}$. Then S is null if and only if S is measurable and $m(S) = 0$.

Proof. For (a) see 4.2.

Now consider (b). We have declared 'S is null' and '$\chi_S = 0$ a.e.' to be synonymous, and defined our integral in such a way that $\chi_S = 0$ a.e. implies that $\int \chi_S$ exists and equals 0. In these circumstances S is measurable, with $m(S) = 0$. For the converse we invoke 14.7: for the non-negative integrable function χ_S we have $\int \chi_S = 0$ only if $\chi_S = 0$ a.e. $\square$

22.3 Combining measurable sets.

(a) Let S and T be measurable sets. Then the following are measurable:

$$S \cap T, \quad S \cup T, \quad S \setminus T, \quad S \triangle T, \quad \mathbb{R} \setminus S.$$

(b) Let $\{S_k\}$ be a sequence of measurable sets. Then $\bigcap_{k=1}^{\infty} S_k$ and $\bigcup_{k=1}^{\infty} S_k$ are measurable.

Proof. For (a) we use 21.5 and the following formulae:
$\chi_{S\cup T} = \chi_S \vee \chi_T$, $\chi_{S\cap T} = \chi_S \wedge \chi_T$, $\chi_{S\setminus T} = (\chi_S - \chi_T)^+$, $\chi_{S\,\Delta\,T} = |\chi_S - \chi_T|$.
For the complement, just note that $\mathbb{R}$ is measurable.

The second assertion in (b) follows from the first, using de Morgan's laws and the fact that the complement of a measurable set is measurable. So consider $S := \bigcap_{k=1}^{\infty} S_k$ and let $f_n = \chi_{T_n}$ where $T_n := \bigcap_{k=1}^{n} S_k$ $(n \geqslant 1)$. For any n we have $f_n \in \mathrm{M}$ by (a), and $\chi_S = \lim f_n$, which is measurable by 21.5. □

The results in 22.3 let us enlarge our repertoire of measurable sets.

22.4 Open sets, closed sets, and Borel sets. In topological parlance a subset U of $\mathbb{R}$ is *open* (in the usual topology on $\mathbb{R}$) if $U = \varnothing$ or U is a union of (non-empty) open intervals; it is *closed* if $\mathbb{R} \setminus U$ is open. Any open interval (a, b) is open in the topological sense, and any closed interval $[a, b]$ is closed.

By definition any interval is measurable. Note that we would not expect an **arbitrary** union of measurable sets to be measurable. (If this were true, **every** set E would be measurable, because $E = \bigcup\{\,\{x\} \mid x \in E\,\}$. The existence of non-measurable sets is discussed in 22.8.) Exercise 10.8 establishes that every non-empty open set may be expressed as a **countable** union of open intervals. Therefore any open set is measurable. By 21.5, any closed set is measurable.

The union of any family $\{U_\imath\}_{\imath\in I}$ of open sets in $\mathbb{R}$ is open, and, by de Morgan's laws, the intersection of a family $\{V_\imath\}_{\imath\in I}$ of closed sets is closed. However an infinite intersection—even a countable intersection—of open sets need not be open. A simple example is $\bigcap_{n=1}^{\infty}(-1/n, 1/n)$, which is the non-open set $\{0\}$. Similarly we may go outside the class of closed sets when we take infinite unions. The class of *Borel sets* is the smallest family of subsets of $\mathbb{R}$ closed under countable unions, countable intersections, and complements, and containing the open intervals (see Exercise 22.2 for other descriptions). Every Borel set is measurable, by 22.3. Exercise 22.10 shows that the converse is true 'modulo null sets'.

We now refine 22.3, considering the measures of the sets involved there.

22.5 Properties of measure.
(a) Let S and T be measurable. Then $S \subseteq T$ implies $m(S) \leqslant m(T)$.
(b) Let S and T be measurable. Then
$$m(S \cup T) + m(S \cap T) = m(S) + m(T).$$
(c) Let $\{S_k\}$ be a sequence of measurable sets. Then
$$m\Big(\bigcup_{k=1}^{\infty} S_k\Big) \leqslant \sum_{k=1}^{\infty} m(S_k),$$
with equality if the sets S_k are pairwise disjoint (*countable additivity*).
(d) For $d \in \mathbb{R}$ and S measurable, $S + d$ is measurable and $m(S + d) = m(S)$.

Proof. For (a) note that $S \subseteq T$ implies $\chi_S \leqslant \chi_T$. We now separate cases. If χ_T is integrable then so is χ_S, by 21.3, and hence $m(S) \leqslant m(T)$ by Property (P).

If χ_T is not integrable then $m(T) = \infty$ and certainly $m(S) \leqslant m(T)$. Part (d) is treated similarly, using Property (T).

Now consider (b). By (a) we have $m(S \cap T) \leqslant m(S) \leqslant m(S \cup T)$ and $m(S \cap T) \leqslant m(T) \leqslant m(S \cup T)$, so that if one side of the equation in (b) has value ∞ then so does the other, with this occurring if and only if $m(S) = \infty$ or $m(T) = \infty$. If both $m(S)$ and $m(T)$ are finite then we integrate the equation $\chi_{S\cup T} + \chi_{S\cap T} = \chi_S + \chi_T$ to obtain from Property (L) the required relationship between measures.

For (c) we first observe that there is nothing to prove unless $\sum m(S_k)$ is finite. In that case we can apply the MCT with $f_n := \chi_{S_1} + \cdots + \chi_{S_n}$, to obtain

$$\int \sum_{k=1}^{\infty} \chi_{S_k} = \int \lim f_n = \lim \int f_n = \sum_{k=1}^{\infty} m(S_k).$$

Also, since $\chi_S \leqslant \lim f_n$, we have $m(S) := \int \chi_S \leqslant \int \sum_{k=1}^{\infty} \chi_{S_k}$. The stated inequality follows. Finally, if the sets S_k are disjoint then we have $\chi_S = \lim f_n$. □

We can establish links between measurable functions and measurable sets, thereby illuminating both concepts.

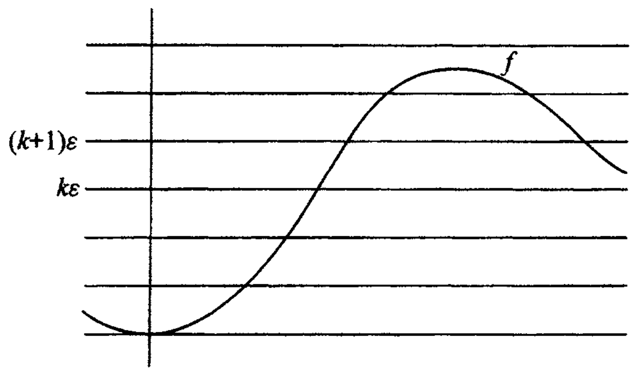


Figure 22.1

22.6 Horizontal slicing. Let $f\colon \mathbb{R} \to \mathbb{R}$. Observe that for any $\varepsilon > 0$, the sets

$$S_k := \{\, x \in \mathbb{R} \mid k\varepsilon \leqslant f(x) < (k+1)\varepsilon \,\} \quad (k \in \mathbb{Z})$$

are disjoint, and $\bigcup_{k\in\mathbb{Z}} S_k = \mathbb{R}$. Now assume f is bounded, say $|f| \leqslant K$. In that case, for any $N \in \mathbb{N}$ with $N \geqslant K/\varepsilon$ we have $\mathbb{R} = \bigcup\{S_k \mid -N \leqslant k \leqslant N\}$. Define

$$\varphi := \sum_{k=-N}^{N} (k\varepsilon)\chi_{S_k}.$$

Then for all $x \in \mathbb{R}$ we have

$$0 \leqslant f(x) - \varphi(x) < \varepsilon.$$

Thus we can approximate f uniformly by φ, achieving any desired degree of accuracy by choosing ε suitably small. So long as φ is integrable we might reasonably regard $\int \varphi$ as an approximation to $\int f$, obtained by slicing the area under the graph of f into suitably narrow horizontal strips; see Fig. 22.1. As we explain in Appendix I, this was the approach originally taken by Lebesgue in defining his integral. It contrasts with our step function approximations, in which we sliced the area up into vertical strips. As a step towards reconciling the two approaches we investigate measurability of the sets and functions involved.

22.7 Proposition. Let $f\colon \mathbb{R} \to \mathbb{R}$. Then f is measurable if and only if

$$S^c := \{\, x \in \mathbb{R} \mid f(x) \geqslant c \,\}$$

is measurable for all $c \in \mathbb{R}$. This is also valid when any of $<$, $\leqslant$, or $>$ is substituted for $\geqslant$.

Proof. Assume S^c is measurable for all $c \in \mathbb{R}$. It is easy to see that, for any $d \in \mathbb{R}$,

$$\{\, x \in \mathbb{R} \mid f(x) < d \,\} = \mathbb{R} \setminus S^d \quad \text{and} \quad \{\, x \in \mathbb{R} \mid f(x) > d \,\} = \bigcup_{n \geqslant 1} S^{d+1/n},$$

so these sets are measurable too. If f is bounded, then the function φ defined in 22.6 is measurable. Taking $\varepsilon = 1/m$ for $m = 1, 2, \dots$, we find a sequence $\{\varphi_m\}$ in M such that $0 \leqslant f(x) - \varphi_m(x) < 1/m$. This means $f = \lim \varphi_m$, and hence f is measurable by 21.5(b). If f is not bounded then we consider the bounded functions $g_k := (f \wedge k) \vee (-k)$ $(k = 1, 2, \dots)$. Clearly

$$\{\, x \in \mathbb{R} \mid g_k(x) \geqslant c \,\} = \begin{cases} \varnothing & \text{if } c > k, \\ S^c & \text{if } -k < c \leqslant k, \\ \mathbb{R} & \text{if } c \leqslant -k, \end{cases}$$

and so is measurable. By the result already proved for bounded functions, each g_k is measurable, and hence their limit, namely f, is measurable too.

Conversely, assume f is measurable and define, for $c \in \mathbb{R}$,

$$f_n = n\left((f \wedge c) - (f \wedge (c - 1/n))\right).$$

Each f_n is measurable, by 21.5(a). We have $f_n(x) = 1$ for any n if $f(x) \geqslant c$, while if $f(x) < c$ we can find N such that $f(x) < c - 1/N$, so that $f_n(x) = 0$ for $n \geqslant N$. Hence $f_n \to \chi_{\{x \mid f(x) \geqslant c\}}$, and this limit is measurable by 21.5(b). □

22.8 A non-measurable set. We can present a non-measurable set E as follows. We partition the interval $[0, 1]$ into a union of disjoint subsets $\{X_\lambda\}_{\lambda \in \Lambda}$ by asserting that x, y lie in the same set X_λ if and only if $x - y$ is rational—that is, we take the partition associated with the equivalence relation $\sim$, where $x \sim y \iff x - y \in \mathbb{Q}$. We let E be a set containing exactly one point from each set X_λ, and assume for a contradiction that E is measurable. We enumerate the rational numbers in $[0,1]$ as $q_1, q_2, \ldots$ and define $E_n := E + q_n$. Then

(a) each E_n is measurable, with $m(E_n) = m(E)$;

(b) the sets $E_1, E_2, \ldots$ are mutually disjoint (because by construction E does not contain two points differing by a non-zero rational);

(c) $[0,1] \subseteq \bigcup_{n\geqslant 1} E_n \subseteq [0,2]$.

Let $F := \bigcup E_n$. By (b) and countable additivity, 22.5(b), $m(F) = \sum m(E_n)$. By (a), this forces $m(E) = 0$ or $m(F) = \infty$. But neither of these is possible, by (c).

In the argument above we specify E by picking out one point from each set X_λ. While this may appear a reasonable procedure, it cannot be **proved** to be possible unless we work with a suitable formulation of set theory. The mathematical theory of sets is based on a list of axioms designed to capture the set-theoretic concepts and constructions we regularly use: membership, union, power set, etc. A limited list of core axioms—for example the Zermelo–Fraenkel axioms—suffices to formalize the elementary properties of sets. To delve deeper one must add more axioms. One such 'optional extra' is the *Axiom of Choice*, known as (AC). This states that for any non-empty family $\{Y_i\}_{i\in I}$ of non-empty sets there is a set containing one element from each Y_i. Thus (AC) permits us to specify our set E. Therefore in a mathematical world with (AC) a non-measurable set E exists. Further, χ_E will be a non-measurable function.

Just as we can model the mathematical world with core set-theoretic axioms plus (AC), it is equally possible to have 'anti-worlds' in which the core axioms and the negation of (AC) are postulated. It is possible to show (under one additional set-theoretic assumption concerning cardinal numbers) that there exists such an anti-world in which there are no non-measurable sets; this was proved by R. Solovay in a paper published in the *Annals of Mathematics* in 1970. These remarks reinforce the point that non-measurable sets and functions cannot be exhibited by using just the machinery of everyday mathematics.

Exercises

22.1 Let $S \subseteq \mathbb{R}$ and let N be null. Prove that $S \cup N$ is measurable if and only if S is measurable.

22.2 Show that the family of Borel sets is the smallest family of subsets of $\mathbb{R}$ closed under countable unions, countable intersections, and complements, and containing $\mathcal{E}$, when $\mathcal{E}$ consists of (i) all closed intervals, (ii) all intervals $(a,b]$ $(-\infty < a < b < \infty)$, or (iii) all intervals.

22.3 [For those who know some set theory] Let c denote the cardinality of $\mathbb{R}$.
   (a) Show that the set of all Borel sets in $\mathbb{R}$ has cardinality 2^c.
   (b) Show that the set of all null sets has cardinality 2^{2^c}. [Hint: consider subsets of the Cantor set.]
   (c) Deduce that not every Lebesgue measurable set is a Borel set.

22.4 Prove that $f\colon \mathbb{R} \to \mathbb{R}$ is measurable if and only if $f^{-1}(A)$ is a measurable set for every open subset A in $\mathbb{R}$, or equivalently for every Borel set A.

22.5 Let $\{S_n\}$ be a sequence of measurable sets such that $S_n \supseteq S_{n+1}$ for all n. Prove that $m(\bigcap S_n) = \inf_n m(S_n)$.

22.6 Imitating the construction of the Cantor set in 10.9, define

$$F_0 := [0,1], \quad F_{n+1} := F_n \cap (1/4F_n \cup (1/4F_n + 3/4)) \ (n \geqslant 0).$$

(a) Prove that F_n is the union of 2^{n-1} disjoint closed intervals each of length 4^{-n}.

(b) Define $F := \bigcap_{n=1}^{\infty} F_n$. Prove that F is a measurable set with $m(F) = 1/2$ and that F contains no non-empty open interval.

22.7 Adapt the construction of the preceding exercise to exhibit a measurable subset C_r of $[0,1]$ with $m(C_r) = 1 - r^{-1}$ and such that C_r contains no non-empty open interval.

22.8 [The Borel–Cantelli Lemmas] Let $\{E_n\}$ be a sequence of measurable subsets of $[0,1]$. Define

$$E := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n.$$

(a) Assume that $\sum m(E_n)$ converges. Prove that $m(E) = 0$.

(b) Assume that $E_1, E_2, \ldots$ are disjoint and that $\sum m(E_n)$ diverges. By considering $\log(m(\bigcap_{n=k}^{p}([0,1] \setminus E_n))\ (p \geqslant k)$ prove that $m(E) = 1$.

22.9 Let E be measurable and bounded. Let $\{g_n\}$ be a sequence in $\mathrm{L}^{\mathrm{step}}$ with $g_n \xrightarrow{\text{a.e.}} \chi_E$, where all g_n are zero outside some fixed bounded set. Define

$$h_n(x) := \begin{cases} 1 & \text{if } g_k(x) > 1/2 \text{ for some } k \geqslant n, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Prove that $\{h_n\}$ is a decreasing sequence of integrable functions, and that $h_n \xrightarrow{\text{a.e.}} \chi_E$.

(b) Deduce that, for a given $\varepsilon > 0$, there exists a measurable set G containing E such that G is a countable union of open intervals and $m(G \setminus E) < \varepsilon$. [Thus E can be approximated in measure arbitrarily closely from the outside by open sets.]

22.10 [Continuation of Exercise 22.9]

(a) Let E be an arbitrary measurable subset of $\mathbb{R}$ and let $\varepsilon > 0$. Show that there exists an open set $G \supseteq E$ such that $m(G \setminus E) < \varepsilon$. By considering $\mathbb{R} \setminus E$, show that there exists a closed set $F \subseteq E$ with $m(E \setminus F) < \varepsilon$. Deduce that there exist a countable intersection of open sets, U, and a countable union of closed sets, V, such that $E \triangle U$ and $E \triangle V$ are null.

(b) Show that a subset E of $\mathbb{R}$ is measurable if and only if E is expressible as $B \triangle N$, where B is a Borel set and N is null.

23 The character of integrable functions

This chapter contains various results which we could have scattered through the text, where the theorems needed to prove them first appeared. Because of their usefulness we have chosen to bring them together for reference, along with some related results already proved. Unless otherwise stated, functions may be real- or complex-valued.

23.1 Zero integrals. Let $f \in \mathrm{L}_{\mathbb{R}}$, $f \geqslant 0$, and $\int f = 0$. Then

(a) $f \equiv 0$ if f is continuous at every point;

(b) $f = 0$ a.e.

The first assertion was proved for functions in L^{C} in 5.2(d). The elementary proof given there applies here too. The second part was derived from the MCT in 14.7. Note that if f is continuous on $\mathbb{R}$ and zero a.e. then $f \equiv 0$ (Exercise 11.1), so (a) follows from (b) (circuitously!).

We next collect together a group of results involving products, which we could not conveniently consider before we had studied measurable functions. In general the product of two integrable functions is **not** integrable: $x^{-1/2}$ is integrable on $(0,1]$, but $x^{-1/2} \cdot x^{-1/2} = x^{-1}$ is not. However we do have the following positive results.

23.2 Integrable products.

(a) Let f be integrable and let g be bounded and measurable. Then fg is integrable.

(b) Let f and g be measurable and such that f^2 and g^2 are integrable. Then fg is integrable and the **Cauchy–Schwarz inequality**, (CS),

$$\left|\int fg\right|^2 \leqslant \int |f|^2 \int |g|^2$$

holds.

(c) Let f be measurable on the bounded interval I and assume $f^2 \in \mathrm{L}(I)$. Then $f \in \mathrm{L}(I)$.

Proof. Note that f, g measurable implies fg measurable, by 21.5. In (a) there exists by hypothesis a finite constant M such that $|fg| \leqslant M\,|f|$, whence the result follows by the Comparison Theorem, 21.3.

For (b) first note that

$$2\,|fg| \leqslant |f|^2 + |g|^2$$

(because $(|f| \pm |g|)^2 \geqslant 0$) and deduce from the Comparison Theorem that fg is integrable. We obtain (c) from this by taking g to be the constant function 1, which is integrable on any bounded interval.

The Cauchy–Schwarz inequality can be proved very simply in the real case by considering $(f \pm g)^2$, as suggested for the L^{C} version in Exercise 5.3. For the complex case we refer to 31.10, where the result is set in its proper context—the theory of inner product spaces. □

23.3 Approximations to integrable functions. Fix $\varepsilon > 0$. Let f be integrable. We have already proved in 12.13 that we can approximate f by a step function φ so that the error, as measured by $\int |f - \varphi|$, is less than ε. For certain applications we need to approximate f in the same sense by a function g drawn not from $\mathrm{L}^{\mathrm{step}}$ but from some other special class, $\mathcal{E}$ say, of integrable functions. The proof strategy is first to approximate f by a step function φ and then to modify φ to a function $g \in \mathcal{E}$, keeping the discrepancy small. We have already presented two useful results of this type.

(a) Given an integrable function f there exists a continuous function g of compact support such that $\int |f - g| < \varepsilon$ (use Exercise 3.6).
(b) Given an integrable function f there exists a C^∞ function g of compact support such that $\int |f - g| < \varepsilon$ (use 11.13).

The following result is important in the theory of Fourier series.

23.4 The Riemann-Lebesgue Lemma. Let f be an integrable function. Then $f(x) \sin nx$ is integrable and

$$\int f(x) \sin nx \, \mathrm{d}x \to 0 \text{ as } n \to \infty.$$

The same conclusion holds if $\sin nx$ is replaced by $\cos nx$ or $\mathrm{e}^{\pm inx}$ or the discrete variable n is replaced by a continuous variable $\lambda \to \infty$ or $-\infty$.

Proof. it is surprisingly awkward to prove by elementary means that $f(x) \sin nx$ is integrable (try it!). Instead we appeal to 23.2(a), using the fact that a continuous function is measurable.

The convergence theorems are no help to us in establishing the limit because the integrand may not tend to 0. Instead (for once!) we use step functions.

Stage 1. Consider $\varphi := \chi_I$ where I is a bounded interval $\langle a, b\rangle$. Then

$$\left| \int_a^b \varphi(x) \sin nx \, \mathrm{d}x \right| = \left| \frac{\cos na - \cos nb}{n} \right| \leqslant \frac{2}{n}.$$

Hence the result holds for characteristic functions of bounded intervals.

Stage 2. The result holds for step functions by linearity.

Stage 3. Let $f \in \mathrm{L}$. By 12.13, given $\varepsilon > 0$, there exists a complex step function φ such that $\int |f - \varphi| < \varepsilon$. Then

$$\begin{aligned}\left|\int f(x)\sin nx \,\mathrm{d}x\right| &\leqslant \left|\int (f(x)-\varphi(x))\sin nx\,\mathrm{d}x\right| + \left|\int \varphi(x)\sin nx\,\mathrm{d}x\right| \\ &\leqslant \int |f(x)-\varphi(x)|\,\mathrm{d}x + \left|\int \varphi(x)\sin nx\,\mathrm{d}x\right| \\ &\leqslant \varepsilon + \varepsilon,\end{aligned}$$

if n is sufficiently large, by choice of φ and Stage 2.

The final assertions are proved in the same manner. □

23.5 Exercise example. Here is a continuity result associated with the integral. We use it in 30.6 and 33.13.

(a) Show that $\lim_{h\to 0}\int|\varphi(x+h)-\varphi(x)|\,\mathrm{d}x = 0$ for any $\varphi \in \mathrm{L}^{\mathrm{step}}$.
(b) Deduce that $\lim_{h\to 0}\int|f(x+h)-f(x)|\,\mathrm{d}x = 0$ for any $f \in \mathrm{L}$.

23.6 Boundedness vs. integrability. Before proceeding, review the comments on bounded functions in 2.11. Now compare the following statements about a real- or complex-valued measurable function f:

(a) there exists $M \in \mathbb{R}$ such that $|f(x)| \leqslant M$ for all $x \in \mathbb{R}$ (f is bounded);
(b) there exists a null set E and $M \in \mathbb{R}$ such that $|f(x)| \leqslant M$ for all $x \notin E$ (f is *essentially bounded*, in the sense that there exists a bounded function g such that $f = g$ a.e.);
(c) given $\delta > 0$ there exists a measurable set G with $m(G) < \delta$ such that f is bounded on $\mathbb{R} \setminus G$.

Obviously (a) $\implies$ (b) $\implies$ (c). Let us now look at some examples.

(1) Let $f(n) = n$ if $n \in \mathbb{N}$ and $f(x) = 0$ for $x \notin \mathbb{N}$. This function is zero a.e. and so is integrable with zero integral. It is clearly not bounded, though it is essentially bounded.
(2) Take $f(x) = x^{-1/2}\chi_{(0,1]}(x)$, which we know is integrable. Certainly $f(x) \nearrow \infty$ as $x \to 0+$ so f is not bounded. For any M, we have $|x^{-1/2}| > M$ if and only if $x \in (0, 1/M^2)$. If (b) were true, we would have $(0, 1/M^2)$ null, in contradiction to 10.10. On the other hand (c) is true, since the interval $(0, 1/M^2)$ has measure $< \delta$ if we choose $M > 1/\delta^{1/2}$.
(3) Take $f(x) = x^{-1}\chi_{(0,1]}(x)$. Then f is not integrable, f is unbounded, and (c) holds.

It is all too commonly asserted that an integrable function f must be bounded—presumably on the mistaken belief that this is necessary to ensure that the area under the graph of $|f|$ is finite. Example (1) shows this is not so. Confronted with this evidence, a muddled thinker will usually back off, and claim instead that an integrable function is bounded off a null set, that is, essentially bounded—still wrong, as the example in (2) proves. We stress: integrability does not imply boundedness, not even off some null set.

By contrast, (c) is true for any integrable function f, as can be proved by applying the downward MCT to $\{\chi_{E_n}\}$, where $E_n := \{x \in \mathbb{R} \mid |f(x)| > n\}$ (do it!). Example (3) shows that (c) is not a sufficient condition for a measurable function to be integrable. Together, (2) and (3) indicate that it is rate of growth not boundedness/unboundedness that determines whether or not a measurable function on a bounded interval is integrable.

There is another sense besides (c) in which an integrable function f is 'nearly bounded'. We know that for each $\varepsilon > 0$ there exists a step function φ such that $\int |f - \varphi| < \varepsilon$. We may interpret this informally as saying that, to within ε, the area under the graph of f is concentrated in some bounded region of the plane, since φ is bounded and vanishes outside a compact interval.

23.7 Behaviour near infinity. Paralleling the preceding section, we may quickly counter the erroneous claim that f integrable implies that $f(x) \to 0$ as $|x| \to \infty$ by considering $f = \chi_{\mathbb{N}}$. Contrary to popular belief even a continuous integrable function need not tend to 0 at infinity; see Exercise 15.7.

In the positive direction, Exercise 1.9 outlined a proof, relying only on the Basic Properties, that if $f\colon \mathbb{R}^+ \to \mathbb{R}^+$ is integrable and decreasing, then $xf(x) \to 0$ as $x \to \infty$—a stronger conclusion than $f(x) \to 0$. (In fact, if the condition that f is integrable is replaced by the condition that $x^p f(x)$ is integrable then $x^{p+1} f(x) \to 0$ as $x \to \infty$ $(p > 0)$.)

A different positive result is particularly useful.

23.8 Proposition. Assume that f and f' are integrable and that f' is continuous. Then $f(x) \to 0$ as $|x| \to \infty$.

Proof. Suppose for a contradiction that $f(x) \not\to 0$ as $x \to \infty$. This means that there exists $\varepsilon > 0$ such that for any X there exists $Y > X$ for which $|f(Y)| \geqslant \varepsilon$. Also $\left|\int_X^Y f'\right| < \frac{1}{2}\varepsilon$ if X is sufficiently large, by 16.1 applied to f'. By the FTC, $\int_X^Y f' = f(Y) - f(X)$. Hence for X sufficiently large we can choose $Y > X$ so that

$$\varepsilon - |f(X)| \leqslant |f(Y)| - |f(X)| \leqslant |f(Y) - f(X)| = \left|\int_X^Y f'\right| < \frac{1}{2}\varepsilon.$$

But this says that $|f(X)| > \frac{1}{2}\varepsilon$ for all X sufficiently large. Then we have the required contradiction because, for Z sufficiently large,

$$\frac{\varepsilon Z}{2} \leqslant \int_Z^{2Z} |f| \leqslant \int |f|. \qquad \square$$

Exercises

23.1 Give an example of

(a) a measurable function f such that f is integrable but f^2 is not;
(b) a measurable function f such that f^2 is integrable but f is not.

23.2 Let f be integrable. Prove that

$$\lim_{n\to\infty}\int_{-n}^{n}\Big(1-\frac{|x|}{n}\Big)f(x)\,\mathrm{d}x=\int_{-\infty}^{\infty}f(x)\,\mathrm{d}x.$$

23.3 Let f be a function that is periodic with period $T>0$ and assume that f is integrable on $[0,T]$. Prove that $\mathrm{e}^{-\alpha x}f(x)$ is integrable on $[0,\infty)$ for any $\alpha>0$.

23.4 Let $g\in\mathrm{L}$ be such that $fg\in\mathrm{L}$ for all $f\in\mathrm{L}$. Prove that there exists $M\in\mathbb{R}$ such that $|g|\leqslant M$ a.e. [The converse of 21.3, and harder.]

23.5 (a) Use Exercise 2.11 and the Riemann–Lebesgue Lemma to prove that

$$\lim_{n\to\infty}\int_0^{\pi/2}(\cot x-x^{-1})\sin 2nx\,\mathrm{d}x=0.$$

(b) Deduce that $\displaystyle\lim_{n\to\infty}\int_0^{\pi/2}\frac{\sin 2nx}{x}\,\mathrm{d}x=\pi/2.$

(c) Deduce that $\displaystyle\lim_{X\to\infty}\int_0^{X}\frac{\sin x}{x}\,\mathrm{d}x=\pi/2.$

(d) By using integration by parts prove that $\displaystyle\int_0^{\infty}\frac{\sin^2 x}{x^2}\,\mathrm{d}x=\pi/2.$

23.6 (a) Use the result of Exercise 6.8 to prove that, for $0<\alpha<\beta<\infty$,

$$\int_\alpha^\beta \sin(nx^2)\,\mathrm{d}x\to 0\quad\text{as } n\to\infty.$$

(b) Deduce that if $\varphi\in\mathrm{L}^{\text{step}}$ is such that $\varphi=0$ outside $[\delta,\infty)$ for some $\delta>0$, then

$$\int\varphi(x)\sin(nx^2)\,\mathrm{d}x\to 0\quad\text{as } n\to\infty.$$

(c) Let f be integrable on $(0,\infty)$. Prove that, given $\varepsilon>0$, there exists $\delta>0$ and a step function φ such that $\varphi=0$ outside $[\delta,\infty)$ and $\int|f-\varphi|<\varepsilon$. Deduce that

$$\int_0^\infty f(x)\sin(nx^2)\,\mathrm{d}x\to 0\quad\text{as } n\to\infty.$$

23.7 Let $-\infty<a<b<\infty$. Formulate and prove an analogue of the Riemann-Lebesgue Lemma for the limit

$$\lim_{n\to\infty}\int_a^b f(x)\,|\sin nx|\,\mathrm{d}x\qquad(f\in\mathrm{L}[a,b]).$$

24 Integration vs. differentiation

At the outset we asserted that differentiation and integration are mutually inverse processes. We now reveal the definitive answers to the following fundamental questions.

- When is a derivative integrable?
- When is an indefinite integral differentiable?
- When does $F(x) - F(a) = \int_a^x F'(t)\,dt$ hold?
- Which functions arise as indefinite integrals?

The theory involved in providing the answers is far from elementary, and scarcely appropriate for a book entitled "Introduction to integration". Accordingly we give an account for spectators rather than for would-be professionals. We draw heavily on earlier results and exercises, give only outline proofs, and cite without proof certain theorems from outside integration theory, concerning increasing functions and functions of bounded variation. We work throughout with real-valued functions on a compact interval $[a, b]$.

We earlier noted in passing that a monotonic function has only a countable set of discontinuities, and so is continuous a.e. We record below a very much deeper result on differentiability, due to Lebesgue. It was first proved in a measure-theoretic way; see for example **[14]** or **[10]**. Since integration enters into the statement only through an 'a.e.', which can be rephrased in terms of open intervals, it is natural to ask whether there is a proof based solely on differential calculus. Such a proof was discovered by Riesz. It can be found in **[17]**.

24.1 Theorem. A monotonic function on $[a, b]$ is differentiable a.e.

Remember that—perhaps surprisingly—not every derivative is Lebesgue integrable: we showed in 13.7 that $x^2 \sin(x^{-2})\chi_{(0,1]}(x)$ is not in $L_{\mathbb{R}}[0, 1]$ (although it does have an improper integral). Here is a positive result.

24.2 Integrable derivatives. Let $G\colon [a, b] \to \mathbb{R}$.
(a) If G' exists a.e. and is bounded then $G' \in L_{\mathbb{R}}[a, b]$;

(b) If G is increasing then G' exists a.e., $G' \in L_{\mathbb{R}}[a, b]$, and $\int_a^b G' \leqslant G(b) - G(a)$.

Proof. (a) Put $G(x) = G(b)$ for $x > b$ and let $g := G'$. We may write

$$g(x) = \lim_{n\to\infty} f_n(x) \text{ a.e.} \quad \text{where } f_n(x) = n(G(x + n^{-1}) - G(x)).$$

We apply the DCT to $\{f_n\}$; note that by the MVT, $|f_n| \leqslant M := \sup |G|'$, and M is integrable on $[a,b]$. Hence g is integrable.

Now consider (b). Theorem 24.1 tells us that G' exists a.e. Since G is increasing, $g := G' \geqslant 0$ a.e. By 21.5(b), g is measurable. We now apply Fatou's Lemma, 21.12, to get

$$\begin{aligned}\int_a^b g \leqslant \liminf \int f_n &= \liminf \int_a^b n(G(x+n^{-1}) - G(x))\,\mathrm{d}x \\ &= \liminf \Big(n\int_b^{b+1/n} G(x)\,\mathrm{d}x - n\int_a^{a+1/n} G(x)\,\mathrm{d}x\Big) \\ &\leqslant G(b) - G(a),\end{aligned}$$

because $g(x) = G(b)$ on $[b, b+1/n]$ and $G(x) \geqslant G(a)$ on $[a, a+1/n]$. This forces g to be integrable. $\square$

Our interest in monotonic functions is explained by the next observation.

24.3 Monotonicity and the indefinite integral. Let $f \in L_{\mathbb{R}}[a,b]$. If $f \geqslant 0$ then $\int_a^x f$ clearly increases as x increases. In general we can write

$$F(x) := \int_a^x f = \int_a^x f^+ - \int_a^x f^- \quad (x \in [a,b]),$$

and so express F as the difference of two increasing functions. We can apply 24.1 to each of these. So F is differentiable a.e. and its derivative is integrable. This is significant progress, though we do not yet know that $F' = f$ a.e. Before working towards this we collect together the evidence we already have.

24.4 The indefinite integral (recap). Let $f \in L_{\mathbb{R}}[a,b]$ and let $F(x) := \int_a^x f$. Then

(a) F is continuous (19.4);
(b) $F'(x)$ exists and equals $f(x)$ at any point x at which f is continuous (12.15);
(c) if $f \in L^{\text{step}}$ then $F'(x)$ exists a.e., the exceptional points being the discontinuity points of f (see Exercise 5.7).

We now present the first of two technical lemmas.

24.5 Lemma I. Let $f \in L_{\mathbb{R}}[a,b]$. Assume that $F(x) := \int_a^x f = 0$ for all $x \in [a,b]$. Then $f = 0$ a.e.
Proof. We write $[a,b]$ as the union of the disjoint sets on which $f = 0$, $f > 0$, and $f < 0$, and assume for a contradiction that, without loss of generality,

$$E := \{\, x \in [a,b] \mid f(x) > 0 \,\}$$

is not null. Certainly E is measurable, by 22.7, and $f\chi_E$ is integrable, by 23.2(a). By Exercise 22.10, we can approximate E from the inside by a closed set, V say, with $m(V) > 0$. Let $U := (\mathbb{R} \setminus V) \cap (a, b)$; this is an open set. Then, by the contrapositive of 14.7 applied to $f\chi_V$,

$$0 = \int_a^b f = \int_V f + \int_U f > \int_U f.$$

Use Exercise 10.8 to write U as the disjoint union of a countable collection of disjoint open intervals $\{(a_n, b_n)\}$. By the MCT for series,

$$\int_U f = \sum_n \int_{a_n}^{b_n} f.$$

There must exist m such that $\int_{a_m}^{b_m} f \neq 0$. By 12.10, $\int_a^{a_m} f \neq 0$ or $\int_a^{b_m} f \neq 0$. Either way this contradicts the hypothesis that $F \equiv 0$. □

The next lemma would be the theorem we are seeking were it not for the restriction to bounded functions.

24.6 Lemma II. Let $f \in \mathrm{L}_{\mathbb{R}}[a, b]$, with f bounded. Assume that

$$F(x) - F(a) = \int_a^x f(t)\,\mathrm{d}t \quad (x \in [a, b]).$$

Then F is differentiable a.e., with $F'(x) = f(x)$ a.e.

Proof. By 24.3, $F'(x)$ exists for almost all $x \in [a, b]$. We must show that $F' = f$ a.e., which we do by showing its indefinite integral is zero, and then appealing to Lemma I.

Fix $c \in [a, b]$ and let $x \in [a, c]$. We can apply the Continuous DCT to $\{f_h\}$, where h is such that $x + h \in [a, c]$ and

$$f_h(x) := \frac{F(x+h) - F(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t.$$

Then $f_h \xrightarrow{\text{a.e.}} F'$. Note also that if $|f| \leqslant M$ then $|f_h| \leqslant M$ too, by Property (P), so M serves as the dominating function. Hence the conditions for the Continuous DCT are met and we deduce that

$$\begin{aligned}
\int_a^c F'(x)\,\mathrm{d}x &= \int_a^c \lim_{h\to 0} f_h(x)\,\mathrm{d}x \\
&= \lim_{h\to 0} \frac{1}{h}\int_a^c (F(x+h) - F(x))\,\mathrm{d}x \\
&= \lim_{h\to 0} \left(\frac{1}{h}\int_c^{c+h} F(x)\,\mathrm{d}x - \frac{1}{h}\int_a^{a+h} F(x)\,\mathrm{d}x\right) && \text{(by (T))} \\
&= F(c) - F(a) && \text{(by 5.4)} \\
&= \int_a^c f(x)\,\mathrm{d}x.
\end{aligned}$$

We conclude that $\int_a^c (F' - f) = 0$ for all $c \in [a, b]$, as required. □

We now show that the boundedness restriction in Lemma II can be removed.

24.7 Indefinite Integral Theorem III. Let $f \in \mathrm{L}_{\mathbb{R}}[a,b]$ and assume that

$$F(x) - F(a) = \int_a^x f(t)\,\mathrm{d}t \quad (x \in [a,b]).$$

Then F is differentiable a.e., with $F'(x) = f(x)$ a.e.

Proof. By considering positive and negative parts we may assume without loss of generality that $f \geqslant 0$. We don't know that f is bounded above so we truncate. Let $f_n := f \wedge n$ and let F_n be its indefinite integral. Then f_n is bounded and integrable and F_n is increasing. By 24.6, $F_n'(x) = f_n(x)$ for x outside some null set E_n. Define $E := \bigcup_{n\geqslant 1} E_n$; this is null. Let $G_n(x) := \int_a^x (f - f_n)$. Since $f - f_n \geqslant 0$, G_n is increasing, and has a derivative a.e., which is non-negative. So, by linearity,

$$F'(x) = G_n'(x) + f_n(x) \geqslant f_n(x) \quad \text{a.e.},$$

so that $\int_a^b F' \geqslant \int_a^b f_n$ for all n. By the MCT, $\int_a^b f_n \nearrow \int_a^b f$. Therefore

$$\int_a^b F' \geqslant \int_a^b f.$$

We proved the reverse inequality in 24.2(b). □

We should like to characterize those functions F for which there exists $f \in \mathrm{L}_{\mathbb{R}}[a,b]$ such that $F' = f$ a.e. We start from the fact that any such F is the difference of two increasing functions.

24.8 Functions of bounded variation. Let g be a real-valued function defined on a compact interval $[a,b]$. Then g is said to be of *bounded variation* if there exist increasing functions g_1 and g_2 such that $g = g_1 - g_2$. The terminology comes from an alternative characterization of the class $\mathrm{BV}[a,b]$ of functions of bounded variation on $[a,b]$. Let $a = \alpha_0 < \cdots < \alpha_N = b$ be a partition $\mathcal{P}$ of $[a,b]$ and let $g\colon [a,b] \to \mathbb{R}$ be any function. Define

$$V_g^{\mathcal{P}} := \sum_{i=0}^{N-1} |g(\alpha_{i+1}) - g(\alpha_i)|\,.$$

Then define the *total variation* of g on $[a,b]$ by

$$V_g := \sup\{\, V_g^{\mathcal{P}} \mid \mathcal{P} \text{ is a partition of } [a,b]\,\};$$

$V_g = \infty$ is a possibility here. Obviously V_g gives a measure of the extent to which the values of g jump up and down. As an example consider $g = \chi_{\mathbb{Q}}$: by taking partitions $\mathcal{P}$ with sufficiently many points chosen to be alternately rational and irrational we can make $V_g^{\mathcal{P}}$ arbitrarily large, so g is not of bounded variation. For a harder example, look at h as shown in Fig. 24.1: $h(x) = x\cos(1/x)$ if $x \neq 0$, $h(0) = 0$. Here h is continuous but $h \notin \mathrm{BV}[0,\pi]$. To check the latter claim take partitions whose points are alternately local maxima and minima of h. These

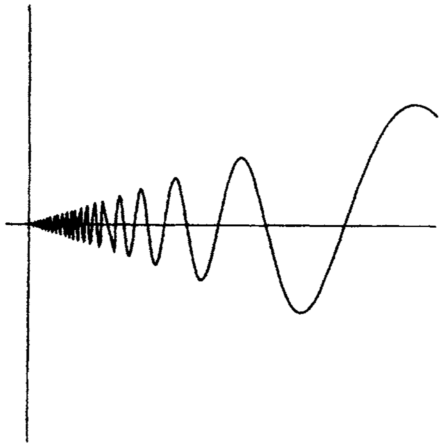


Figure 24.1

extrema occur at the points $\{\alpha_k\}$ where $\alpha := 1/(k\pi)$, and are maxima for k even and minima for k odd. Then

$$\sum_{k=1}^{n} |h(\alpha_k) - h(\alpha_{k+1})| = \sum_{k=1}^{n} \left(\frac{1}{k\pi} + \frac{1}{(k+1)\pi}\right) \nearrow \infty \quad \text{as } n \to \infty.$$

Now assume $g\colon [a,b] \to \mathbb{R}$ is monotonic. Then $V_g^{\mathcal{P}} = |g(b) - g(a)|$ for every partition $\mathcal{P}$, so that $V_g = |g(b) - g(a)|$. Also, by the triangle inequality $V_{g_1 \pm g_2} \leqslant V_{g_1} + V_{g_2}$ for any functions g_1 and g_2. Combining these observations we see that V_g is finite whenever g is the difference of two monotonic functions. Conversely it can be proved that if V_g is finite, then g can be expressed as $g_1 - g_2$ where g_1 and g_2 are increasing: one can take $g_1(x)$ to be the total variation of g on $[a,x]$, for $a \leqslant x \leqslant b$, and $g_2 := g_1 - g$ (see, for example, [14] for the details).

Every indefinite integral is continuous and of bounded variation. Our example h shows that not every continuous function can be represented as an indefinite integral. A natural conjecture might now be that the indefinite integrals on $[a,b]$ are exactly the functions in $\mathrm{C}[a,b] \cap \mathrm{BV}[a,b]$. This is still not true. What we need is a subset of this set consisting of functions satisfying a condition in which continuity and bounded variation are combined.

24.9 The indefinite integral re-examined. Let $F(x) := \int_a^x f$ on $[a,b]$. Given $\delta > 0$, choose $n \in \mathbb{N}$ and points a_i and b_i $(i = 1, 2, \ldots, n)$ so that

$$a \leqslant a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n \leqslant b \quad \text{and} \quad \sum_{i=1}^{n} (b_i - a_i) < \delta.$$

That is, we are considering an arbitrary finite collection of disjoint subintervals

(a_i, b_i) of $[a, b]$ of total length $\leqslant \delta$. Fix $\varepsilon > 0$. Now

$$\begin{aligned}\sum_{i=1}^{n} |F(b_i) - F(a_i)| &= \sum_{i=1}^{n} \left|\int_{a_i}^{b_i} f\right| \\ &\leqslant \sum_{i=1}^{n} \int_{a_i}^{b_i} (|f| - (|f| \wedge n)) + \sum_{i=1}^{n} \int_{a_i}^{b_i} (|f| \wedge n) \\ &\leqslant \int_a^b (|f| - (|f| \wedge n)) + n \sum_{i=1}^{n} (b_i - a_i).\end{aligned}$$

The first of these terms can be made $< \varepsilon/2$ by taking n sufficiently large (Exercise 14.5). Then, provided $2n\delta < \varepsilon$, the second term is also $< \varepsilon/2$. (We remark that the argument simplifies if f itself is bounded, so the truncation process is not needed.)

24.10 Absolute continuity. A function $G\colon [a, b] \to \mathbb{R}$ is *absolutely continuous*. if given $\varepsilon > 0$ there exists $\delta > 0$ such that $\sum_i |G(b_i) - G(a_i)| < \varepsilon$ for every finite collection of disjoint open subintervals (a_i, b_i) of $[a, b]$ of total length $< \delta$. Clearly every absolutely continuous function is continuous and of bounded variation (and hence differentiable a.e.). We proved above that every indefinite integral is absolutely continuous. The converse is true as well. For this it is enough to prove that an absolutely continuous function G is the indefinite integral of its derivative. This is done by showing that $G(x) - \int_a^x G'$ has zero derivative. We do not give the details (these can be found, for example, in [2]). Instead we present an instructive example.

24.11 A function based on the Cantor set. Refer back to 10.9. Every point x of the Cantor set, C, can be expressed uniquely as

$$x = \sum_{n=0}^{\infty} 3^{-n} a_n \quad \text{where } a_n \in \{0, 2\}.$$

Define $G\colon [0, 1] \to \mathbb{R}$ as follows. If $x \in C$, put

$$G(x) := \sum_{n=0}^{\infty} 2^{-n} b_n \quad \text{where } b_n = 0 \text{ if } a_n = 0 \text{ and } b_n = 1 \text{ if } a_n = 2.$$

Write $[0, 1] \setminus C$ as the union of disjoint open intervals, O_n, as in 10.9. If $x \notin C$, then define $G(x) = G(\alpha_x)$ where $x \in O_{n_x}$ and α_x is the left-hand endpoint of O_{n_x}. Thus the graph of G, insofar as we can depict it, is shown in Fig. 24.2. The following properties are not too difficult to check:

(i) G is an increasing continuous map of $[0, 1]$ onto $[0, 1]$;
(ii) G is not absolutely continuous;
(iii) $G' = 0$ a.e., the exceptional set being the null set C;
(iv) $\int_0^1 G' = 0 \neq 1 = G(1) - G(0)$.

In conclusion we record the following definitive result linking the concepts discussed in this chapter.

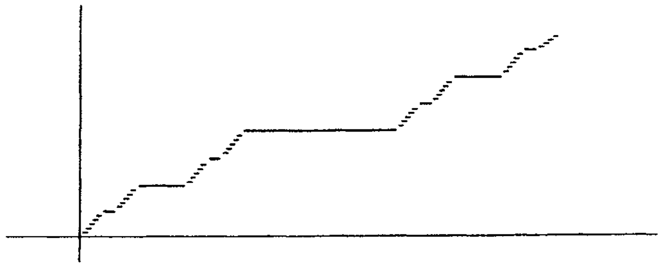


Figure 24.2

24.12 Theorem. The following are equivalent for a function $G\colon [a,b] \to \mathbb{R}$:

(i) G is absolutely continuous;

(ii) $G \in \mathrm{BV}[a,b]$ and $G(x) - G(a) = \int_a^x G'(t)\,\mathrm{d}t$;

(iii) $G(x) - G(a) = \int_a^x f(t)\,\mathrm{d}t$ for some $f \in \mathrm{L}_{\mathbb{R}}[a,b]$.

25 Integrable functions on $\mathbb{R}^k$

Areas are to integrals of functions of one variable as volumes are to integrals of functions of two variables. A first course in multi-variable calculus oriented towards applied mathematics customarily treats 'double' and 'triple' integrals at the same level of mathematical sophistication as does a first course in 1-variable calculus. In developing the 1-variable Lebesgue theory, we provided L^C as a stepping stone to L, which we shall here call $L(\mathbb{R})$. A similar stepping stone is of little assistance in higher dimensions. However we have set up our theory of $L(\mathbb{R})$ in such a way that it extends to higher dimensions with only notational changes.

Remember our strategy for constructing $L(\mathbb{R})$, as summarized in 12.14.

Stage 1. For $f = \chi_I$, where $I = \langle a, b\rangle$ is a bounded interval in $\mathbb{R}$ with endpoints a and b, define $\int f := (b - a)$, motivated by the definition of the area of a rectangle.

Stage 2. For f a step function, that is, a finite linear combination of the functions from Stage 1, define $\int f$ linearly.

Stage 3. For $f \in L^{\text{inc}}$, that is, f the limit a.e. of an increasing sequence $\{\varphi_n\}$ of functions from Stage 2 for which $\lim \int \varphi_n$ exists, define $\int f$ 'MCT-fashion', that is, $\int f = \int \lim \varphi_n := \lim \int \varphi_n$.

Stage 4. For $f \in L_{\mathbb{R}}$, that is, the difference of two functions from Stage 3, define $\int f$ linearly.

Stage 5. For $f \in L$, that is, $\operatorname{Re} f$ and $\operatorname{Im} f$ functions from Stage 4, define $\int f := \int \operatorname{Re} f + i \int \operatorname{Im} f$.

Where have we used the fact that $f\colon \mathbb{R} \to \mathbb{R}$ (or $f\colon \mathbb{R} \to \mathbb{C}$) and that $\int f$ is to be the area under the graph? Answer: only at Stage 1, and at Stage 3, where we insinuate 'a.e.'. To get a 2-dimensional integral, all we have to do is to replace our 1-dimensional bounded intervals in $\mathbb{R}$ by their 2-dimensional analogues—bounded rectangles in $\mathbb{R}^2$—and proceed exactly as before! For example, in place of 1-dimensional step functions we have functions which graphically look like a city of high-rise office blocks, as in 25.1. We can proceed likewise in higher dimensions. We draw attention to a few key points.

25.1 Definitions. For $k \geqslant 2$ an *interval in* $\mathbb{R}^k$ is a set of the form

$$\mathbf{I} = I_1 \times \cdots \times I_k := \{\, (x_1, \ldots, x_k) \mid x_j \in I_j \text{ for all } j \,\},$$

where each $I_j = \langle a_j, b_j\rangle$ is an interval in $\mathbb{R}$, as defined in 3.1. The interval $\mathbf{I}$ is respectively *open, closed, bounded* if each I_j is open, closed, bounded and an

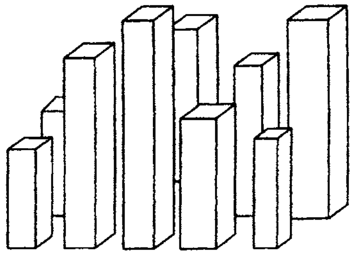


Figure 25.1

interval in $\mathbb{R}^2$ will be called a *rectangle*. For a bounded interval $\mathbf{I}$ we define the *measure of* $\mathbf{I}$ to be

$$m(\mathbf{I}) := (b_1 - a_1) \times \cdots \times (b_k - a_k).$$

A subset E of $\mathbb{R}^k$ is *null* if given $\varepsilon > 0$ there exists a sequence of intervals $\mathbf{I}_1, \mathbf{I}_2, \ldots$ such that

$$E \subseteq \bigcup_i \mathbf{I}_i \quad \text{and} \quad \sum_i m(\mathbf{I}_i) \leqslant \varepsilon.$$

We say a statement holds a.e.$[\mathbb{R}^k]$ if it is true except on a null set in $\mathbb{R}^k$. Informally, a set is null in $\mathbb{R}$ if it has zero length, and null in $\mathbb{R}^2$ if it has zero area. Note that this means that the interval $[0, 1]$ is not null when regarded as a subset of $\mathbb{R}$ but is null when regarded as a subset of the plane, as it has zero area; see Exercise 25.7.

In Stage 1 of our construction for the integral on $\mathbb{R}^k$ we take

$$\int \chi_{\mathbf{I}} := m(\mathbf{I})$$

for any bounded interval $\mathbf{I}$ in $\mathbb{R}^k$. We now define $\mathrm{L}^{\mathrm{step}}(\mathbb{R}^k)$, $\mathrm{L}^{\mathrm{inc}}(\mathbb{R}^k)$, $\mathrm{L}_{\mathbf{R}}(\mathbb{R}^k)$, $\mathrm{L}(\mathbb{R}^k)$ and the integrals on them in the manner indicated in Stages (2)–(5) above. The consistency proofs which validate these definitions work just as before, apart from notational changes.

Measurable functions on $\mathbb{R}^k$ are defined to be those which are a.e.$[\mathbb{R}^k]$ limits of functions in $\mathrm{L}^{\mathrm{step}}(\mathbb{R}^k)$. That is, $f\colon \mathbb{R}^k \to \mathbb{R}$ is measurable if there exists a sequence $\{\varphi_m\}$ of functions in $\mathrm{L}^{\mathrm{step}}(\mathbb{R}^k)$ such that $\varphi_n(\mathbf{x}) \to f(\mathbf{x})$ for $\mathbf{x} \notin E$, where E is null in $\mathbb{R}^k$. A subset S of $\mathbb{R}^k$ is *measurable* if χ_S is a measurable function, and the *volume of* S (or *area of* S in case $k = 2$) is defined to be $\int \chi_S$ if this is finite and ∞ otherwise.

Highlights of the higher dimensional theory are presented below. For more on measurable functions see 25.6.

25.2 Theorem. Let $k \geqslant 2$.

(a) Technical theorems I, II, and III hold for $L_{\mathbb{R}}(\mathbb{R}^k)$.
(b) Properties (L) and (M) hold for $L(\mathbb{R}^k)$ and (P) holds also for $L_{\mathbb{R}}(\mathbb{R}^k)$.
(c) A continuous real-valued (complex-valued) function on a closed bounded interval in $\mathbb{R}^k$ belongs to $L_{\mathbb{R}}(\mathbb{R}^k)$ ($L(\mathbb{R}^k)$).
(d) The MCT holds for $L_{\mathbb{R}}(\mathbb{R}^k)$ and the DCT for $L(\mathbb{R}^k)$.
(e) Measurable functions on $\mathbb{R}^k$ have the properties stated for $k = 1$ in 21.5, a function which is continuous a.e.$[\mathbb{R}^k]$ is measurable, and the Truncation Lemma (21.2) extends in the obvious way.

25.3 Remarks: $L(\mathbb{R}^k)$ vs. $L(\mathbb{R})$. A notable omission from the list in Theorem 25.2 is the Fundamental Theorem of Calculus (the corresponding result in higher dimensions is Stokes's Theorem, which we shall not discuss in this book). Without the FTC we would have been severely restricted in testing for integrability of functions on $\mathbb{R}$ and in evaluating integrals. So how do we recognize which functions belong to $L(\mathbb{R}^k)$ and how do we integrate them? In elementary multi-variable calculus, areas and volumes are computed by integrating with respect to one variable at a time. For example, the area bounded by the ellipse $x^2/a^2 + y^2/b^2 = 1$ $(a, b > 0)$ is computed as

$$\int_{-b}^{b} \left\{ \int_{-a\sqrt{b^2-y^2}/b}^{a\sqrt{b^2-y^2}/b} \mathrm{d}x \right\} \mathrm{d}y = \frac{2a}{b} \int_{-b}^{b} \sqrt{b^2 - y^2}\, \mathrm{d}y = 2ab \int_{-\pi/2}^{\pi/2} \cos^2\theta \, \mathrm{d}\theta = \pi ab.$$

The functions likely to be involved in such calculations are well-behaved continuous functions on closed bounded sets. The simplest such function is of the type $\chi_{\mathbf{I}}$ where $\mathbf{I} = \langle a, b\rangle \times \langle c, d\rangle$ is a bounded rectangle. We have $\chi_{\mathbf{I}}(x, y) = \chi_{\langle a,b\rangle}(x)\chi_{\langle c,d\rangle}(y)$, whence

$$\int \chi_{\mathbf{I}} := (b-a)(d-c) = \int \left\{ \int \chi_{\mathbf{I}}(x, y)\, \mathrm{d}x \right\} \mathrm{d}y.$$

—so far, so good! We are led to hope that all multi-dimensional Lebesgue integrals can be reduced to iterated 1-dimensional integrals. We shall indeed prove in the next chapter that

$$f \in L(\mathbb{R}^2) \implies \int_{\mathbb{R}^2} f = \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}} f(x, y)\, \mathrm{d}x \right\} \mathrm{d}y.$$

This is a major theorem Fubini's Theorem) and not a truism: the integral on the left-hand side is not **defined** as the expression on the right-hand side.

The proof of Fubini's Theorem proceeds in stages from characteristic functions of rectangles, through $L^{step}(\mathbb{R}^2)$ and $L^{inc}(\mathbb{R}^2)$ to $L(\mathbb{R}^2)$. We carry out the first steps now. First we state more explicitly some facts we used above. Informally, the first statement of the following lemma says that a photograph of the Manhattan skyline looks like the graph of a 1-dimensional step function.

25.4 Lemma (Fubini's Theorem for step functions). Let $\varphi \in \mathrm{L}^{\mathrm{step}}(\mathbb{R}^2)$. Then $\varphi^{[y]}(x) := \varphi(x, y)$ defines a function $\varphi^{[y]} \in \mathrm{L}^{\mathrm{step}}(\mathbb{R})$ for each $y \in \mathbb{R}$, and

$$\Phi(y) := \int_{\mathbb{R}} \varphi^{[y]}(x)\,\mathrm{d}x$$

defines a function $\Phi \in \mathrm{L}^{\mathrm{step}}(\mathbb{R})$ and

$$\int \Phi = \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}} \varphi^{[y]}(x)\,\mathrm{d}x \right\} \mathrm{d}y = \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}} \varphi(x, y)\,\mathrm{d}x \right\} \mathrm{d}y.$$

A similar statement holds with the roles of x and y reversed.

Proof. When $\varphi = \chi_{\langle a,b\rangle \times \langle c,d\rangle}$, we have $\varphi^{[y]} = \chi_{\langle a,b\rangle}$ if $y \in \langle c, d\rangle$ and $\varphi^{[y]} = 0$ otherwise. Further, $\Phi = (b-a)\chi_{\langle c,d\rangle}$. The general case follows from the definition of the integral on $\mathrm{L}^{\mathrm{step}}(\mathbb{R})$. □

In going from $\mathrm{L}^{\mathrm{step}}(\mathbb{R}^2)$ to $\mathrm{L}^{\mathrm{inc}}(\mathbb{R}^2)$ we meet a.e. statements. As a consequence we shall need a technical result on null sets which is intuitively entirely reasonable; see 25.2. It says that almost all vertical (or horizontal) slices across a set of zero area have zero length.

In general, the 'almost all' in the conclusion of the Slice Lemma gives a clue that the proof isn't merely a definition-chase. Instead we have to appeal to Technical Theorems I and II, which characterize nullness in terms of step functions.

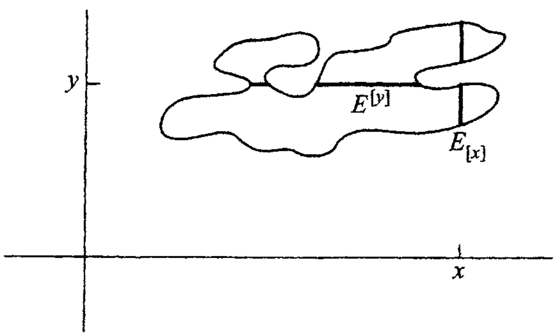


Figure 25.2

25.5 Slice Lemma. Let E be a null set in $\mathbb{R}^2$. Then, for almost all $x \in \mathbb{R}$,

$$E_{[x]} := \{\, y \in \mathbb{R} \mid (x, y) \in E \,\}$$

is a null set in $\mathbb{R}$. Likewise

$$E^{[y]} := \{\, x \in \mathbb{R} \mid (x, y) \in E \,\}$$

is null for almost all y.

Proof. By Technical Theorem II (9.11) for $\mathrm{L}_{\mathbb{R}}(\mathbb{R}^2)$, there exists an increasing sequence $\{\psi_n\}$ of functions in $\mathrm{L}^{\mathrm{step}}(\mathbb{R}^2)$ for which $\{\int \psi_n\}$ is bounded and $\{\psi_n(x, y)\}$ diverges if $(x, y) \in E$. Let

$$\Psi_n(y) := \int \psi_n(x, y)\,\mathrm{d}x.$$

By Lemma 25.4, $\Psi_n \in \mathrm{L}^{\mathrm{step}}(\mathbb{R})$ and $\int \Psi_n = \int \psi_n$, and this converges by assumption. The sequence $\{\Psi_n\}$ is increasing, with its sequence of integrals bounded. By Technical Theorem I for $\mathrm{L}_{\mathbb{R}}(\mathbb{R})$ (the MCT for step functions),

$$F := \{\, y \in \mathbb{R} \mid \{\Psi_n(y)\} \text{ diverges} \,\}$$

is null. Fix $y \in F$. Then $\{\int \psi_n^{[y]}\}$ diverges, where $\psi_n^{[y]}(x) := \psi_n(x,y)$. Apply Technical Theorem I again. We see that $\{\, x \in \mathbb{R} \mid \{\psi_n(x,y)\} \text{ diverges} \,\}$ is null. This tells us that $E^{[y]}$ is null for any y not in the null set F. □

25.6 Remarks on measurability. We have seen that non-measurable functions on $\mathbb{R}$ are elusive creatures, which we can only capture with the aid of the Axiom of Choice (see 22.8). This means that we do not encounter non-measurable functions except in theoretical situations. Then measurability questions do occur. In particular we may need to know that certain procedures on measurable functions in $\mathbb{R}$ create measurable functions on $\mathbb{R}^2$. Some such results are easy: for example, given measurable functions $f, g\colon \mathbb{R} \to \mathbb{R}$, then $h\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $h(x,y) := f(x)g(y)$ is measurable (Exercise 25.7). Others are plausible but tricky. For example, if $f\colon \mathbb{R} \to \mathbb{R}$ is measurable, then $(x,y) \mapsto f(x-y)$ defines a measurable function on $\mathbb{R}^2$. We prove this in 27.8 using the machinery on change of variables developed in Chapter 27.

We conclude the chapter with some important examples of measurable sets in $\mathbb{R}^k$, assuming some knowledge of open and closed sets in $\mathbb{R}^k$. The assertions generalize 22.4, and we leave the proofs as an exercise.

25.7 Lemma.

(a) Every non-empty open subset U of $\mathbb{R}^k$ is the countable union of open intervals $\{\mathbf{I}_n\}$, and hence is measurable. Further, these intervals can be chosen to be pairwise disjoint, and bounded, with each of the associated closed intervals $\{\bar{\mathbf{I}}_n\}$ also contained in U.

(b) Every non-empty closed subset S of $\mathbb{R}^k$ is measurable.

Exercises

The exercises for this chapter focus on links between the higher dimensional theory and the 1-dimensional case, and on new features. Routine exercises on $\mathrm{L}^{\mathrm{step}}(\mathbb{R}^2)$, for example, have not been included, as they would be boring and not very illuminating.

25.1 Show that if E_1 and E_2 are null sets in $\mathbb{R}$ then $E_1 \times E_2$ is a null set in $\mathbb{R}^2$.

25.2 Show that the following subsets of $\mathbb{R}^2$ are null:

(i) $\{(x,y) \in \mathbb{R} \mid x \in \mathbb{Q} \text{ or } y \in \mathbb{Q} \text{ (or both)}\}$,

(ii) $\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$.

25.3 Let S be a subset of $\mathbb{R}^2$ which is the union, not necessarily disjoint, of a finite or countable family of sets each of which is either a straight line (of finite or infinite length) or a circular arc. Prove that S is null.

25.4 Let $\mathbf{I}$ be a rectangle in $\mathbb{R}^2$. Prove that there exists an open rectangle $\mathbf{J} \subseteq \mathbf{I}$ and a closed rectangle $\mathbf{K} \supseteq \mathbf{I}$ such that $\mathbf{I} \setminus \mathbf{J}$ and $\mathbf{K} \setminus \mathbf{I}$ are null sets.

25.5 Prove the (easy) converse of the Slice Lemma 10.11: that E is null in $\mathbb{R}^2$ if $E_{[x]}$ is null for almost all x and $E^{[y]}$ is null for almost all y.

25.6 Formulate and prove an analogue for $\mathbb{R}^k$ of the translation-invariance property (T), as given in 12.3.

25.7 Let $\varphi, \psi \in \mathrm{L}^{\mathrm{step}}(\mathbb{R})$ and define $\theta(x,y) := \varphi(x)\psi(y)$. Prove that $\theta \in \mathrm{L}^{\mathrm{step}}(\mathbb{R}^2)$.
Now let $f, g\colon \mathbb{R} \to \mathbb{R}$ be measurable functions. Define

$$h(x,y) := f(x)g(y).$$

Prove that $h\colon \mathbb{R}^2 \to \mathbb{R}$ is measurable.

25.8 Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the map $T\colon (x,y) \mapsto (r,\theta)$ where $x = r\cos\theta$, $y = r\sin\theta$ ($r \geqslant 0$, $0 \leqslant \theta < 2\pi$, with θ chosen arbitrarily when $r = 0$). Prove that T maps null sets to null sets.
[Prove more generally that any homeomorphism $T\colon Y \to V$ between open subsets Y and V of $\mathbb{R}^2$ maps null sets to null sets. Recall that a homeomorphism is a continuous one-to-one onto map with a continuous inverse.]

25.9 Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be measurable and assume that $x \mapsto f(x,y)$ is integrable for all $y \notin N$, where N is null. Prove that $F(y) := \int f(x,y)\,dx$ defines a measurable function F.

26 Fubini's Theorem and Tonelli's Theorem

This chapter presents two theorems which allow us to relate the integral on $\mathbb{R}$ with the integral on $\mathbb{R}^k$ for $k > 1$ and to evaluate and manipulate higher dimensional integrals. These theorems are attributed to Fubini and Tonelli, whose names sound rather appropriately like those of a conjuring act. Tonelli's magic spell proves integrability assertions, while Fubini's interchanges the order of integration with respect to different variables. Like all conjuring tricks, Fubini's and Tonelli's theorems are not as mysterious as they might seem. They are a direct consequence of the way in which the integrals are set up. In Chapter 25 we introduced $\mathrm{L}(\mathbb{R}^k)$. Here we shall work throughout with $k = 2$, noting that extensions to higher dimensions are easy to formulate if needed.

26.1 Repeated integrals and 2-dimensional integrals. Assume we are given a function $f\colon \mathbb{R}^2 \to \mathbb{C}$. The expressions

$$\int \left\{ \int f(x,y)\,\mathrm{d}x \right\} \mathrm{d}y \quad \text{and} \quad \int \left\{ \int f(x,y)\,\mathrm{d}y \right\} \mathrm{d}x$$

are called *repeated integrals*. An integral $\int\{\int f(x,y)\,\mathrm{d}x\}\,\mathrm{d}y$ is understood to be evaluated in two stages, from the inside out; in the inner integral, $\int f(x,y)\,\mathrm{d}x$, y is held fixed. When we say that the repeated integral exists we mean that this two-stage process is meaningful. Let us spell out carefully what is involved here. We require

(i) $f^{[y]}\colon x \mapsto f(x,y)$ is integrable for all $y \notin N$, where N is some null set, and

(ii) F, defined for $y \notin N$ by $F(y) := \int f(x,y)\,\mathrm{d}x$, belongs to L.

(Remember once again the convention adopted in 12.9 concerning functions defined a.e.) Assuming (i) and (ii) hold,

$$\int \left\{ \int f(x,y)\,\mathrm{d}x \right\} \mathrm{d}y = \int \left\{ \int f^{[y]}(x)\,\mathrm{d}x \right\} \mathrm{d}y = \int F(y)\,\mathrm{d}y.$$

We stress that this process involves only integrals of functions of **one** variable.

Now assume that $f \in \mathrm{L}(\mathbb{R}^2)$. We adopt the notation

$$\int f \quad \text{or} \quad \int f(x,y)\,\mathrm{d}(x,y);$$

here the $\mathrm{d}(x,y)$ notation is intended to indicate an 'element of area', in the same way that in elementary integral calculus $\mathrm{d}x$ is thought of as an 'element of length'. (Remember that $\mathrm{L}(\mathbb{R}^2)$ is built using characteristic functions of bounded rectangles as the 'atoms'.) We have pointed out before, and emphasize again, that for $f \in \mathrm{L}(\mathbb{R}^2)$,

$$\int f(x,y)\,\mathrm{d}(x,y) \quad \text{and} \quad \int\Big\{\int f(x,y)\,\mathrm{d}x\Big\}\,\mathrm{d}y$$

are not equal **by definition**. Indeed we have no a priori assurance that the repeated integral even exists. However we have proved in 25.4 that when $f \in \mathrm{L}^{\mathrm{step}}(\mathbb{R}^2)$ the repeated integral $\int\{\int f(x,y)\,\mathrm{d}x\}\,\mathrm{d}y$ exists and equals the 2-dimensional integral $\int f$. We are now ready to prove this for an arbitrary $f \in \mathrm{L}(\mathbb{R}^2)$.

In the statement of Fubini's Theorem, (a) asserts existence of one of the repeated integrals, (b) relates the repeated and 2-dimensional integrals, and (c) gives a frequently used consequence.

26.2 Fubini's Theorem. Let $f \in \mathrm{L}(\mathbb{R}^2)$. Then

(a) $f^{[y]} \in \mathrm{L}(\mathbb{R})$ for $y \notin N$, where N is a null set in $\mathbb{R}$, and there exists $F \in \mathrm{L}(\mathbb{R})$ with $F(y) := \int f(x,y)\,\mathrm{d}x$ $(y \notin N)$;

(b) $\displaystyle\int f(x,y)\,\mathrm{d}(x,y) = \int F(y)\,\mathrm{d}y = \int\Big\{\int f(x,y)\,\mathrm{d}x\Big\}\,\mathrm{d}y.$

Corresponding assertions hold when the roles of x and y are reversed. Consequently,

(c) $\displaystyle\int\Big\{\int f(x,y)\,\mathrm{d}x\Big\}\,\mathrm{d}y = \int\Big\{\int f(x,y)\,\mathrm{d}y\Big\}\,\mathrm{d}x.$

Proof. We already know that the theorem is true for $f \in \mathrm{L}^{\mathrm{step}}(\mathbb{R}^2)$. We now make the transition to $\mathrm{L}^{\mathrm{inc}}(\mathbb{R}^2)$. Let $\{\varphi_n\}$ be an $\mathrm{L}^{\mathrm{inc}}(\mathbb{R}^2)$-sequence for $f \in \mathrm{L}^{\mathrm{inc}}(\mathbb{R}^2)$, with $\varphi_n \to f$ off a null set E in $\mathbb{R}^2$. Then $\varphi_n(x,y) \to f(x,y)$ for $x \notin E^{[y]} := \{\, x \in \mathbb{R} \mid (x,y) \in E\,\}$, which is null for all y not in some null set F_1 in $\mathbb{R}$ (by the Slice Lemma, 25.5). We define $\Phi_n(y) := \int \varphi_n(x,y)\,\mathrm{d}x$ $(y \in \mathbb{R})$, and recall that $\Phi_n \in \mathrm{L}^{\mathrm{step}}(\mathbb{R})$ (25.4). In the argument below we interchange limit and integral twice. For this we use the MCT, applied once to $\{\Phi_n\}$ and once to $\{\varphi_n^{[y]}\}$, where $\varphi_n^{[y]}(x) := \varphi_n(x,y)$, for y not in a forbidden null set. [We actually need only Technical Theorem I, 14.3, which is the forerunner to the MCT, and is the special case of it for step functions.] Now

$$\begin{aligned}
\int f(x,y)\,\mathrm{d}(x,y) &:= \lim \int \varphi_n(x,y)\,\mathrm{d}(x,y) \\
&= \lim \int\Big\{\int \varphi_n(x,y)\,\mathrm{d}x\Big\}\,\mathrm{d}y \qquad &&\text{(by 25.4)} \\
&= \lim \int \Phi_n(y)\,\mathrm{d}y \\
&= \int \lim \Phi_n(y)\,\mathrm{d}y \qquad &&\text{(by 9.9).}
\end{aligned}$$

It is part of the conclusion of 9.9 that there exists a null set F_2 such that

$\lim \Phi_n(y)$ is finite for $y \notin F_2$. For $y \notin F_1 \cup F_2$ we have

$$\begin{aligned}
\lim \Phi_n(y) &:= \lim \int \varphi_n(x,y)\,\mathrm{d}x \\
&= \int \lim \varphi_n(x,y)\,\mathrm{d}x && \text{(by 9.9)} \\
&= \int f(x,y)\,\mathrm{d}x \quad \text{for } y \notin E^{[y]} && \text{(by 10.11 and 25.5).}
\end{aligned}$$

Further, we have shown that $\int f(x,y)\,\mathrm{d}x$ exists for almost all y, the exceptional values lying in $F_1 \cup F_2 \cup E^{[y]}$, which is null. This completes the proof for $f \in \mathrm{L}^{\mathrm{inc}}(\mathbb{R}^2)$. The extension to $\mathrm{L}(\mathbb{R}^2)$ follows by linearity. □

26.3 Example (non-integrability by the contrapositive of Fubini's Theorem—repeated integrals unequal). Let $f(x,y) = \dfrac{x-y}{(x+y)^3}$ on $[0,1] \times [0,1]$. Then

$$\begin{aligned}
\int_0^1 \left\{ \int_0^1 f(x,y)\,\mathrm{d}x \right\} \mathrm{d}y &= \int_0^1 \left\{ \int_0^1 \left(\frac{1}{(x+y)^2} - \frac{2y}{(x+y)^3} \right) \mathrm{d}x \right\} \mathrm{d}y \\
&= \int_0^1 \left[-\frac{1}{(x+y)} + \frac{y}{(x+y)^2} \right]_{x=0}^{1} \mathrm{d}y \\
&= \int_0^1 \left(-\frac{1}{1+y} + \frac{y}{(1+y)^2} \right) \mathrm{d}y \\
&= \int_0^1 -\frac{1}{(1+y)^2}\,\mathrm{d}y \\
&= \left[\frac{1}{(1+y)} \right]_0^1 = -\frac{1}{2}.
\end{aligned}$$

By symmetry,

$$\int_0^1 \left\{ \int_0^1 f(x,y)\,\mathrm{d}y \right\} \mathrm{d}x = \frac{1}{2}.$$

Therefore the repeated integrals

$$\int_0^1 \left\{ \int_0^1 f(x,y)\,\mathrm{d}x \right\} \mathrm{d}y \quad \text{and} \quad \int_0^1 \left\{ \int_0^1 f(x,y)\,\mathrm{d}y \right\} \mathrm{d}x$$

are **not** equal. This may come as a surprise. There is nothing amiss, but we must conclude from Fubini's Theorem that f cannot belong to $\mathrm{L}(\mathbb{R}^2)$. This is a situation analogous to that which we encountered in Chapter 13 with improper integrals. The Lebesgue theory doesn't allow within its scope quite all the functions we might wish—the price we pay for its power.

The over-scrupulous might complain that in the calculation of the repeated integral above we have been more cavalier than previously in our use of the (extended) FTC, and not worried about 'nonsense on null sets'. You should have acquired enough experience by now safely to do likewise.

26.4 Exercise example. By proving the repeated integrals are unequal prove that f is not integrable when

$$f(x,y) = \frac{x^2 - y^2}{(x^2+y^2)^2} \qquad \text{on } [0,1] \times [0,1].$$

[Hint for performing the integration: for fixed $y > 0$, $x^2(x^2+y^2)^{-2}$ can be integrated with respect to x by using integration by parts.]

26.5 Example (non-integrability by the contrapositive of Fubini's Theorem applied to $|f|$). Let $f(x,y) = \dfrac{xy}{(x^2+y^2)^2}$ on $[-1,1] \times [-1,1]$. It is routine to check that

$$\int_{-1}^{1} \left\{ \int_{-1}^{1} f(x,y)\,\mathrm{d}x \right\} \mathrm{d}y = \int_{-1}^{1} \left\{ \int_{-1}^{1} f(x,y)\,\mathrm{d}y \right\} \mathrm{d}x = 0.$$

However,

$$\int_{-1}^{1} \left\{ \int_{-1}^{1} |f(x,y)|\,\mathrm{d}x \right\} \mathrm{d}y = 2\int_{0}^{1} (y^{-1} - y(1+y^2)^{-1})\,\mathrm{d}y.$$

This cannot be finite, as $y(1+y^2)^{-1}$ is continuous and so integrable on $[0,1]$, while $\int_0^1 y^{-1}\,\mathrm{d}y$ does not exist. Arguing by contradiction from Fubini's Theorem applied to $|f|$, we deduce that $|f|$ cannot be integrable. Thus f cannot be integrable. Notwithstanding, the repeated integrals of f both exist, and they are equal.

26.6 Exercise example. By showing that one of the repeated integrals is not finite prove that f is not integrable when $f(x,y) = \mathrm{e}^{-xy}$ on $[0,1] \times [1,\infty)$.

You should have noticed that all the preceding examples concerned non-integrable functions. We cannot get full benefit from Fubini's Theorem without having a viable method for proving that functions are integrable. Here the second member of the duo comes on stage.

26.7 Tonelli's Theorem. Suppose $f \colon \mathbb{R}^2 \to \mathbb{C}$ is measurable and one of

$$\int \left\{ \int |f(x,y)|\,\mathrm{d}x \right\} \mathrm{d}y \quad \text{and} \quad \int \left\{ \int |f(x,y)|\,\mathrm{d}y \right\} \mathrm{d}x$$

exists. Then $f \in \mathrm{L}(\mathbb{R}^2)$ and Fubini's Theorem applies.

[**Note**. Remember that we give the value ∞ to the integral of a non-negative measurable function which is not integrable; see 21.11. In each of the repeated integrals the inner integral is either infinite or passes a non-negative measurable function on to the outer one (Exercise 25.9). Thus the repeated integrals of $|f|$ exist if and only if they are finite. If one is finite then both are, by Fubini's Theorem applied to $|f|$.]

Proof. By the Truncation Lemma for functions on $\mathbb{R}^2$ (see 21.2 and 25.2),

$$f_n := f^{\square_n} = (|f| \wedge n)\chi_{[-n,n]\times[-n,n]} \in \mathrm{L}_{\mathbb{R}}(\mathbb{R}^2)$$

for each n. We now apply the MCT to $\{f_n\}$. Certainly (M1) (integrability) and (M2) (monotonicity) are satisfied, and $f_n \to |f|$. Assume without loss of generality that $\int \{\int |f(x,y)| \,\mathrm{d}x\} \,\mathrm{d}y$ is finite. Then

$$\begin{aligned}\int f_n(x,y)\,\mathrm{d}(x,y) &= \int \left\{\int f_n(x,y)\,\mathrm{d}x\right\}\mathrm{d}y \qquad \text{(by Fubini's Theorem)}\\ &\leqslant \int \left\{\int |f(x,y)|\,\mathrm{d}x\right\}\mathrm{d}y\end{aligned}$$

whence (M3) (bounded integrals) holds. By the MCT, $|f|$ is integrable. Since f is measurable, f itself is integrable. □

Here is a 2-dimensional analogue of Theorem 11.2(b).

26.8 Corollary. Let S be a closed bounded set in $\mathbb{R}^2$ and let $f\colon S \to \mathbb{C}$ be continuous. Then $f\chi_S$ is integrable.

Proof. The Boundedness Theorem, 2.23, extends to continuous function on closed bounded subsets of $\mathbb{R}^2$ (see, for example, [1]), so there exists a finite constant M such that $|f(x,y)| \leqslant M$ for all $(x,y) \in S$. Certainly $f\chi_S$ is measurable, using 25.7. Also there exists a bounded rectangle $\mathbf{I}$ such that $S \subseteq \mathbf{I}$. Hence

$$\int \left\{\int |f(x,y)\chi_S(x,y)|\,\mathrm{d}x\right\}\mathrm{d}y \leqslant \int \left\{\int M\chi_{\mathbf{I}}(x,y)\,\mathrm{d}x\right\}\mathrm{d}y = M\,m(\mathbf{I}) < \infty.$$

Therefore $f\chi_S$ is integrable by Tonelli's Theorem. □

26.9 Example. We prove that f is integrable when

$$f(x,y) = \begin{cases} x^2\mathrm{e}^{-xy} & \text{if } 0 \leqslant x \leqslant y,\\ 0 & \text{otherwise.}\end{cases}$$

Since $f \geqslant 0$ and f is measurable (true but not obvious—use 25.7 or 27.8), f is integrable by Tonelli's Theorem if we can show that either of its repeated integrals is finite. Since y appears only in the exponential we prefer to integrate with respect to y first:

$$\begin{aligned}\int \left\{\int f(x,y)\,\mathrm{d}y\right\}\mathrm{d}x &= \int_0^\infty \left\{\int_x^\infty x^2\mathrm{e}^{-xy}\,\mathrm{d}y\right\}\mathrm{d}x\\ &= \int_0^\infty \left[-x\mathrm{e}^{-xy}\right]_{y=x}^{\infty}\mathrm{d}x\\ &= \int_0^\infty x\mathrm{e}^{-x^2}\,\mathrm{d}x.\end{aligned}$$

The last integral is certainly finite, so f is integrable.

26.10 Exercise example. Use Tonelli's Theorem to prove that $e^{-xy}\sin y$ is integrable on $[1,\infty] \times [0,\infty)$.

By validating the interchange of the order of integration in a repeated integral we can evaluate a variety of otherwise awkward integrals. The strategy in given by the following amalgam of 26.2 and 26.7.

26.11 Abbreviated Fubini+Tonelli Theorem (FTT). Let $f\colon \mathbb{R}^2 \to \mathbb{C}$ be measurable and assume $\int \{\int |f(x,y)|\,dy\}\,dx$ (or $\int \{\int |f(x,y)|\,dx\}\,dy$) is finite. Then $f \in L(\mathbb{R}^2)$, the repeated integrals of f are well defined, and

$$\int\left\{\int f(x,y)\,dx\right\}dy = \int\left\{\int f(x,y)\,dy\right\}dx.$$

26.12 Example (Tonelli and Fubini in tandem). The function $xe^{-x^2(1+y^2)}$ on $[0,\infty) \times [0,\infty)$ is continuous, hence measurable, and non-negative. Consider

$$\begin{aligned}\int_0^\infty\left\{\int_0^\infty xe^{-x^2(1+y^2)}\,dx\right\}dy &= \int_0^\infty\left[-\frac{1}{2(1+y^2)}e^{-x^2(1+y^2)}\right]_{x=0}^\infty dy\\ &= \int_0^\infty \frac{1}{2(1+y^2)}\,dy\\ &= \left[\frac{1}{2}\tan^{-1}y\right]_0^\infty = \frac{\pi}{4}.\end{aligned}$$

We deduce from Tonelli's Theorem that $xe^{-x^2(1+y^2)}$ is integrable on $[0,\infty) \times [0,\infty)$. By Fubini's Theorem,

$$\int_0^\infty\left\{\int_0^\infty xe^{-x^2(1+y^2)}\,dy\right\}dx = \frac{\pi}{4}.$$

Now make the substitution $u = xy$ in the inner integral (with x fixed). The displayed integral becomes

$$\int_0^\infty\left\{\int_0^\infty e^{-x^2}e^{-u^2}\,du\right\}dx = \left(\int_0^\infty e^{-v^2}\,dv\right)^2.$$

Therefore our conjuring tricks have produced

$$\int_0^\infty e^{-v^2}\,dv = \frac{\sqrt{\pi}}{2}.$$

26.13 Exercise example.

(a) Prove that e^{-xy} is integrable over $[0,\infty) \times [1,a]$ $(1 < a < \infty)$.

(b) Use Fubini's Theorem to deduce that

$$\int_0^\infty \frac{e^{-x} - e^{-ax}}{x}\,dx = \log a \qquad (a > 1).$$

(c) By considering repeated integrals, show that $e^{-xy} - ae^{-axy}$ is not integrable over $[0,1] \times [1,\infty)$ for any $a > 1$.

26.14 Example: a fresh look at integration by parts. Assume f, g are real-valued integrable functions on a compact interval $[a,b]$ in $\mathbb{R}$, with indefinite integrals

$$F(x) = \int_a^x f(t)\,\mathrm{d}t \quad \text{and} \quad G(x) = \int_a^x g(t)\,\mathrm{d}t \qquad (a \leqslant x \leqslant b).$$

We apply FTT to q, where $q(s,t) := f(s)g(t)k(s,t)$, k being the characteristic function of $\{(s,t) \mid a \leqslant s \leqslant t \leqslant b\}$. By 25.7 and 25.6 q is measurable. Also

$$\begin{aligned}\int\left\{\int |f(s)g(t)k(s,t)|\,\mathrm{d}s\right\}\mathrm{d}t &\leqslant \int_a^b\left\{\int_a^b |f(s)g(t)|\,\mathrm{d}t\right\}\mathrm{d}s \\ &= \int_a^b |f(s)|\,\mathrm{d}s \int_a^b |g(t)|\,\mathrm{d}t,\end{aligned}$$

and this is finite by hypothesis. By Tonelli's Theorem, q is integrable. Now

$$\begin{aligned}\int_a^b F(t)g(t)\,\mathrm{d}t &= \int_a^b\left\{\int_a^t f(s)g(t)\,\mathrm{d}s\right\}\mathrm{d}t && \text{(by definition of } F\text{)} \\ &= \int_a^b\left\{\int_a^b f(s)g(t)k(s,t)\,\mathrm{d}s\right\}\mathrm{d}t && \text{(by definition of } k\text{)} \\ &= \int_a^b\left\{\int_a^b f(s)g(t)k(s,t)\,\mathrm{d}t\right\}\mathrm{d}s && \text{(by Fubini's Theorem)} \\ &= \int_a^b\left\{\int_s^b f(s)g(t)\,\mathrm{d}t\right\}\mathrm{d}s && \text{(by definition of } k\text{)} \\ &= \int_a^b\left\{\int_a^b f(s)g(t)\,\mathrm{d}t\right\}\mathrm{d}s \\ &\qquad - \int_a^b\left\{\int_a^s f(s)g(t)\,\mathrm{d}t\right\}\mathrm{d}s \\ &= F(b)G(b) - \int_a^b f(s)G(s)\,\mathrm{d}s && \text{(by definition of } F \text{ and } G\text{).}\end{aligned}$$

In the case that f and g are continuous we recapture the usual integration by parts formula given in 6.11 by combining the above formula with the Indefinite Integral Theorem, 5.4.

Note how we used k to build the variable limits of integration into the integrand.

The final results in this chapter might more properly belong in Chapter 33. They are such good illustrations of appeal to the Fubini and Tonelli Theorems that we present them now.

26.15 The Fourier convolution (calling on all Fubini's powers). Let $f, g \in \mathrm{L}$. We define the *convolution* $f * g$ by

$$(f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)\,\mathrm{d}t,$$

or try to—we don't yet know that this integral makes sense. Fubini's Theorem will allow us to prove that the integral does exist for almost all x, and that it defines a function $f * g \in \mathrm{L}$. Such functions occur naturally in a number of applications; see 33.11 and Chapter 34.

We define q by $q(x,t) := f(t)g(x-t)$, which is measurable so long as $(x,t) \mapsto g(x-t)$ is. This a plausible but non-trivial result which we prove in 27.8. We now apply Tonelli's Theorem to q. We have

$$\begin{aligned}
\int \left\{ \int |f(t)g(x-t)|\,\mathrm{d}x \right\} \mathrm{d}t &= \int \left\{ \int |f(t)g(x-t)|\,\mathrm{d}x \right\} \mathrm{d}t \\
&= \int |f(t)| \left\{ \int |g(x-t)|\,\mathrm{d}x \right\} \mathrm{d}t \\
&= \int |f(t)| \left\{ \int |g(u)|\,\mathrm{d}u \right\} \mathrm{d}t && \text{(by (T))} \\
&= \int |f| \int |g| && \text{(by (L))}.
\end{aligned}$$

Hence, by Tonelli's Theorem, $f(t)g(x-t)$ is integrable as a function of (x,t). Fubini's Theorem then tells us that the convolution integral $\int f(t)g(x-t)\,\mathrm{d}t$ exists for almost all x and is integrable. Hence $f * g$ exists. (Here we are again taking advantage of our convention about functions only needing to be defined a.e.; recall the remarks in 12.9.)

It is easy to see how 'a.e.' can arise. Consider $(f * f)(0) = \int f(t)f(-t)\,\mathrm{d}t$. If $f \in \mathrm{L}_{\mathbb{R}}$ and f is even, but $f^2 \notin \mathrm{L}_{\mathbb{R}}$, then this integral fails to exist. For an example consider our old friend $f(x) := |x|^{-1/2}\,\chi_{[-1,1]}(x)$. In general 'singularities' of f and g may not be sufficiently serious enough to prevent f and g being integrable individually, yet can coalesce in the integrand of the convolution to cause this not to be integrable; see Exercise 26.8.

26.16 Convolution Theorem for the Fourier transform. For $f \in \mathrm{L}$ the Fourier transform, $\hat{f}$, is defined by

$$\hat{f}(y) = \int_{-\infty}^{\infty} f(x)\mathrm{e}^{-\mathrm{i}xy}\,\mathrm{d}x.$$

Because f is integrable and the exponential factor is measurable and bounded (it has modulus 1), $\hat{f}(y)$ is defined for all y. The Convolution Theorem states that, for all $f, g \in \mathrm{L}$, $\widehat{f * g} = \hat{f}\hat{g}$ a.e. To see how this comes about, observe that, for each fixed $y \in \mathbb{R}$,

$$
\begin{aligned}
(\widehat{f*g})(y) &= \int \left\{ \int f(t)g(x-t)\mathrm{e}^{-\mathrm{i}yx}\,\mathrm{d}t \right\} \mathrm{d}x \\
&\stackrel{*}{=} \int \left\{ \int f(t)g(x-t)\mathrm{e}^{-\mathrm{i}yx}\,\mathrm{d}x \right\} \mathrm{d}t \\
&= \int f(t)\mathrm{e}^{-\mathrm{i}yt} \left\{ \int g(x-t)\mathrm{e}^{-\mathrm{i}y(x-t)}\,\mathrm{d}x \right\} \mathrm{d}t \\
&= \int f(t)\mathrm{e}^{-\mathrm{i}yt} \left\{ \int g(u)\mathrm{e}^{-\mathrm{i}yu}\,\mathrm{d}u \right\} \mathrm{d}x \\
&= \hat{f}(y)\hat{g}(y).
\end{aligned}
$$

provided the repeated integrals make sense and the interchange of the order of integration $(*)$ is valid.

Now we give the justification. With y fixed, we let

$$q(x,t) := f(t)g(x-t)\mathrm{e}^{-\mathrm{i}yx},$$

noting that this becomes $f(t)g(x-t)$ when $y = 0$. We first apply Tonelli's Theorem to q, which is measurable by 27.8 and 25.6. We have

$$
\begin{aligned}
\int \left\{ \int \left| f(t)g(x-t)\mathrm{e}^{-\mathrm{i}yx} \right| \mathrm{d}x \right\} \mathrm{d}t &= \int \left\{ \int |f(t)g(x-t)|\,\mathrm{d}x \right\} \mathrm{d}t \\
&= \int |f(t)| \left\{ \int |g(x-t)|\,\mathrm{d}x \right\} \mathrm{d}t \\
&= \int |f| \int |g| \qquad \text{(by (T))}.
\end{aligned}
$$

Hence, by Tonelli's Theorem, $f(t)g(x-t)\mathrm{e}^{-\mathrm{i}yx}$ is integrable as a function of (x,t), for each y. Applying Fubini's Theorem in the case $y = 0$ we obtain that the convolution integral $\int f(t)g(x-t)\,\mathrm{d}t$ exists for almost all x and is integrable (as proved already in 26.15). Finally Fubini's Theorem justifies $(*)$, for any y.

26.17 Calling on Fubini and Tonelli: summary. Fubini's Theorem reconciles repeated 1-dimensional integrals with 2-dimensional integrals. Tonelli's Theorem, which only involves 1-dimensional integrals, is the recommended way to show a given function is in $\mathrm{L}(\mathbb{R}^2)$. The measurability condition in Tonelli's Theorem is usually satisfied in concrete examples because the integrand is continuous a.e. In abstract applications it may be harder to check; see 25.6.

We may summarize the uses of Fubini's Theorem and Tonelli's Theorem as follows. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ or $\mathbb{C}$ be measurable.

- Testing for integrability: use Tonelli's Theorem.
- Testing for non-integrability: try one of the following.
  (a) If $\int \{\int f(x,y)\,\mathrm{d}x\}\,\mathrm{d}y \neq \int \{\int f(x,y)\,\mathrm{d}y\}\,\mathrm{d}x$ then Fubini's Theorem implies $f \notin \mathrm{L}(\mathbb{R}^2)$. The repeated integrals may be equal even for non-integrable functions so this test may give no result.
  (b) Try to show that $\int \{\int |f(x,y)|\,\mathrm{d}x\}\,\mathrm{d}y$ or $\int \{\int |f(x,y)|\,\mathrm{d}y\}\,\mathrm{d}x$ is not finite. Note that if f were integrable, $|f|$ would be too, and argue by contradiction, applying Fubini to $|f|$.

- A necessary and sufficient condition for integrability: a measurable function f on $\mathbb{R}^2$ belongs to $\mathrm{L}(\mathbb{R}^2)$ if and only if $\int \{\int |f(x,y)|\,\mathrm{d}x\}\,\mathrm{d}y$ and $\int \{\int |f(x,y)|\,\mathrm{d}y\}\,\mathrm{d}x$ are both finite. This is just an amalgam of Tonelli's Theorem and the contrapositive of Fubini's Theorem, applied to $|f|$. It exposes the way in which Tonelli and Fubini complement one another.
- To show that $\int \{\int f(x,y)\,\mathrm{d}x\}\,\mathrm{d}y = \int \{\int f(x,y)\,\mathrm{d}y\}\,\mathrm{d}x$, use FTT: Tonelli's Theorem to show $f \in \mathrm{L}(\mathbb{R}^2)$ and then Fubini's Theorem to equate the repeated integrals. As a working rule, assume that **whenever** two integrals are juxtaposed worthwhile conclusions will only come when their order is reversed. Beware concealing interchanges of the order of integration amongst other manipulations such as changes of variable.
- A by-product of Fubini's Theorem. Suppose you are asked to show that a certain function $\int f(x,y)\,\mathrm{d}x$ of y exists a.e. The 'a.e.' is almost certainly a signal that an argument is needed like that given in 26.15 for the Fourier convolution.
- When working with repeated integrals always build variable limits of integration into the integrand using a characteristic function; the limits will then take care of themselves when the order of integration is changed. For an illustration, see Exercise 26.9.

Another valuable technique for showing the existence of multi-dimensional integrals, and for evaluating them, is the use of a change of variables, to polar coordinates, for example. This topic is the subject of the next chapter.

Exercises

26.1 In which of the following cases is f integrable on the set specified? Justify each answer by use of Tonelli's Theorem or an argument by contradiction using Fubini's Theorem.

$$\text{(i) } f(x,y) = \begin{cases} (x-1)^{-3} & \text{if } 0 < y < |x-1|, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{(ii) } f(x,y) = \begin{cases} y^{-2} & \text{if } 0 < x < y < 1, \\ -x^{-2} & \text{if } 0 < y < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{(iii) } f(x,y) = y\,|1-xy|^{-3/2} \quad \text{on } [1,\infty) \times [1,\infty),$$

$$\text{(iv) } f(x,y) = \mathrm{e}^{-x^2y^2} \quad \text{on } [1,\infty) \times [1,\infty).$$

[Hint for (iv): you don't have to evaluate your repeated integral—an upper bound will do.]

26.2 By integrating x^y on $[0,1] \times [a,b]$ prove that

$$\int_0^1 \frac{x^b - x^a}{\log x}\,\mathrm{d}x = \log\left(\frac{1+b}{1+a}\right) \quad (b > a > 0).$$

26.3 By integrating suitable functions on $[0,\infty) \times [a,b]$ evaluate the following integrals for $b > a > 0$:

$$\text{(i)} \int_0^\infty \frac{\mathrm{e}^{-ax} - \mathrm{e}^{-bx}}{x}\,\mathrm{d}x, \qquad \text{(ii)} \int_0^\infty \frac{\tan^{-1}(bx) - \tan^{-1}(ax)}{x}\,\mathrm{d}x.$$

26.4 Let J_0 be the Bessel function defined in Exercise 19.1. Prove that

$$\int_0^\infty J_0(x)\mathrm{e}^{-ax}\,\mathrm{d}x = \frac{1}{\sqrt{1+a^2}} \quad (a > 0).$$

[Hint: remember 6.10 and 15.9(2).]

26.5 (a) Prove that $\mathrm{e}^{-xy^2}\sin x$ is integrable on $[0,n] \times [0,\infty)$ for any $n \in \mathbb{N}$.
(b) Assuming that

$$\int_0^\infty \frac{1}{1+x^4}\,\mathrm{d}x = \frac{\pi}{2\sqrt{2}}$$

(a result customarily evaluated using the methods of complex analysis (see **[13]**, Exercise 8.3)), prove that

$$\lim_{n\to\infty} \int_0^n \frac{\sin x}{\sqrt{x}}\,\mathrm{d}x = \sqrt{\frac{\pi}{2}}.$$

[Note that $\sin x/\sqrt{x}$ is not integrable on $[0,\infty)$; cf. 13.5.]

26.6 Let $k : \mathbb{R}^2 \to \mathbb{C}$ be a measurable function for which $k^{[y]} \in \mathrm{L}$ for all $y \in \mathbb{R}$ and there exists $M \in \mathbb{R}$ such that $\int \left|k^{[y]}\right| \leqslant M$ for all y. Prove that for any $f \in \mathrm{L}(\mathbb{R})$ almost all of the integrals $\int k(x,y)f(y)\,\mathrm{d}y$ exist, and that

$$(Kf)(x) := \int k(x,y)f(y)\,\mathrm{d}y \quad (x \in \mathbb{R},\ f \in \mathrm{L}(\mathbb{R}))$$

defines a linear map $K\colon \mathrm{L} \to \mathrm{L}$.

26.7 Let $f, g, h \in \mathrm{L}$.
(a) Prove that $f * g = g * f$ a.e.
(b) Prove that $(f * g) * h = f * (g * h)$ a.e.

26.8 Show how to construct a function $f \in \mathrm{L}_{\mathbb{R}}$ such that the integral defining $f * f$ fails to exist at infinitely many points. [Hint: adapt the example given in the remarks following 26.16.]

26.9 Let f and g be measurable and such that the Laplace transforms $\overline{f}(p) := \int_0^\infty f(t)\mathrm{e}^{-pt}\,\mathrm{d}t$ and $\overline{g}(p) := \int_0^\infty g(t)\mathrm{e}^{-pt}\,\mathrm{d}t$ exist as Lebesgue integrals for $\operatorname{Re} p \geqslant c$ (c a positive constant). Prove that

$$h(t) := \int_0^t f(u)g(t-u)\,\mathrm{d}u$$

exists for almost all $t \geqslant 0$ and that $\overline{h}(p) = \overline{f}(p)\overline{g}(p)$ for $\operatorname{Re} p \geqslant c$. [This is the convolution theorem for the Laplace transform.]

27 Transformations of $\mathbb{R}^k$

The preceding chapter emphasized the connections between integrals on $\mathbb{R}$ and on $\mathbb{R}^2$. There are aspects of the integral on $\mathbb{R}^2$ (and likewise on $\mathbb{R}^k$ for $k > 2$) which are interesting and important in their own right, notably the relationship between Lebesgue measure and the rich geometry of $\mathbb{R}^k$. We have defined the measure of a bounded measurable subset S of $\mathbb{R}^k$ to be $\int \chi_S$. This agrees, by definition, with the expected area when S is a bounded interval. Is the same true for other familiar figures, such as parallelograms? What happens to measure under transformations of the plane? Substitution is a valuable technique for evaluating integrals on $\mathbb{R}$. In the same way, it is useful to be able to make changes of variables in higher dimensional integrals. In full generality these topics involve multi-dimensional calculus beyond the scope of this book. Our aim here is to validate in the Lebesgue framework the use of the transformations which are most useful in practice. Beginners may wish to omit the proof of the main theorem, 27.4.

27.1 Change of variables: preliminary remarks. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a continuous map. The theorem we should like to prove would take the following form (cf. 6.6):

$$\int_{T(U)} g(u,v)\,\mathrm{d}(u,v) = \int_U g(T(x,y))\mu(x,y)\,\mathrm{d}(x,y),$$

where

(i) U is a suitable measurable set, and $T(U) := \{\, T(x,y) \mid (x,y) \in U \,\}$ its image under T,

(ii) g is a suitable integrable function, and

(iii) μ is a non-negative 'density function', which should satisfy, for every bounded interval $\mathbf{I}$,

$$(*)\quad m(T(\mathbf{I})) := \int \chi_{T(\mathbf{I})}(u,v)\,\mathrm{d}(u,v) = \int_{T(\mathbf{I})} 1.\,\mathrm{d}(u,v) = \int_{\mathbf{I}} \mu(x,y)\,\mathrm{d}(x,y).$$

Users of multi-variable calculus will realize that our candidate for μ will be $|\mathbf{J}_T|$, the modulus of the Jacobian determinant of T, namely

$$\mathbf{J}_T = \frac{\partial(u,v)}{\partial(x,y)} := \begin{vmatrix} u_x & v_x \\ u_y & v_y \end{vmatrix},$$

where $T(x,y) = (u(x,y), v(x,y))$.

Notice that all the integrals above are two-dimensional integrals, **not** repeated integrals: $(*)$ tells us how areas of bounded rectangles transform under T.

The change of variables will be easiest to validate in case we can establish $(*)$ by elementary means and U is built from bounded intervals in a simple way. We shall normally take U to be open or closed, and use Lemma 25.7. We also wish, for example, to be able to switch from polar coordinates to cartesian coordinates or vice versa, so that we shall want our transformation T to be reversible. So we want T to be one-to-one, at least off a null set. We now look at some simple transformations T meeting these requirements.

27.2 Linear maps. First consider a change of origin. Let T map (x, y) to $(x + A, y + B)$, where $(A, B) \in \mathbb{R}^2$. Then the formula for change of variables is simply the translation-invariance property (T) for $\mathrm{L}(\mathbb{R}^2)$, and μ is the constant function 1.

Now consider the map $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T(x, y) := (ax + by, cx + dy),$$

where we assume that $\Delta_T := |ad - bc| \neq 0$, so that T has an inverse map

$$T^{-1}(x, y) = \frac{1}{ad - bc}(dx - by, -cx + ay).$$

Under T, a bounded rectangle $\mathbf{I}$ maps to a parallelogram $\mathbf{J} = T(\mathbf{I})$. Working out the coordinates of the vertices of $\mathbf{J}$ in terms of those of $\mathbf{I}$ we see easily that the area of $\mathbf{J}$ is the area of $\mathbf{I}$ multiplied by Δ_T. Remember that we don't yet know that an area of a set in the plane, calculated geometrically, is necessarily the same as the integral of its characteristic function. However we are encouraged to think that we should take μ to be the constant function Δ_T. A formal confirmation, based on Fubini's Theorem, that $(*)$ then holds is sought in Exercise 27.1.

27.3 Polars $\longleftrightarrow$ cartesians. Write

$$T(r, \theta) = (x, y) := (r\cos\theta, r\sin\theta) \quad (r \geqslant 0,\ 0 \leqslant \theta \leqslant 2\pi).$$

Then T would be a one-to-one map of $[0, \infty) \times [0, 2\pi]$ onto $\mathbb{R}^2$ if it were not for the duplication of values on the positive real axis. Since this set, L say, is null in $\mathbb{R}^2$, it causes us no problem. Off L, T is invertible, with

$$T^{-1}(x, y) = \left(\sqrt{x^2 + y^2}, \tan^{-1}\left(\frac{y}{x}\right)\right).$$

We have continuous one-to-one transformations

$$\{\,(r, \theta) \mid 0 < r,\ 0 < \theta < 2\pi\,\} \underset{T^{-1}}{\overset{T}{\rightleftarrows}} \{\,(x, y) \mid x < 0 \text{ or } y \neq 0\,\}.$$

Consider the image under T of a rectangle $\mathbf{I}$, in (r, θ) coordinates, defined by

$$r_1 \leqslant r \leqslant r_2, \quad \theta_1 \leqslant \theta \leqslant \theta_2.$$

Under T, $\mathbf{I}$ maps to a region $\mathbf{J}$ in the (x, y)-plane of the form indicated in Fig. 27.1. Thinking of this as a portion of an annulus, and assuming that the area of a circle radius R is πR^2, we see that the area of $T(\mathbf{I}) = \mathbf{J}$ is

$$\frac{(\theta_2 - \theta_1)}{2\pi} . \pi (r_2^2 - r_1^2).$$

The area of $\mathbf{I}$ is $(\theta_2 - \theta_1)(r_2 - r_1)$. In the limit as $r_2 \to r_1$, the quotient of these areas tends to r_1, suggesting that the appropriate density function μ should be

$$\mu(r, \theta) = r.$$

Note that this is just the Jacobian determinant

$$\mathbf{J}_T = \begin{vmatrix} x_r & y_r \\ x_\theta & y_\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix}.$$

As in 27.2, we confirm that μ satisfies $(*)$ in 27.1 by calculating $\int \chi_{\mathbf{J}}$ and $\int_{\mathbf{I}} \mu$ by Fubini's Theorem. By 26.8, both $\chi_{\mathbf{J}}(x, y)$ and $\mu(r, \theta)\chi_{\mathbf{I}}(r, \theta)$ belong to $\mathrm{L}(\mathbb{R}^2)$. Fubini's Theorem gives immediately

$$\int_{\mathbf{I}} r \,\mathrm{d}(r, \theta) = \int_{\theta_1}^{\theta_2} \left\{ \int_{r_1}^{r_2} r \,\mathrm{d}r \right\} \mathrm{d}\theta = (\theta_2 - \theta_1)\frac{(r_2^2 - r_1^2)}{2}.$$

To simplify the calculation of $\int \chi_{\mathbf{J}}$ we take $r_1 = 0$, $r_2 = \rho$, $\theta_1 = 0$ and $\theta_2 = \alpha$; Properties (L) and (T) then give the result required in general. By Fubini's Theorem, and some elementary trigonometry to find the limits in the repeated integrals,

$$\begin{aligned}
\int \chi_{\mathbf{J}} &= \int_0^{\rho\sin\alpha} \left\{ \int_{y\cot\alpha}^{\sqrt{\rho^2 - y^2}} \mathrm{d}x \right\} \mathrm{d}y \\
&= \int_0^{\rho\sin\alpha} \left(\sqrt{\rho^2 - y^2} - y\cot\alpha \right) \mathrm{d}y \\
&= -\frac{\rho^2 \sin\alpha\cos\alpha}{2} + \int_0^\alpha \rho^2 \cos^2\theta \,\mathrm{d}\theta \qquad \text{(putting } y = \rho\sin\theta\text{)} \\
&= \frac{\rho^2 \alpha}{2}.
\end{aligned}$$

Note that the above change of variable is in a 1-dimensional integral, and is legitimate by 6.6. This completes the verification of $(*)$ for T.

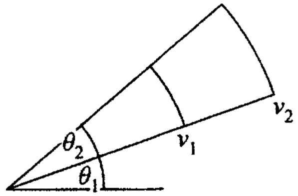


Figure 27.1

Similarly we can check that off the positive real axis $(*)$ holds for the inverse transformation, T^{-1}, from cartesians to polars, with $\mu(x,y) = 1/\sqrt{x^2+y^2}$.

We are now ready for a simple change of variables theorem.

27.4 Theorem. Let U be an open subset of $\mathbb{R}^2$, let $T\colon U \to V := T(U)$ be a continuous one-to-one map and assume that there is a non-negative measurable function μ such that

$$m(T(\mathbf{I})) = \int_{\mathbf{I}} \mu \qquad (\mathbf{I} \text{ any bounded rectangle in } U).$$

Then $g \circ T \in \mathrm{L}(U)$ for any $g \in \mathrm{L}(V)$ and

$$\int_{T(U)} g(u,v)\,\mathrm{d}(u,v) = \int_U g(T(x,y))\mu(x,y)\,\mathrm{d}(x,y).$$

Sketch proof. We shall use the usual step-by-step approach, building up to a general integrable function g via functions $\chi_{\mathbf{J}}$ ($\mathbf{J}$ a bounded rectangle), step functions, and $\mathrm{L}^{\mathrm{inc}}$ functions.

Stage 1. Take $g = \chi_{\mathbf{J}}$, where $\mathbf{J}$ is a bounded open rectangle. We have $\chi_{\mathbf{J}}(T(x,y)) = 1$ if and only if $(x,y) \in T^{-1}(\mathbf{J})$. We require to show that

$$m(\mathbf{J} \cap T(U)) = \int_{T^{-1}(\mathbf{J}) \cap U} \mu.$$

Note that $T^{-1}(\mathbf{J}) \cap U$ is an open subset of U since T is continuous, and that $T(T^{-1}(\mathbf{J}) \cap U) = \mathbf{J} \cap T(U)$ as T is one-to-one. Therefore we want to show that

$$m(T(W)) = \int_W \mu \quad (W \text{ open, } W \subseteq U),$$

with the convention that if either side is not finite then the other is infinite too.

Fix an open subset B of U. Then B can be written as a countable union of disjoint bounded open rectangles $\{\mathbf{I}_n\}$, as in Lemma 25.7. Let

$$\mathbf{K}_n := \mathbf{I}_1 \cup \ldots \cup \mathbf{I}_n.$$

Because the rectangles are disjoint, $\{\chi_{\mathbf{K}_n}\}$ is an increasing sequence of integrable functions, converging to χ_B. Also $\{\chi_{T(\mathbf{K}_n)}\}$ is increasing, since the one-to-one map T preserves disjointness, and converges to $\chi_{T(B)}$. In case $m(T(B))$ and $\int_B \mu$ are finite,

$$\begin{aligned} m(T(B)) &= \sum_{k=1}^{\infty} m(T(\mathbf{I}_k)) && (\text{using MCT on } \{\chi_{T(\mathbf{K}_n)}\}) \\ &= \sum_{k=1}^{\infty} \int_{\mathbf{I}_k} \mu && (\text{by assumption}) \\ &= \int_B \mu && (\text{using MCT on } \{\mu\chi_{\mathbf{K}_n}\}). \end{aligned}$$

If one of $m(T(B))$ and $\int_B \mu$ is infinite, then so is the other, by 13.2 (see 21.11).
Stage 2. Any bounded rectangle may be written as the difference of open subsets of U with one contained in the other (Exercise 25.4). By (L) the theorem holds for $g = \chi_{\mathbf{J}}$ for any bounded rectangle $\mathbf{J}$.

Stages 3 and 4. To finish the proof it is enough, by linearity again, to assume $g \in \mathrm{L}^{\mathrm{inc}}(\mathbb{R}^2)$. Take an $\mathrm{L}^{\mathrm{inc}}(\mathbb{R}^2)$-sequence $\{\varphi_n\}$ for g, with $\varphi_n \nearrow g$ except on a null set E. We first use Technical Theorems I and II (in the same manner as in the proof of Fubini's Theorem) to show $T(E)$ is null. Because E is null, 9.11 supplies a non-negative decreasing sequence $\{\psi_n\}$ in $\mathrm{L}^{\mathrm{step}}(\mathbb{R}^2)$ such that $\{\int \psi_n\}$ converges and $\{\psi_n(u,v)\}$ diverges for $(u,v) \in E$. From stage 1,

$$\int_U \mu(x,y)\psi_n(T(x,y))\,\mathrm{d}(x,y) = \int_{T(U)} \psi_n(u,v)\,\mathrm{d}(u,v),$$

so that $\{\int_U (\psi_n \circ T)\mu\}$ converges. But $\{\psi_n(T(x,y))\mu(x,y)\}$ diverges for $(x,y) \in F := T^{-1}(E) \cap P$, where $P := \{\,(x,y) \mid \mu(x,y) \neq 0\,\}$. By 9.9, F is null. Also, using the fact that μ is non-negative,

$$\varphi_n(T(x,y))\mu(x,y) \nearrow g(T(x,y))\mu(x,y) \quad ((x,y) \notin F),$$

Finally, the MCT (twice) implies that

$$\int_{T(U)} g = \lim \int_{T(U)} \psi_n = \lim \int_U (\psi_n \circ T)\mu = \int_U (g \circ T)\mu. \qquad \square$$

27.5 Theorem. Let $f\colon U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^2$ and let $U^* := \{\,(r,\theta) \mid (r\cos\theta, r\sin\theta) \in U\,\}$. Suppose $f(x,y) = f^*(r,\theta)$ (where $x = r\cos\theta, y = r\sin\theta$). Then $f(x,y)$ is integrable on U if and only if $rf^*(r,\theta)$ is integrable on U^* and in that case

$$\int_U f(x,y)\,\mathrm{d}(x,y) = \int_{U^*} rf^*(r,\theta)\,\mathrm{d}(r,\theta).$$

Proof. To obtain the 'if and only if' statement we need to apply Theorem 27.4 both to the transformation T from polars to cartesians and to the inverse transformation from cartesians to polars. The only awkwardness is that our open set U may intersect the non-negative real axis, $L := \{\,(x,0) \mid x \geqslant 0\,\}$, which we excluded in order to obtain a one-to-one map T. However this causes no problem because $W := U \setminus U \cap L$ is open and L is null in $\mathbb{R}^2$. $\square$

27.6 Example. We give another, proof that $\int_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}/2$; this has the merit of being very natural. Let

$$f(x,y) = \mathrm{e}^{-(x^2+y^2)} \quad \text{on } U := (0,\infty) \times (0,\infty),$$

so that

$$rf^*(r,\theta) = r\mathrm{e}^{-r^2} \quad \text{on } U^* := (0,\infty) \times (0,\pi/2).$$

The latter function is continuous, and hence measurable. Also

$$\int_0^{\pi/2}\left\{\int_0^{\infty}\left|re^{-r^2}\right| dr\right\} d\theta = \int_0^{\pi/2}\left[-\frac{1}{2}e^{-r^2}\right]_0^{\infty} d\theta = \frac{\pi}{4}.$$

By Tonelli's Theorem, $rf^*(r,\theta) = re^{-r^2}$ is integrable on U^*. We conclude that $f(x,y) = e^{-(x^2+y^2)}$ is integrable on U and

$$\begin{aligned}
\int_0^{\pi/2}\left\{\int_0^{\infty} re^{-r^2}\, dr\right\} d\theta &= \int_{U^*} re^{-r^2}\, d(r,\theta) && \text{(by Fubini's Theorem)}\\
&= \int_U e^{-(x^2+y^2)}\, d(x,y) && \text{(by 27.5)}\\
&= \int_0^{\infty}\left\{\int_0^{\infty} e^{-(x^2+y^2)}\, dx\right\} dy && \text{(by Fubini's Theorem)}\\
&= \int_0^{\infty}\left\{e^{-y^2}\int_0^{\infty} e^{-x^2}\, dx\right\} dy\\
&= \left(\int_0^{\infty} e^{-u^2}\, du\right)^2.
\end{aligned}$$

The left-hand side is $\frac{1}{4}\pi$, so $\int_0^{\infty} e^{-u^2}\, du = \frac{1}{2}\sqrt{\pi}$.

Remark. Note the way in which we have combined the transformation theorem 27.5 with the theorems of Fubini and Tonelli in order to show that repeated integrals transform as anticipated under our change of coordinates. We reiterate that Theorem 27.4 and its corollary Theorem 27.5 concern $L(\mathbb{R}^2)$ integrals and 2-dimensional measure. They are not theorems about repeated integrals.

27.7 Example. Let $f(x,y) = (x-y)(x^2+y^2)^{\alpha}$ on $S := [0,1]\times[0,1]$. We claim that f is integrable on S if and only if $\alpha > -3/2$. First observe that f is continuous except at $(0,0)$. By 26.8, f is integrable on any closed subset of S excluding $(0,0)$. Therefore it is the behaviour of f near $(0,0)$ we must consider. It will be enough to show that $f\chi_U$ is integrable if and only if $\alpha > -3/2$, where

$$U := \{(x,y) \mid x^2+y^2 < 1\} \quad \text{and} \quad U^* = \{(r,\theta) \mid r < 1\}.$$

Further,

$$rf^*(r,\theta) = (\cos\theta - \sin\theta)\, r^{2\alpha+2},$$

and

$$\int_0^{2\pi}\left\{\int_0^1\left|(\cos\theta-\sin\theta)r^{2\alpha+2}\right| dr\right\} d\theta = \int_0^{2\pi}|\cos\theta-\sin\theta|\, d\theta \int_0^1 r^{2\alpha+2}\, dr < \infty$$

if and only if $2\alpha+2 > -1$, by 15.1. Hence $rf^*(r,\theta)$ is integrable if $\alpha > -3/2$ by Tonelli's Theorem, and is not integrable if $\alpha \leqslant -3/2$ by the contrapositive of Fubini's Theorem.

As another application of Theorem 27.4 we prove the measurability result previewed in 25.6. We also use it in Chapter 33.

27.8 Lemma. Let $f\colon \mathbb{R} \to \mathbb{R}$ be measurable. Define $g\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$g(x,y) := f(x-y).$$

Then g is measurable.

Proof. We use the Truncation Lemma in its 2-dimensional form. It will be sufficient to check that, for all n,

$$g^{\square_n} := (|g| \wedge n)\chi_{\mathbf{I}_n} \in \mathrm{L}(\mathbb{R}^2) \quad \text{where } \mathbf{I}_n := [-n,n] \times [-n,n].$$

Let T be the linear change of variables given by

$$T(x,y) = (x-y, x+y) = (u,v),$$

so $x = \frac{1}{2}(u+v)$ and $y = \frac{1}{2}(v-u)$, and $\mu = 2$. The function h given by $h(u,v) = f(u)$ is measurable, by Exercise 25.7, and so each truncation $h^{\square_k}$ $(:= (|h| \wedge k)\chi_{\mathbf{I}_k})$ is integrable. For each n, $T(\mathbf{I}_n)$ is a closed bounded rectangle and so its characteristic function is bounded and measurable (see 27.2). Then

$$h\chi_{T(\mathbf{I}_n)} = h^{\square_{2n}}\chi_{T(\mathbf{I}_n)} \in \mathrm{L}(\mathbb{R}^2).$$

Therefore Theorem 27.4 tells us that $(h \circ T)\mu\chi_{\mathbf{I}_n}$ is integrable too. But

$$h(T(x,y)) = h(u,v) = f(u) = f(x-y),$$

so $g^{\square_n} = |h \circ T|\,\chi_{\mathbf{I}_n}$ is integrable for any n. By the Truncation Lemma, g is measurable. $\square$

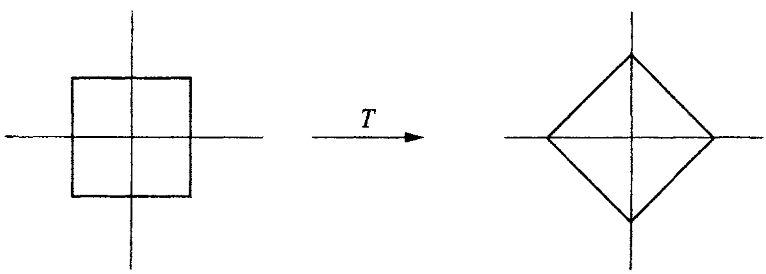


Figure 27.2

We conclude our account of multi-dimensional integrals by stating without proof a full-blown transformation theorem. The proof can be found, for example, in [**18**]. A *coordinate transformation* of an open subset U of $\mathbb{R}^k$ is a one-to-one map $T\colon U \to T(U) \subseteq \mathbb{R}^k$ such that T is continuously differentiable, with the Jacobian determinant $\mathbf{J}_T$ non-zero at each point of U. These conditions are sufficient to ensure that T has an inverse map with the same properties.

We have shown, for $k=2$, that the special transformations considered in 27.2 and 27.3 are coordinate transformations in the sense of the definition above. Further, the density function identified in each case is the Jacobian determinant of the transformation. Our simple transformation theorem presupposes a density function; the general theorem shows that every coordinate transformation has one. Further, the general theorem treats general measurable sets and functions, whereas 27.4 restricts attention to bounded rectangles and integrable functions. As a special case of (b) below we see that a coordinate transformation maps null sets to null sets. The most useful transformations of $\mathbb{R}^3$ are linear transformations, and switches between cartesian coordinates and either cylindrical or spherical polar coordinates. All these fall within the scope of Theorem 27.9. For more details see [1] or [18].

27.9 Theorem. Let U be an open subset of $\mathbb{R}^k$ and let $T\colon U \to \mathbb{R}^k$ be a coordinate transformation.

(a) The Jacobian determinant $|\mathbf{J}_T|$ is a density function μ for T, so that

$$m(T(\mathbf{I})) = \int_{\mathbf{I}} |\mathbf{J}_T|$$

for any bounded interval $\mathbf{I} \subseteq U$.

(b) For any measurable set S with $S \subseteq U$, $T(S)$ is measurable and

$$m(T(S)) = \int_S \mu.$$

(c) For $g\colon T(U) \to \mathbb{R}$,

$$\int_{T(U)} g = \int_U (g \circ T)\mu,$$

in the sense that if either side is defined, then so is the other and the integrals coincide.

Exercises

27.1 This exercise concerns the claim made in 27.3 concerning the integral of the characteristic function of a bounded parallelogram $\mathbf{J}$, obtained by applying the linear map $Y\colon (x,y) \mapsto (ax+by, cx+dy)$ to the bounded rectangle $\mathbf{I}$. Note that we may assume without loss of generality that $\mathbf{J}$ is closed (see Exercise 25.3) and, thanks to Properties (L) and (T) we may take $\mathbf{I} = [0,1] \times [0,1]$, so that $\mathbf{J}$ is the closed parallelogram with vertices $(0,0)$, (a,c), (b,d), and $(a+b, c+d)$. Take $0 < a < c$ and $0 < d < b$ (the calculations for other cases are similar).

(a) Use 26.8 to verify that $\chi_{\mathbf{J}}$ is integrable.

(b) Apply Fubini's Theorem to obtain

$$\int \chi_{\mathbf{J}}(x,y)\,\mathrm{d}(x,y) = \int \left\{ \int \chi_{\mathbf{J}}(x,y)\,\mathrm{d}x \right\} \mathrm{d}y$$

and check that this repeated integral equals $bc - ad$. [Hint: subdivide $\mathbf{J}$ into suitable subsets, corresponding to values of s lying in $[0,d)$, $[d,b)$, and $[b, b+d]$; a diagram will help.]

27.2 Prove by switching to polars that

(i) $\dfrac{(x-y)\sin(xy)}{(x^2+y^2)^2}$ is integrable over $[-1,1]\times[-1,1]$,

(ii) $\dfrac{(x-y)\cos(xy)}{(x^2+y^2)^2}$ is not integrable over $[-1,1]\times[-1,1]$.

27.3 Let

$$f_{\alpha,\beta}(x,y)=\frac{(xy)^\beta}{(x^2+y^2+1)^\alpha}\quad (0<\alpha<\infty,\ 0<\beta<\infty).$$

(a) Prove that $f\in \mathrm{L}(\mathbb{R}^2)$ if $\alpha-\beta>1$.

(b) Evaluate $\int f_{\alpha,1}$ $(\alpha>1)$.

(c) Use the Cauchy–Schwarz inequality to prove that, for $\alpha-\beta>1$,

$$\int f_{\alpha,\beta}\leqslant\frac{2^{1-\beta}}{(\alpha-\beta)(\alpha-\beta-1)}.$$

27.4 For $p,q>0$, define the Beta function by

$$B(p,q)=2\int_0^{\pi/2}\cos^{2p-1}\theta\sin^{2q-1}\theta\,\mathrm{d}\theta.$$

Prove that $\Gamma(p+q)B(p,q)=\Gamma(p)\Gamma(q)$, where Γ is the Gamma function defined in Exercise 16.3.

28 The spaces L^1, L^2, and L^p

We have hitherto exploited the vector space structure of L, and dealt with various approximating sequences to integrable functions: $\mathrm{L}^{\mathrm{inc}}$-sequences, truncations, etc. This chapter brings these ideas together in a systematic way and sets spaces of integrable functions in their rightful place in the mathematical orchestra. We can play here only a snatch of the overture to the symphony in which integration theory and functional analysis come together in concert. More advanced texts take up the refrain from where we are forced to leave it.

28.1 Normed and seminormed spaces. To say that one function, g, is a good approximation to another, f, we need a precise way of measuring the error, $f - g$. A natural way to do this is to assign a non-negative number, $\|k\|$, to $k := f - g$, measuring its size. The properties we shall require of the function $k \mapsto \|k\|$ mimic those of the modulus function. Given points a, b in $\mathbb{C}$, the modulus $|a - b|$ may be interpreted as the distance between a and b, and $|a|$ as the length of a, regarded as a vector. Further, modulus interacts well with the vector space structure of $\mathbb{C}$: it defines what is known as a norm. Given a real (or complex) vector space X a function $\|.\|: X \to \mathbb{R}^+$ is a *norm* if

(NS1) for all $a \in X$, $\|a\| = 0$ if and only if $a = 0$;
(NS2) for all $a \in X$ and $\lambda \in \mathbb{R}$ ($\lambda \in \mathbb{C}$), $\|\lambda a\| = |\lambda|\,\|a\|$;
(NS3) for all $a, b \in X$, $\|a + b\| \leqslant \|a\| + \|b\|$ (triangle inequality).

The structure $(X; \|.\|)$ is called a *normed space.* Any norm on X gives rise to a metric as it is defined in topology and analysis: $d(a, b) := \|a - b\|$. If $\|.\|$ satisfies just (NS2) and (NS3) then it is said to be a *seminorm.*

Henceforth we generally work with complex-valued functions and consequently with complex normed spaces. All these spaces, and all the results we prove about them, have real analogues.

28.2 Examples.

(1) Consider the real vector space $\mathbb{R}^2$. We can define a norm, $\|.\|_2$, by $\|\mathbf{a}\|_2 := \sqrt{a_1^2 + a_2^2}$, where $\mathbf{a} = (a_1, a_2)$. The norm $\|.\|_2$ is associated with the Euclidean distance, d, on $\mathbb{R}^2$:

$$\|\mathbf{a} - \mathbf{b}\|_2 = d(\mathbf{a}, \mathbf{b}) := \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}.$$

It is related to the usual scalar product $\mathbf{a} \cdot \mathbf{b} := a_1 b_1 + a_2 b_2$ on $\mathbb{R}^2$ through the equation $\|\mathbf{a}\|_2^2 = \mathbf{a} \cdot \mathbf{a}$. This link leads to special properties of the norm. For instance, we have Pythagoras' Theorem:

$$\|\mathbf{a} + \mathbf{b}\|_2^2 = \|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2 \quad \text{if } \mathbf{a} \cdot \mathbf{b} = 0.$$

We can define another norm, $\|.\|_1$, on $\mathbb{R}^2$ by $\|\mathbf{a}\|_1 := |a_1| + |a_2|$. Certainly $\|.\|_2$ and $\|.\|_1$ are not the same: $\|(1,1)\|_1 = 2$ and $\|(1,1)\|_2 = \sqrt{2}$.

(2) $\|f\|_\infty := \sup\{|f(x)| \mid x \in S\}$ defines a norm (the *supremum norm*) on any vector space of bounded real- or complex-valued functions on a set S. In particular, $\mathrm{C}[a,b]$ is a normed space for the supremum norm.

(3) Take $\|f\|_1 := \int |f|$ on L. Here we have a seminorm, but not a norm: (NS1) fails because $\|f\|_1 = 0$ if and only if $f = 0$ **a.e.** Below we show how to convert L into a normed space, L^1, and also how to define a 'Euclidean-like' space, L^2, of 'square-integrable functions'.

28.3 From a seminorm to a norm: $\mathcal{L}^1$ and L^1. Consider $\mathcal{L}^1 := (\mathrm{L}, \|\ .\ \|_1)$. As we have said, this fails to be a normed space only because (NS1) fails. Suppose we neglect to distinguish between functions which are equal a.e.—an entirely reasonable attitude since we don't care what a function does on a null set. Then we treat $\{\, g \in \mathcal{L}^1 \mid g = f \text{ a.e.}\,\}$ as a single function, $[f]$. The set $\mathrm{L}^1 := \{\,[f] \mid f \in \mathrm{L}\,\}$ can be made into a genuine normed space on which $\|[f]\| := \|f\|_1$. The formal details of this procedure are outlined in Exercise 28.2. It is customary to blur the distinction between $\mathcal{L}^1$ and L^1, treating the elements of L^1 as the elements of L, but remembering that '=' means '= a.e.'.

28.4 $\mathcal{L}^2$ and L^2. Paralleling the Euclidean space $(\mathbb{R}^2, \|.\|_2)$ we define

$$\mathcal{L}^2 := \{\, f\colon \mathbb{R} \to \mathbb{C} \mid f \text{ is measurable and } |f|^2 \in \mathrm{L}\,\},$$

on which we take

$$\|f\|_2 := \left(\int |f|^2\right)^{1/2}.$$

We assert that $\mathcal{L}^2$ is a vector space over $\mathbb{C}$ for any $p \geqslant 1$, on which $\|.\|_2$ is a seminorm. The non-routine parts of this claim are

(i) $f, g \in \mathrm{L}^2$ implies $f + g \in \mathrm{L}^2$, and

(ii) the triangle inequality, (NS3).

Let $f, g \in \mathcal{L}^2$. Certainly $f+g$ is measurable. Also $(f+g)^2 = f^2 + 2fg + g^2$, so $f + g \in \mathcal{L}^2$ provided fg is integrable. We proved in 23.2(b) that it is; the assumption that functions in $\mathcal{L}^2$ are measurable being needed here. Moreover,

$$\begin{aligned}
\int |f+g|^2 &= \int |f^2 + g^2 + 2fg| \\
&\leqslant \int (|f|^2 + |g|^2 + 2|fg|) && \text{(by } (\Delta) \text{ and (P))} \\
&= \int |f|^2 + \int |g|^2 + 2\int |fg| && \text{(by (L))} \\
&\leqslant \int |f|^2 + \int |g|^2 + 2\left(\int |f|^2 \int |g|^2\right)^{1/2} && \text{(by CS)} \\
&= \left(\left(\int |f|^2\right)^{1/2} + \left(\int |g|^2\right)^{1/2}\right)^2.
\end{aligned}$$

We make the transition from the seminormed space $\mathcal{L}^2$ to the normed space L^2 in the manner described in 28.3. In Chapters 31 and 32 we exploit the fact that

$$\langle f, g\rangle := \int f(x)\overline{g(x)}\,\mathrm{d}x$$

defines an inner product on L^2, with $\langle f, f\rangle = \|f\|_2^2$.

The spaces $\mathrm{L}^2(I)$, where I is an interval in $\mathbb{R}$, and $\mathrm{L}^2(\mathbb{R}^k)$ $(k > 1)$ are defined in the expected way.

28.5 L^1 and L^2 compared. Neither of L^1 and L^2 is contained in the other:

$$x^{-1/2}\chi_{(0,1)}(x) \in \mathrm{L}^1 \setminus \mathrm{L}^2 \quad \text{and} \quad x^{-1}\chi_{(1,\infty)}(x) \in \mathrm{L}^2 \setminus \mathrm{L}^1.$$

Of course, the complex-valued versions of $\mathrm{L}^{\mathrm{step}}$ and L^{C} sit inside both L^1 and L^2.

On the other hand, if I is a **bounded** interval, then $\mathrm{L}^2(I) \subseteq \mathrm{L}^1(I)$. To prove this, first note that the constant function 1 is integrable over I. Then use the fact that $|f| = |f|.1 \leqslant \frac{1}{2}(|f|^2 + 1)$ and invoke the Comparison Theorem. Further, by the Cauchy–Schwarz inequality we have, for $f \in \mathrm{L}^2(I)$,

$$\|f\|_1^2 = \int_I |f.1|^2 \leqslant \left(\int_I |f|^2\right)\left(\int_I |1|^2\right) = \ell(I)^2\|f\|_2^2.$$

28.6 Convergence in a normed space. The notion of convergence of a sequence in $\mathbb{C}$ extends immediately to any normed (or merely seminormed) space X: we say $\{a_n\}$ *converges* to the limit $a \in X$ if given $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $n \geqslant N$ implies $\|a_n - a\| < \varepsilon$. If $\|.\|$ is a norm then the limit is unique.

Different norms on the same space in general describe different modes of convergence. For example in $\mathrm{C}[0,1]$ equipped with the supremum norm, $\{f_n\}$ converges to f if and only if $\{f_n\}$ converges uniformly to f on $[0,1]$ (recall 8.5). On the other hand, in $(\mathrm{C}[0,1], \|\cdot\|_2)$, $\{f_n\}$ converges to f if and only if $\int_0^1 |f - f_n|^2 \to 0$. This time we average out the error over the interval $[0,1]$. Such 'mean-square convergence' may well be familiar to students of probability, statistics, or numerical analysis.

28.7 Examples. Below we illustrate by examples how the convergence or non-convergence of a given sequence may depend on what mode of convergence is selected. This might be pointwise or uniform convergence, convergence in a variety of norms, or convergence in measure (important, but outside the scope of this book). The proliferation of concepts and behaviours should be regarded as an asset rather than a burden: it gives scope for mathematical modelling of varied phenomena—for example in probability—in a variety of ways.

(1) Let $f_n := n^{-1}\chi_{[0,n]}$, so that $f_n(x) \to 0$ for all $x \in \mathbb{R}$. In fact, $f_n \xrightarrow{\text{u}} 0$ on $\mathbb{R}$. Then $\|f_n\|_\infty = n^{-1} \to 0$ but $\|f_n\|_1 = 1$ for all n. Hence uniform convergence does not imply convergence in the L^1 norm.

(2) Let $\{f_n\}$ be defined on $[-1,1]$ by

$$f_n(x) := \begin{cases} -1 & \text{if } -1 \leqslant x \leqslant -\frac{1}{n}, \\ nx & \text{if } -\frac{1}{n} \leqslant x \leqslant \frac{1}{n}, \\ 1 & \text{if } \frac{1}{n} \leqslant x \leqslant 1 \end{cases}$$

(see Fig. 28.1). Pointwise on $[-1,1]$, $f_n \to f := \chi_{(0,1]} - \chi_{[-1,0)}$. Also $f \in \mathrm{L}$ and $\|f_n - f\|_1 = 1/n \to 0$ as $n \to \infty$. Note however that $\{f_n\}$ is a sequence of continuous functions whose limit f is not continuous, so that $f_n \xrightarrow{\mathrm{u}} f$ is necessarily false (invoke 8.8).

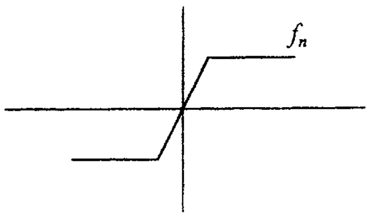


Figure 28.1

(3) A rather contrived example shows that for a sequence which converges pointwise at no point may nonetheless be norm convergent. For each $m = 1, 2, \ldots$ partition $[0, m)$ into m^2 translates of the interval $[0, m^{-1})$,

$$J_{m,r} := [0, m^{-1}) + (r-1)m^{-1} \quad (r = 1, \ldots, m^2).$$

Let $\{I_n\}$ be the sequence of intervals obtained by concatenating into a single sequence the finite strings of intervals $J_{m,1}, \ldots, J_{m,m^2}$ $(m = 1, 2, \ldots)$. Thus $\ell(I_1) = 1$, $\ell(I_2) = \cdots = \ell(I_5) = 2^{-1}$, $\ell(I_6) = \cdots = \ell(I_{14}) = 3^{-1}$, and so on. Then $\{\chi_{I_n}\}$ converges to 0 in L^1 and in L^2. Pointwise, $\{\chi_{I_n}(x)\}$ fails to converge for any $x \geqslant 0$. To see this, think how I_n shifts backwards and forwards along $\mathbb{R}^+$ as n increases: for each point x, there exist arbitrarily large r and s with $x \in I_r$ and $x \notin I_s$.

Taken together, (1) and (3) show that neither of convergence for the L^1 norm and pointwise convergence implies the other. [It can be proved that a sequence $\{f_n\}$ convergent in L^1 is forced to have a **subsequence** $\{f_{n_k}\}$ which converges pointwise a.e.]

28.8 Exercise example.

(a) Let $f_n := n^{-k}\chi_{[0,n]}$, where $k > 0$. Prove that $f_n \xrightarrow{\mathrm{u}} 0$ on $\mathbb{R}$. For which values of k if any is it true that (i) $\lim \|f_n\|_1 = 0$, (ii) $\lim \|f_n\|_2 = 0$?

(b) Let $f_n = n\chi_{(0,1/n)}$. Show that $\lim f_n(x) = 0$ for all x but that there is no $f \in \mathrm{L}^1$ for which $\|f_n - f\|_1 \to 0$.

28.9 Completeness. Fundamental to the study of convergence in $\mathbb{R}$ is the Cauchy Convergence Principle: $\{x_n\}$ converges to some limit $x \in \mathbb{R}$ if and only if $\{x_n\}$ is a *Cauchy sequence*, that is, given $\varepsilon > 0$ there exists N such that $m, n \geqslant N$ implies $|x_n - x_m| < \varepsilon$. Replacing modulus signs by norm signs we obtain the notion of a Cauchy sequence in any normed space $(X; \|.\|)$. We say $(X; \|.\|)$ is *complete* if every Cauchy sequence converges to a limit in X. The Cauchy Convergence Principle says exactly that $(\mathbb{R}; |.|)$ is complete, and we may ask whether other spaces of interest to us are also complete. A complete normed space is called a *Banach space*.

28.10 An example of non-completeness. Consider again Example 28.2(2). Regard $\{f_n\}$ as lying in $\mathrm{C}[-1,1]$, which we equip with $\|.\|_1$. It is easy to see that $\{f_n\}$ is a Cauchy sequence. For any $n \geqslant m$, $f_n - f_m = 0$ except on $(-1/m, 1/m)$ where the discrepancy is not more than 1. Therefore $\|f_n - f_m\|_1 \leqslant 2m^{-1}$.

We have already shown that the pointwise limit, f, of $\{f_n\}$ is such that $\|f_n - f\|_1 \to 0$ but that $f \notin \mathrm{C}[-1,1]$. We now make the stronger claim that there is no function $g \in \mathrm{C}[-1,1]$ such that $\|f_n - g\|_1 \to 0$. Suppose, for a contradiction, that g does exist. Then

$$\int |f - g| \leqslant \int |f - f_n| + \int |f_n - g|,$$

Since the right-hand side can be made arbitrarily small by taking n large enough, we conclude that $\int |f-g| = 0$ so that $f = g$ a.e. This forces $f = g$ on each of $[-1,0)$ and $(0,1]$, on which f is continuous (recall Exercise 11.1). But continuity of g would then imply both $g(0) = -1$ and $g(0) = 1$, which gives the required contradiction. We conclude that $\mathrm{C}[-1,1]$ is not complete for $\|.\|_1$.

28.11 Exploiting non-completeness: density. The preceding example tells us that by taking limits of continuous functions in $\mathrm{L}^1[-1,1]$ we may go outside $\mathrm{C}[-1,1]$. This fact is good news! Without it we would not be able to approximate integrable functions on $[-1,1]$ by continuous functions in a manner appropriate to integration, that is, measuring distance by $\|.\|_1$.

In general we say that a subset $\mathcal{E}$ of a normed space $(X; \|.\|)$ is *dense* if for each $x \in X$ we can find, given any $\varepsilon > 0$, a point $y \in \mathcal{E}$ such that $\|x - y\| < \varepsilon$. For instance, $\mathbb{Q}$ is dense in $(\mathbb{R}, |\,.\,|)$. We may say that our entire theory of integration is founded on the notion of density. We set up the integral in an elementary, hands-on, way for a suitable class $\mathcal{E}$ of functions, and by taking limits extended the integral to a class $\mathcal{F}$ in which $\mathcal{E}$ is dense for a suitable norm. In our approach $\mathcal{E}$ is $\mathrm{L}^{\mathrm{step}}$ and the space $(\mathcal{F}, \|.\|)$ is $(\mathrm{L}^1, \|.\|_1)$. Other approaches start from different sets $\mathcal{E}$ but still finish up with the same space of integrable functions; see Appendix I.

A common strategy for proving a result about $\mathrm{L}_{\mathbb{R}}$ is to prove it first for functions in some amenable dense subset. The archetypal example is the proof of the Riemann–Lebesgue Lemma, 23.4. Other examples are given in 23.5 and Exercise 30.3. Here are the density results we have already established.

(a) $\mathrm{C}[a,b]$ is dense in $\mathrm{L}^1[a,b]$, for any compact interval $[a,b]$ (Exercise 3.5).

(b) Each of the following is dense in L^1:
   (i) $\mathrm{L}^{\mathsf{step}}$;
   (ii) bounded integrable functions of compact support (use the Truncation Lemma, 21.2, or (iii) below);
   (iii) C^∞ functions of compact support.

All the statements above remain true—and are very useful—when L^1 is replaced by L^2. We prove the L^2 version of (b)(i), from which the other assertions follow quite easily. The proof is technical, and much harder than that for the L^1 case.

28.12 Lemma. Let $f \in \mathrm{L}^2$. Then, given $\varepsilon > 0$, there exists a step function φ such that $\|f - \varphi\|_2 < \varepsilon$.

Proof. We prove the result in the case that $f \geqslant 0$. To complete the proof when f is real-valued we consider f^+ and f^-, and in the complex-valued case we take real and imaginary parts.

Our strategy is first to manufacture an approximating sequence for f which lies in $\mathrm{L}^1 \cap \mathrm{L}^2$. For this we take the functions

$$k_n := f \wedge n(f^2)^{\square_n} \quad (n = 1, 2, \dots);$$

remember that the truncation $g^{\square_n}$ is defined to be $(n \wedge |g|)\chi_{[-n,n]}$. We have

(a) k_n is measurable;

(b) $0 \leqslant k_n \leqslant n(f^2)^{\square_n}$ so, by the Truncation Lemma, 21.2, applied to f^2, and the Comparison Theorem, $k_n \in \mathrm{L}$;

(c) $0 \leqslant k_n \leqslant f$ so $k_n^2 \leqslant f^2$ and, again by the Comparison Theorem, $k_n \in \mathrm{L}^2$;

(d) $f - k_n \in \mathrm{L}^2$ and $|f - k_n|^2 \leqslant f^2$ (all by (c));

(e) $k_n \xrightarrow{\text{a.e.}} f$ as $n \to \infty$ (because of the multiplier n in the second term in k_n);

(f) $0 \leqslant k_n \leqslant n^2$ (because $(f^2)^{\square_n} \leqslant n$).

Using (d) and (e) we can apply the DCT to $\{|f - k_n|^2\}$ to get $\lim \|f - k_n\|_2^2 = \lim \int |f - k_n|_2^2 = 0$. Choose N such that $\|f - k_N\|_2 < \varepsilon$. Now we apply 12.13 to the integrable function k_N to find $\psi \in \mathrm{L}^{\mathsf{step}}$ such that $\int |k_N - \psi| < \varepsilon^2/(2N^2)$. Replacing ψ by the truncation $\varphi := \psi^{\square_N}$ can only improve the approximation because $|k_N - \varphi| \leqslant |k_N - \psi|$. Since each of k_N and φ is bounded by N^2,

$$\int |k_N - \varphi|^2 \leqslant 2N^2 \int |k_N - \varphi| < \varepsilon^2.$$

Therefore $\|k_N - \varphi\|_2 < \varepsilon$ and by (NS3) we get $\|f - \varphi\|_2 < 2\varepsilon$. $\square$

28.13 The benefit of completeness. Just as completeness of $\mathbb{R}$ enriches its theory, completeness of other normed spaces ensures that they have good properties. Indeed we shall shortly see that the key convergence theorems are closely tied to the completeness of L^1. Further, there is a very well-developed and powerful theory of complete normed spaces (Banach spaces) in general, on which advanced integration theory draws very heavily.

28.14 A criterion for completeness. For a complex series, absolute convergence implies convergence: $\sum |a_k|$ converges implies that $\sum a_k$ converges. 'Seeing double' and replacing modulus signs by norm signs, the same is true in any complete normed space X. To check this, let $a_k \in X$ $(k = 1, 2, \dots)$, write $s_n := a_1 + \cdots + a_n$ and $S_n := \|a_1\| + \cdots + \|a_n\|$ and assume $S_n \to S$. By the triangle inequality first in X and then in $\mathbb{R}$,

$$\begin{aligned}\|s_n - s_m\| &= \|a_n + \cdots + a_{m+1}\| \leqslant \|a_n\| + \cdots + \|a_{m+1}\| \\ &= |S_n - S_m| \leqslant |S_n - S| + |S_m - S|,\end{aligned}$$

and the last expression tends to 0 as $m, n \to \infty$. Thus $\{s_n\}$ is a Cauchy sequence in X, and hence converges.

More interesting is the fact that there is a converse to the result just proved. We shall show that a normed space X is complete if for any sequence $\{y_k\}$ in X, $\sum \|y_k\|$ convergent implies $\sum y_k$ converges in X. Let $\{x_n\}$ be a Cauchy sequence in X. We must show $\{x_n\}$ converges. To do this we use the telescoping technique, manufacturing a series whose partial sums form a subsequence of $\{x_n\}$ (cf. 2.19). Choose $n_1 < n_2 < \dots$ in $\mathbb{N}$ such that

$$\|x_m - x_n\| < 2^{-k} \quad \text{for } n, m \geqslant n_k.$$

Let $y_k := x_{n_k} - x_{n_{k-1}}$ $(k \geqslant 2)$, $y_1 = 0$. Then $\sum \|y_k\|$ converges, by comparison with $\sum 2^{-k}$. By hypothesis, $\sum y_k$ converges. That is, there exists $x \in X$ such that

$$x_{n_k} = y_1 + \cdots + y_k \to x \quad \text{as } k \to \infty.$$

But a Cauchy sequence $\{x_n\}$ with a subsequence converging to x itself converges to x (why?). □

28.15 Theorem. The normed space L^1 is complete.

Proof. By the preceding result it suffices to show that if $\{f_k\}$ is a sequence in L^1 such that $\sum \|f_k\|_1$ converges then $\sum f_k$ converges to a function in L^1. This, phrased in terms of $\mathcal{L}^1$, is just what 17.5 gives. □

The corresponding L^2 result underpins Chapter 32. The strategy for proving it is essentially the same as that used in 17.5.

28.16 Completeness of L^2. Let $\{u_k\}$ be a sequence in $\mathcal{L}^2$ such that $\sum \|u_k\|_2$ converges. Then

(a) $\sum u_k$ is pointwise (absolutely) convergent a.e. to a function f in $\mathcal{L}^2$;
(b) $\|f - \sum_{k=1}^{n} u_k\|_2 \to 0$ as $n \to \infty$.

As a consequence the normed space L^2 is complete (the **Riesz–Fischer Theorem**).

Proof. We first apply the MCT to $\{f_n\}$, where

$$f_n := \Big(\sum_{k=1}^{n} |u_k|\Big)^2,$$

noting that (M3) holds because

$$\int f_n \leqslant \sum_{k=1}^{n} \|u_k\|_2 \leqslant \sum_{k=1}^{\infty} \|u_k\|_2.$$

Therefore there exists a real-valued function f such that $f_n \to f$ a.e., and f is integrable Hence $\sum_{k=1}^{\infty} u_k$ is absolutely convergent a.e. and there exists f such that $f = \sum_{k=1}^{n} u_k$ a.e. Also, f is measurable and $f^2 \in \mathrm{L}$. This proves (a).

For (b) we apply the DCT to $\{|f - \sum_{k=1}^{n} u_k|^2\}$. To get a dominating function we use the extended triangle inequality (Exercise 2.9). We have

$$\Big|f - \sum_{k=1}^{n} u_k\Big|^2 = \Big|\sum_{k=n+1}^{\infty} u_k\Big|^2 \leqslant \Big(\sum_{k=n+1}^{\infty} |u_k|\Big)^2 \leqslant |f|^2.$$

Hence $\lim_{n\to\infty} \int |f - \sum_{k=1}^{n} u_k|^2 = \int \lim_{n\to\infty} |f - \sum_{k=1}^{n} u_k|^2 = 0.$ □

The remainder of this chapter is not needed subsequently.

28.17 The Lebesgue spaces L^p. Let $1 \leqslant p < \infty$. Define

$$\mathcal{L}^p := \{\, f\colon \mathbb{R} \to \mathbb{C} \mid f \text{ is measurable and } |f|^p \in \mathrm{L}\,\},$$

on which we take

$$\|f\|_p := \Big(\int |f|^p\Big)^{1/p}.$$

We should like to prove $\mathcal{L}^p$ is a seminormed space, in anticipation of forming an associated normed space, L^p. So, as in the case $p = 2$, we want to prove that $\mathcal{L}^p$ is closed under pointwise addition and that (NS3) holds. These facts follow from two famous inequalities.

28.18 Hölder's inequality. Let $f \in \mathrm{L}^p$ and $g \in \mathrm{L}^q$ where $p^{-1} + q^{-1} = 1$ $(1 < p < \infty)$. Then $fg \in \mathrm{L}^1$ and $\|fg\|_1 \leqslant \|f\|_p \|g\|_q$.

Proof. If $\|f\|_p = 0$ then $f = 0$ a.e. and the result is trivial, and likewise if $\|g\|_q = 0$. So assume $\|f\|_p \neq 0$ and $\|g\|_q \neq 0$. Now apply the inequality

$$AB \leqslant \frac{A^p}{p} + \frac{B^q}{q} \quad (A, B \geqslant 0)$$

from Exercise 2.10 with $A = f(x)/\|f\|_p$ and $B = g(x)/\|g\|_q$. We obtain

$$\frac{|f(x)g(x)|}{\|f\|_p\|g\|_q} \leqslant \frac{|f(x)|^p}{p\|f\|_p} + \frac{|g(x)|^q}{q\|g\|_q}.$$

Since f and g are measurable, so is fg, and hence by the Comparison Theorem, fg is integrable. Further, the inequality above gives

$$\frac{\|fg\|_1}{\|f\|_p\|g\|_q} \leqslant \frac{1}{p} + \frac{1}{q} = 1,$$

from which Hölder's inequality follows. □

28.19 Minkowski's inequality. Let $f, g \in \mathrm{L}^p$ $(1 < p < \infty)$. Then $f + g \in \mathrm{L}^p$ and $\|f + g\|_p \leqslant \|f\|_p + \|g\|_p$.

Proof. By 21.5, $(f+g)^p$ is measurable. For any x,

$$|f(x) + g(x)|^p \leqslant 2\max\{|f(x)|^p, |g(x)|^p\} \leqslant 2^p\left(|f(x)|^p + |g(x)|^p\right).$$

By the Comparison Theorem, $(f+g)^p$ is integrable. Moreover,

$$|f+g|^p = |f+g|\,|f+g|^{p-1} \leqslant |f|\,|f+g|^{p-1} + |g|\,|f+g|^{p-1}.$$

Since $p = (p-1)q$, $(f+g)^{p-1} \in \mathrm{L}^q$. Applying Hölder's inequality to f and $(f+g)^{p-1}$ we have

$$\int |f|\,|f+g|^{p-1} \leqslant \|f\|_p \left(\int |f+g|^{(p-1)q}\right)^q = \|f\|_p \|f+g\|_p^{p/q}.$$

We may handle $|g|\,|f+g|^{p-1}$ similarly. We obtain

$$\|f+g\|_p^p \leqslant (\|f\|_p + \|g\|_p)\|f+g\|_p^{p/q}.$$

If $\|f+g\|_p \neq 0$, then the required inequality follows on dividing by $\|f+g\|_p^{p/q}$, because $p - (p/q) = 1$. The result is trivial if $\|f+g\|_p = 0$. □

28.20 The theory of L^p spaces: a fleeting glimpse. The normed space L^p is obtained from the seminormed space $\mathcal{L}^p$ in the expected way, by identifying functions which are equal a.e. The proof of completeness for the case $p = 2$ is easily adapted to prove that L^p is complete for any $p \geqslant 1$; we leave this as an exercise.

There is a close relationship between the spaces L^p and L^q when the indices p and q are linked by $p^{-1} + q^{-1} = 1$ $(p, q > 1)$. At its most elementary this is visible in Hölder's inequality. At a deeper level, the Banach spaces L^p and L^q are dual to one another in the sense that the term 'dual' is used in functional analysis. Specifically, given any continuous linear map $\varphi\colon \mathrm{L}^p \to \mathbb{C}$ there exists $g \in \mathrm{L}^q$ such that

$$\varphi(f) = \int f\overline{g} \quad (f \in \mathrm{L}^p).$$

For a full discussion of this important result we refer the reader to more advanced texts, such as **[15]** or **[14]**. This pairing of L^p and $\mathrm{L}^{p/(p-1)}$ for $p > 1$ leaves L^1 without a mate. An appropriate partner for L^1 is known as L^∞. It is obtained from a seminormed space, $\mathcal{L}^\infty$, by the usual expedient of identifying functions equal a.e. Here $\mathcal{L}^\infty$ consists of the measurable functions which are essentially bounded (see 23.6 for the definition). For such a function f, $\|f\|_\infty := \sup\{\,|f(x)| \mid x \notin E\,\}$, where E is some null set off which f is bounded.

Finally we should mention in passing the Sobolev spaces. These are important in areas of mathematics where geometry, analysis, and differential equations come together. They provide a framework within which to analyse, for example, the degree of smoothness of the limit of a sequence of functions whose sequences of derivatives, up to the kth, converge appropriately. The Sobolev spaces interact with the L^p spaces through one of several different, but proveably equivalent, definitions. The space $W^{k,p}$ can be defined as the Banach space obtained as the completion to a Banach space of a normed space on which $\|f\| = \sum_{j=0}^{k} \|f^{(j)}\|_p$.

Exercises

28.1 Let X be a complete normed space and let Y be a subspace of X.

(a) Show that Y is complete if and only if whenever $\{y_n\}$ is a sequence in Y such that $y_n \to x \in X$ then $x \in Y$. [This is the requirement that Y is closed in the metric topology derived from the norm.]

(b) Show that if Y is both dense and complete then $Y = X$.

28.2 Let $\|.\|$ be a seminorm on a complex vector space $\mathcal{X}$. Define a relation $\sim$ on $\mathcal{X}$ by $u \sim v$ if and only if $\|u - v\| = 0$.

(a) Prove that $\sim$ is an equivalence relation.

(b) Let $[u] := \{\,v \in \mathcal{X} \mid v \sim u\,\}$ be the equivalence class of $u \in \mathcal{X}$ and let X denote the set of such equivalence classes. Define

$$[u] + [v] := [u + v], \quad \alpha[u] := [\alpha u] \qquad (u, v \in \mathcal{X},\ \alpha \in \mathbb{C}).$$

Check that these operations are well defined (that is, $[u] = [u'], [v] = [v']$ implies $[u + v] = [u' + v']$ and $[\alpha u] = [\alpha u']$). Show further that these operations make X into a vector space. [Those who know about quotient vector spaces may prefer to show that $[0]$ is a subspace, U say, of $\mathcal{X}$, and to think of X as the quotient space $\mathcal{X}/U$.]

(c) Define $\|[u]\| := \|u\|$ for $u \in \mathcal{X}$. Show that this makes X into a normed space.

28.3 Let I be a bounded interval. Exploit the fact that $L^2(I) \subseteq L^1(I)$ to give a proof, simpler than that in 28.12, of the density of $L^{\text{step}}(I)$ in $L^2(I)$.

28.4 Let I be a bounded interval and let $\{f_n\}$ be a sequence in $L^p(I)$ $(p \geqslant 2)$. Prove that if $f_n \to f$ in the L^p norm then $f_n \to f$ in the L^1 norm, and that the former occurs whenever $f_n \xrightarrow{u} f$ on I.

28.5 Let $f \in L^2$. By using the fact that $\ell^{-2}|f|^2 \geqslant 1$ on the set where $|f| \geqslant \ell$, prove Chebychev's inequality,

$$m(\{x \in \mathbb{R} \mid f(x) \geqslant \ell\}) \geqslant \ell^{-2} \int |f|^2 .$$

28.6 Let $1 \leqslant r \leqslant p$ and let $f \in \mathrm{L}^p(I)$ where I is a bounded interval.

(a) Prove that $|f|^r \leqslant 1 + |f|^p$ and deduce that $f \in \mathrm{L}^r$.

(b) Apply Hölder's inequality to the functions $|f|^r$ and 1, with p/r in place of p, to show that

$$\|f\|_r \leqslant \|f\|_p \ell(I)^{1/r-1/p}.$$

28.7 This exercise is concerned with a Fredholm integral operator on $\mathrm{C}[0,1]$. Let k be a real-valued continuous function on $[0,1]^2$ and let M be such that $|k(x,y)| \leqslant M$ for all $x, y \in [0,1]$. Define

$$(Kf)(x) := \int_0^1 k(x,y) f(y)\,\mathrm{d}y \quad (f \in \mathrm{C}[0,1]).$$

Use the Continuous DCT to prove that $Kf \in \mathrm{C}[0,1]$. (You may assume that, given $\varepsilon > 0$, there exists $\delta > 0$ such that $|k(s,y) - k(t,y)| < \varepsilon$ whenever $|s-t| < \delta$ and for all y [this comes from the fact that k is **uniformly** continuous on $[0,1]^2$—a 2-dimensional analogue of 9.5].) Show that K defines a linear map from $\mathrm{C}[0,1]$ to $\mathrm{C}[0,1]$ such that $\|Kf\|_\infty \leqslant M\|f\|_\infty$ for all $f \in \mathrm{C}[0,1]$.

28.8 Prove Minkowski's inequality for integrals, *viz.*, for $f \in \mathrm{L}^2(\mathbb{R}^2)$,

$$\sqrt{\int \left|\int f(x,y)\,\mathrm{d}x\right|^2 \mathrm{d}y} \leqslant \left(\int \left(\sqrt{\int |f(x,y)|^2\,\mathrm{d}x}\right) \mathrm{d}y\right)^2.$$

[Hint: start with a step function.]

28.9 This exercise is concerned with a Fredholm integral operator on $\mathrm{L}^2[0,1]$. Let k be a bounded measurable function on $[0,1]^2$ and let M be such that $|k(x,y)| \leqslant M$ for all $x, y \in [0,1]$. Define

$$(Kf)(x) := \int_0^1 k(x,y) f(y)\,\mathrm{d}y \quad (f \in \mathrm{L}^2[0,1]).$$

(a) Show that Kf is a well-defined and measurable function for each $f \in \mathrm{L}^2[0,1]$ (cf. Exercise 26.6).

(b) Use Minkowski's inequality for integrals to prove that K defines a linear map from $\mathrm{L}^2[0,1]$ to $\mathrm{L}^2[0,1]$ such that $\|Kf\|_2 \leqslant M\|f\|_2$ for all $f \in \mathrm{L}^2[0,1]$.

(c) Now assume that $k(x,y) = \overline{k(y,x)}$ for all x, y. Use FTT to prove that

$$\langle Kf, g\rangle = \langle f, Kg\rangle \quad (f, g \in \mathrm{L}^2[0,1]),$$

where $\langle\,.\,,\,.\,\rangle$ is the inner product defined in 28.4.

29 Fourier series: pointwise convergence

Fourier series have widespread applications in science and engineering, by providing a mathematical model for wave-forms. Suppose, for example, we wish to analyse the sound from an organ. We might do this by taking snapshots of the sound over time intervals of length L (a fraction of a second) and, by decomposing it into its constituent frequencies, discover which notes were being played. Mathematically, each sampling corresponds to expressing a function as the superposition of the wave-forms $\cos(2\pi kx/L)$, $\sin(2\pi kx/L)$. By rescaling so that $L = 2\pi$, the mathematical problem becomes that of decomposing functions on, say, $[-\pi, \pi]$, as infinite linear combinations of the functions $\cos kx$ and $\sin kx$ $(k = 0, 1, 2, \ldots)$.

In this chapter and the next we give a glimpse into the rich mathematical theory of Fourier series, exploiting the integration theory of the earlier chapters. Our account is self-contained but assumes that most readers will have met Fourier series before, at the level of rigour of, for example, [**12**]. We urge that it be read alongside a specialized text, for example [**11**], which gives a lively presentation of applications, from the temperature of subsoil to tides.

Fourier series representations are usually introduced to students of applied mathematics as a technique for solving certain ordinary and partial differential equations. In elementary courses the emphasis is on getting answers. Although formulae are presented with such caveats as 'for suitably well-behaved functions f', these qualifications do not greatly concern the average student, and are all too often forgotten or ignored. It is a truism that solutions to differential equations are differentiable. Thus it is good news for applied mathematicians that Fourier series behave well, pointwise, for functions which have a modicum of smoothness (existence of a derivative, piecewise, will certainly suffice (see 29.11)). In sharp contrast, Fourier series for functions which are merely continuous or merely integrable can behave in bizarre ways. We mention the extent of these pathologies in 29.18 and 30.16, but only to set in context the positive results on which we concentrate.

29.1 Fourier series introduced. Take an integrable function f on an interval of length 2π, for which we usually choose $[-\pi, \pi)$. The form of Fourier series most frequently encountered in elementary courses is the trigonometric series

$$(\dagger) \qquad \frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

$$\text{where } a_n := \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt \, \mathrm{d}t \quad \text{and} \quad b_n := \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt \, \mathrm{d}t.$$

The asssumption that $f \in \mathrm{L}[-\pi, \pi]$ guarantees that the integrals defining the Fourier coefficients a_n and b_n exist, because $\cos nt$ and $\sin nt$ are bounded and measurable (see 23.2. To motivate the choice of coefficients let us write

$$h(x) := \frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

and make the (optimistic!) assumption that the series converges and that summation and integration can be switched. Then, for $m \geqslant 0$,

$$\begin{aligned}\int_{-\pi}^{\pi} h(x) \cos mx \,\mathrm{d}x &= \int_{-\pi}^{\pi} \Big(\frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)\Big) \cos mx \,\mathrm{d}x \\ &= \frac{1}{2}a_0 \int_{-\pi}^{\pi} \cos mx \,\mathrm{d}x + \sum_{n=1}^{\infty} (a_n \int_{-\pi}^{\pi} \cos nx \cos mx \,\mathrm{d}x \\ &\qquad + b_n \int_{-\pi}^{\pi} \sin nx \cos mx \,\mathrm{d}x) \\ &= \pi a_m,\end{aligned}$$

and likewise, for $m \geqslant 1$,

$$\int_{-\pi}^{\pi} h(x) \sin mx \,\mathrm{d}x = \pi b_m,$$

thanks to the formulae given in Exercise 6.2. We stress that there is no reason to assume in general that $f(x) = h(x)$.

From the user's point of view the Fourier series (†) leaves something to be desired. It is annoying to have both sine and cosine functions in play together, to have the anomalous term $\frac{1}{2}a_0$, and to have to keep track of factors of π. (To circumvent the last of these irritations some accounts of Fourier theory rescale Lebesgue measure by a factor of $1/(2\pi)$ so that $\langle -\pi, \pi \rangle$ has measure 1.) Aesthetically, it is more pleasing to work with the complex exponential Fourier series of f, *viz.*

$$(\ddagger) \qquad \sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{\mathrm{i}nx}, \quad \text{with } c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \mathrm{e}^{-\mathrm{i}nt} \,\mathrm{d}t.$$

The interchangeability of (†) and (‡) comes from the formula $\mathrm{e}^{-\mathrm{i}nx} = \cos nx + \mathrm{i} \sin nx$ ($n \in \mathbb{Z}$). In this chapter we work with (†), to keep in tune with applications-oriented treatments. The results of Chapter 30 are of more theoretical interest, and there we opt for (‡), which gives a slicker presentation than (†). We put both forms in their natural theoretical context in Chapter 31.

29.2 Periodic functions on $\mathbb{R}$. Because cos and sin are periodic of period 2π, the series (†) has the same property. So, on $\mathbb{R}$, Fourier series can only serve to represent functions which are 2π-periodic. We can think of $\mathbb{R}$ as the union of the intervals $[(2k-1)\pi, (2k+1)\pi)$ ($k \in \mathbb{Z}$); these intervals slot together without

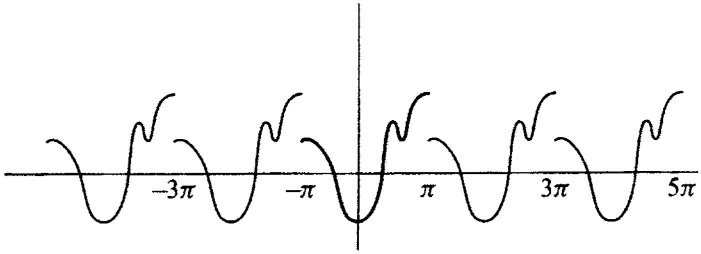


Figure 29.1

overlap. Any function g defined on $[-\pi, \pi)$ extends in a unique way to a well-defined function $g^{\circlearrowright}$ on $\mathbb{R}$ by $g^{\circlearrowright}(x + 2n\pi) := g(x)$ $(x \in [-\pi, \pi),\ n \in \mathbb{Z})$, and $g^{\circlearrowright}$ is 2π-periodic. In the other direction, any periodic function on $\mathbb{R}$ of period 2π is the periodic extension of its restriction to $[-\pi, \pi)$.

We cannot expect functions of the form $g^{\circlearrowright}$ to be integrable on $\mathbb{R}$. In fact $g^{\circlearrowright} \in \mathrm{L}$ only if $g \equiv 0$ (Exercise 13.5). We define

$$\mathrm{L}^{\circlearrowright}[-\pi, \pi] := \{f : \mathbb{R} \to \mathbb{C} \mid f = g^{\circlearrowright} \text{ for some } g \in \mathrm{L}[-\pi, \pi)\}.$$

Suppose that g is continuous on $[-\pi, \pi)$. Sadly, $f := g^{\circlearrowright}$ is not in general continuous on $\mathbb{R}$: see Fig. 29.1 for an example. However the left- and right-hand limits $f(x-) := \lim_{y \to x-} f(y)$ and $f(x+) := \lim_{y \to x+} f(y)$ do exist for every $x \in \mathbb{R}$, and a jump discontinuity occurs at $(2k-1)\pi$ precisely when $f(\pi-) \neq f(-\pi+)$. Because $[-\pi, \pi)$ is not a compact interval, $\mathrm{L}[-\pi, \pi]$ does not contain all continuous functions on $[-\pi, \pi)$ (consider $1/(\pi - x)$, for example). However if $g^{\circlearrowright}$ is continuous on $\mathbb{R}$ then $g^{\circlearrowright} \in \mathrm{L}^{\circlearrowright}[-\pi, \pi]$.

The next example introduces a set of functions we shall use subsequently for illustrations and exercises.

29.3 Exercise example. Sketch $f := g^{\circlearrowright}$ for the following functions g on $[-\pi, \pi)$.

(i) $g(x) = \begin{cases} 0 & (-\pi \leqslant x < 0), \\ x & (0 \leqslant x < \pi), \end{cases}$ (ii) $g = \chi_{(0,\pi)} - \chi_{(-\pi,0)}$,

(iii) $g(x) = x^2$, (iv) $g(x) = \mathrm{e}^x$,

(v) $g(x) = \pi - |x|$.

29.4 Exercise example (for those with little prior experience of Fourier series). Verify that the Fourier series for the functions in 29.3 are as follows.

(i) $\displaystyle \frac{\pi}{4} - \frac{2}{\pi} \sum_{m=1}^{\infty} \frac{\cos(2m-1)x}{(2m-1)^2} + \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\sin nx}{n}$,

(ii) $\displaystyle \frac{4}{\pi} \sum_{m=1}^{\infty} \frac{1}{2m-1} \sin(2m-1)x$,

(iii) $\dfrac{\pi^2}{3} + 4\displaystyle\sum_{n=1}^{\infty}(-1)^n \frac{\cos nx}{n^2}$,

(iv) $2(\mathrm{e}^{\pi} - \mathrm{e}^{-\pi})\Big(\dfrac{1}{2\pi} + \displaystyle\sum_{n=1}^{\infty}(-1)^n \frac{\cos nx}{\pi(n^2+1)} + \sum_{n=1}^{\infty}(-1)^{n-1}\frac{n \sin nx}{\pi(n^2+1)}\Big)$,

(v) $\dfrac{\pi}{2} + \dfrac{4}{\pi}\displaystyle\sum_{m=1}^{\infty} \frac{\cos(2m-1)x}{(2m-1)^2}$.

[We stress, again, that at this stage we are making no claim whatsoever that a function and its Fourier series are equal everywhere, or even anywhere. Indeed, it is not obvious in every case here that the Fourier series is even convergent. For information on these points see 29.12.]

29.5 Even and odd functions. Recall that $f\colon \mathbb{R} \to \mathbb{R}$ is even (odd) if $f(x) = f(-x)$ ($f(x) = -f(-x)$) for all $x \in \mathbb{R}$. Each function $\cos nx$ is even and each $\sin nx$ is odd. Therefore a Fourier series with $a_n = 0$ for all n represents an odd function, while one with $b_n = 0$ for all n represents an even function. For examples see 29.4. We can extend any integrable function $g\colon [0,\pi] \to \mathbb{R}$ to an even 2π-periodic function f by defining $g(x) := g(-x)$ for $x \in [-\pi, 0)$ and taking $f := g^{\circ}$ in the usual way. Then $f \in \mathrm{L}^{\circ}[-\pi,\pi]$. The Fourier coefficients take the form

$$a_n = \frac{2}{\pi}\int_0^{\pi} f(x)\cos nx \,\mathrm{d}x \quad (n = 0,1,2,\dots) \quad \text{and} \quad b_n = 0 \quad (n = 1,2,\dots).$$

The resulting series is called a (*Fourier*) *cosine series*. Similarly we may obtain a *sine series* for an integrable function defined on $[0,\pi)$ and extended to be odd and 2π-periodic. The coefficients are

$$a_n = 0 \quad (n = 0,1,2,\dots) \quad \text{and} \quad b_n = \frac{2}{\pi}\int_0^{\pi} f(x)\sin nx \,\mathrm{d}x \quad (n = 1,2,\dots).$$

29.6 The partial sums of a Fourier series. Let $f \in \mathrm{L}^{\circ}[-\pi,\pi]$. Consider the partial sums of the Fourier series (†) for f:

$$s_n(x) := \frac{1}{2}a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx).$$

Certainly $s_n \in \mathrm{L}^{\circ}[-\pi,\pi]$. Substituting for a_k and b_k and using property (L),

$$\begin{aligned} s_n(x) &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\Big(\frac{1}{2} + \sum_{k=1}^{n} (\cos kt \cos kx + \sin kt \sin kx)\Big)\,\mathrm{d}t \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\Big(\frac{1}{2} + \sum_{k=1}^{n} \cos k(t-x)\Big)\,\mathrm{d}t. \end{aligned}$$

We want to roll up this sum into a more compact form. Define the *Dirichlet kernel*

$$D_n(x) := \frac{1}{2} + \sum_{k=1}^{n} \cos kx = \begin{cases} \dfrac{\sin(n+\frac{1}{2})x}{2\sin\frac{1}{2}x} & \text{if } x \neq 2m\pi,\ m \in \mathbb{Z}, \\ n + \frac{1}{2} & \text{otherwise.} \end{cases}$$

(To add up the sum, express $\sin\frac{1}{2}x\cos kx$ as $2(\sin(k+\frac{1}{2})x - \sin(k-\frac{1}{2})x)$; for an alternative strategy see Exercise 29.1.) Note that

(i) D_n is continuous and 2π-periodic on $\mathbb{R}$,
(ii) $D_n(x) = D_n(-x)$ for all $x \in \mathbb{R}$,
(iii) $\int_{-\pi}^{\pi} D_n(x)\,\mathrm{d}x = \pi$.

To obtain these, use the first expression for $D_n(x)$. Now

$$\begin{aligned} \pi s_n(x) &= \int_{-\pi}^{\pi} f(t)D_n(t-x)\,\mathrm{d}t \\ &= \int_{-\pi-x}^{\pi-x} f(x+u)D_n(u)\,\mathrm{d}u && \text{(by (T))} \\ &= \int_{-\pi}^{\pi} f(x+u)D_n(u)\,\mathrm{d}u && \text{(by periodicity)} \\ &= \int_{0}^{\pi} f(x+u)D_n(u)\,\mathrm{d}u + \int_{-\pi}^{0} f(x+u)D_n(u)\,\mathrm{d}u \\ &= \int_{0}^{\pi} (f(x+u)+f(x-u))D_n(u)\,\mathrm{d}u && \text{(using (ii)).} \end{aligned}$$

Using (iii) above we obtain finally, for any choice of $s(x)$,

$$s_n(x) - s(x) = I_n(x) := \frac{1}{\pi}\int_0^{\pi} (f(x+u)+f(x-u)-2s(x))\frac{\sin(n+\frac{1}{2})u}{2\sin\frac{1}{2}u}\,\mathrm{d}u.$$

We are now ready to seek conditions under which $s_n(x) \to s(x)$ as $n \to \infty$, initially with no presumption on what value $s(x)$ should take. The convergence theorems are of no assistance here. The kernel function D_n is oscillatory, and the pointwise limit of the integrand of $I_n(x)$ generally fails to exist.

Consider the graph of D_n on $[-\pi, \pi]$. Figure 29.2 shows D_6 and D_{18}. The function D_n has a peak at 0, of height $n + \frac{1}{2}$, but for large n the oscillations rapidly die down away from 0. This suggests that in $I_n(x)$ all the key action occurs near $u = 0$. Note further that, near 0 the denominator $\sin\frac{1}{2}u$ is close to $u/2$, the latter being easier to work with. The Localization Theorem below makes all this precise. Its proof relies on the Riemann–Lebesgue Lemma, given in 23.4. In view of its importance here we restate the lemma in a form tailored to its Fourier series applications.

29.7 Riemann–Lebesgue Lemma (restated). Let $f \in \mathrm{L}^{\circ}[-\pi,\pi]$, and let $\{\lambda_n\}$ be a real sequence such that $\lambda_n \to \infty$ as $n \to \infty$. Then $\int_{-\pi}^{\pi} f(t)\sin(\lambda_n t)\,\mathrm{d}t \to 0$.

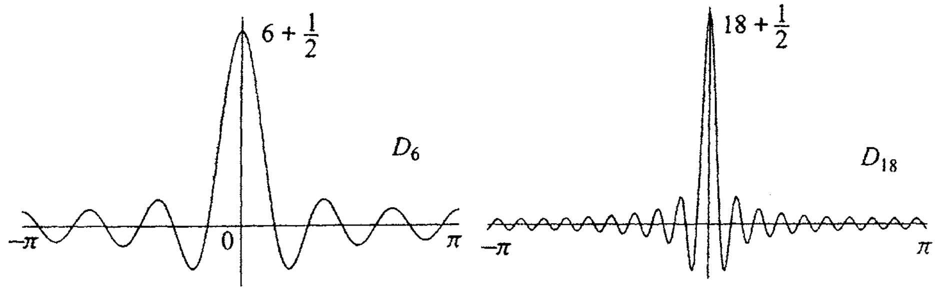


Figure 29.2

In particular, the Fourier coefficients, a_n and b_n, of f tend to 0 as $n \to \infty$.

29.8 Riemann's Localization Theorem. Let $f \in \mathrm{L}^{\mathrm{O}}[-\pi, \pi]$. Then, for any fixed δ with $0 < \delta < \pi$, $s_n(x) \to s(x)$ if and only if

$$\lim_{n\to\infty} \frac{1}{2\pi} \int_0^\delta \frac{(f(x+u) + f(x-u) - 2s(x))}{u} \sin(n + \tfrac{1}{2})u \, du = 0.$$

Proof. We express $s_n(x) - s(x)$ as in 29.6 and split the range of integration into $[0, \delta)$ and $[\delta, \pi]$. On the latter interval, $(f(x+u) + f(x-u) - 2s(x))/2\sin(u/2)$ is integrable because $1/\sin(u/2)$ is a bounded measurable function there, and f is integrable. Thus by the Riemann–Lebesgue Lemma,

$$\lim_{n\to\infty} (s_n(x) - s(x)) = \lim_{n\to\infty} \frac{1}{2\pi} \int_0^\delta \frac{f(x+u) + f(x-u) - 2s(x)}{\sin \frac{1}{2}u} \sin(n + \tfrac{1}{2})u \, du.$$

It now suffices to show that

$$\lim_{n\to\infty} \int_0^\delta (f(x+u) + f(x-u) - 2s(x)) \left(\frac{1}{\sin u/2} - \frac{1}{u/2} \right) \sin(n + \tfrac{1}{2})u \, du = 0.$$

On $(0, \delta)$, $\operatorname{cosec} u/2 - 2/u$ is continuous and bounded (boundedness by L'Hôpital's rule and 2.21; see Exercise 2.11). Hence the required conclusion follows by the Riemann–Lebesgue Lemma. □

29.9 Corollary: Dini's test. Let $f \in \mathrm{L}^{\mathrm{O}}[-\pi, \pi]$ and $x \in \mathbb{R}$. Then $s_n(x) \to s(x)$ if there exists $\delta > 0$ such that $(f(x+u) + f(x-u) - 2s(x))/u$ is integrable on $(0, \delta]$. In particular, $s_n(x) \to s(x)$ if there exist positive constants M and δ

(which may depend on x) such that $|f(x+u)+f(x-u)-2s(x)| \leqslant M|u|$ on $(0,\delta)$.

Proof. The first part is immediate from 29.8 and 29.7. For the final part use the Comparison Theorem, noting that $M\chi_{(0,\delta)}$ is integrable. □

The Localization Theorem is a startling result. The coefficients in the Fourier series depend on the values of f **throughout** $[-\pi,\pi)$, yet the convergence behaviour of the Fourier series at a given point x depends only on the values f takes **close to** x.

We should now like to identify classes of functions—the larger and more familiar the better—to which 29.8 applies. Unfortunately the 19th century mathematicians' quest for progressively weaker sufficient conditions for convergence of $\{s_n(x)\}$ at a particular point x did not lead to their grail—necessary and sufficient conditions for pointwise convergence. Indeed it turns out that this is unattainable; see 30.16. We shall head in the opposite direction and impose rather strong conditions on f which guarantee convergence of $\{s_n(x)\}$ for all $x \in \mathbb{R}$—giving in 29.11 a workable theorem for practical use. We begin with an easy deduction from 29.8.

29.10 Piecewise differentiable functions. This subsection deals with the kind of functions f that occur regularly in applications. As usual we need $f \in \mathrm{L}^{\circ}[-\pi,\pi]$.

Of course our primary candidate for $s(x)$ above is $f(x)$. We have

$$\frac{f(x+u)+f(x-u)-2f(x)}{u} = \frac{f(x+u)-f(x)}{u} - \frac{f(x-u)-f(x)}{-u}.$$

The quotients on the right-hand side tend respectively to the right-hand and left-hand derivatives of f at x if these exist, which is certainly the case if $f'(x)$ exists. Remember that existence of a limit implies boundedness locally; see 2.21. Thus if $f'(x)$ exists there are constants M and δ (which may depend on x) such that $|(f(x+u)+f(x-u)-2f(x))/u| \leqslant M$ for $u \in (0,\delta)$. We conclude, by the last part of 29.9, that the Fourier series converges to $f(x)$ whenever f is differentiable at x. Of course, f is automatically continuous at x in this case.

We quickly see that $s(x) = f(x)$ may not be the right choice in general. Observe that if we change the value of f at the single point x we do not change the Fourier coefficients, and hence the value of $\lim s_n(x)$ should not depend on the value that f actually takes at x. To make further progress we make the additional assumption that the right- and left-hand limits, $f(x+)$ and $f(x-)$, exist. Then $f(x+u)+f(x-u) \to f(x+)+f(x-)$ as $u \to 0+$. This suggests that we should try to take $s(x) := \frac{1}{2}(f(x+)+f(x-))$, the average of the right- and left-hand limits. If f is continuous at x this is just $f(x)$.

Even when we start with a differentiable function on $[-\pi,\pi)$, its periodic extension may have jump discontinuities, at which it is certainly not differentiable, or may have discontinuities in its derivative. A real-valued function g on $[-\pi,\pi)$

(and also its periodic extension $f = g^{\circlearrowleft}$) is said to be *piecewise differentiable* if there exist finitely many points $-\pi = y_0 < y_1 < \cdots < y_N = \pi$ such that

(i) the left-hand limit $g(y_i-)$ exists for $i = 1, \ldots, N$ and the right-hand limit $g(y_i+)$ exists for $i = 0, \ldots, N-1$, and

(ii) for $i = 0, \ldots, N-1$, the function g_i defined on $[y_i, y_{i+1}]$ by

$$g_i(x) := \begin{cases} g(x) & \text{if } y_i < x < y_{i+1}, \\ g(y_i+) & \text{if } x = y_i, \\ g(y_{i+1}-) & \text{if } x = y_{i+1} \end{cases}$$

has a derivative at each point of $[y_i, y_{i+1}]$, the derivatives at the endpoints being one-sided derivatives.

Here are some simple observations.

(a) For each i, g_i is continuous and hence integrable on $[y_i, y_{i+1}]$. Therefore $f \in \mathrm{L}^{\circlearrowleft}[-\pi, \pi]$ because, for each i, $f = g_i$ a.e. on $[y_i, y_{i+1}]$.

(b) The conditions for piecewise differentiability do not involve the values of f at the points y_i; these values may be assigned in any manner.

(c) Periodicity implies that the behaviour demanded of f on $[-\pi, \pi]$ is replicated on the translates $[(2k-1)\pi, (2k+1)\pi]$ $(k \in \mathbb{Z})$. On any bounded interval f is non-differentiable at only a finite number of points, at most.

Informally, a piecewise differentiable function is one whose graph is not too wild to be drawn. Any discontinuities are jump discontinuities (by (i)), and at any point at which the curve fails to have a tangent there is a jump discontinuity in the function or in its gradient or both. See Fig. 29.3 for a typical example.

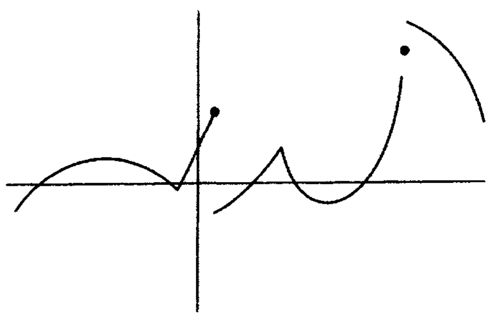


Figure 29.3

The result in 29.11 is a good example of a theorem in which the conditions are strong enough to yield the conclusion reasonably easily, yet weak enough that the theorem applies to most functions of practical importance.

29.11 Pointwise convergence in practice. Let $f\colon \mathbb{R} \to \mathbb{R}$ be 2π-periodic and piecewise differentiable. Then, at each $x \in \mathbb{R}$,

$$s_n(x) \to s(x) := \begin{cases} \frac{1}{2}(f(x+) + f(x-)) & \text{in general,} \\ f(x) & \text{if } f \text{ is continuous at } x. \end{cases}$$

Proof. We invoke the final part of 29.9. We have already indicated that $f \in \mathrm{L}^{\circ}[-\pi,\pi]$. Suppose $x \in [-\pi,\pi)$. Then $x \in [y_i, y_{i+1})$ for some i, and $x+u \in [y_i, y_{i+1})$ too, for $u > 0$, u sufficiently small. Then

$$\frac{f(x+u) - f(x+)}{u} = \frac{g_i(x+u) - g_i(x)}{u} \to g_i'(x) \quad \text{as } u \to 0+,$$

whence there exist constants M and δ such that $|f(x+u) - f(x+)| \leqslant M\,|u|$ for $0 < u < \delta$ (M and δ may depend on x). The difference $f(x-u) - f(x-)$ can be handled in a similar manner. Hence 29.9 can be applied, with $s(x) = \frac{1}{2}(f(x+) + f(x-))$. $\square$

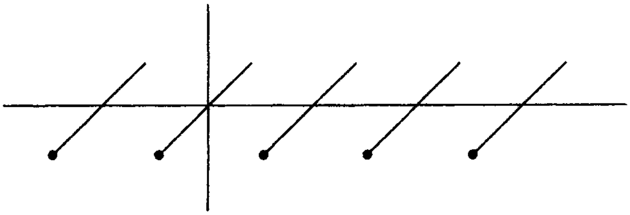


Figure 29.4

29.12 Example. Consider the function shown in Fig. 29.4: the periodic extension of $g(x) = x$ on $[-\pi,\pi)$. Certainly this is 2π-periodic and piecewise differentiable, so 29.11 applies. It tells us that the Fourier series converges to $f(x)$ at continuity points, that is, whenever x is not an odd multiple of π; when $x = (2k-1)\pi$ the Fourier series converges to $\frac{1}{2}(f(-\pi+) + f(\pi-))$, namely 0. Routine calculation gives $a_n = 0$ and $b_n = 2(-1)^{n-1}/n$. Thus, on $[-\pi,\pi]$,

$$2\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \sin nx = \begin{cases} x & (-\pi < x < \pi), \\ 0 & (x = \pm\pi). \end{cases}$$

Putting $x = \pi/2$ we get as a by-product

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = \frac{\pi}{4}.$$

29.13 Remarks. You are strongly advised always to sketch $f = g^{\circ}$ when dealing with the Fourier series of a given function g on $[-\pi,\pi)$. It is f, rather than g, to which 29.11 applies. If you forget this you are liable to assign the wrong value to the Fourier series at $\pm\pi$. Look again at the preceding example. The graph of the periodic extension clearly has jump discontinuities at the points $(2k-1)\pi$ $(k \in \mathbb{Z})$. At these points the Fourier series we have computed has value 0, since each term does. This is in line with the theory. This says that the Fourier series does the fair-minded thing at the discontinuities: it converges to the average of the left- and right-hand limits, namely 0. Suppose we had considered just the given function $g(x) = x$ on $[-\pi, \pi)$. The statement 'g is differentiable on $[-\pi, \pi)$ and hence the Fourier series for g at x converges to $g(x)$ $(-\pi \leqslant x < \pi)$' is patently wrong.

29.14 Exercise example. For the functions g in 29.3 apply 29.11 to show that the sum of the Fourier series at each $x \in [-\pi, \pi]$ is as indicated:

$$\text{(i)} \begin{cases} 0 & (-\pi < x < 0), \\ x & (0 \leqslant x < \pi), \\ \dfrac{\pi}{2} & (x = \pm\pi), \end{cases} \qquad \text{(ii)} \begin{cases} 1 & (0 < x < \pi), \\ -1 & (-\pi < x < 0), \\ 0 & (x = 0, \pm\pi), \end{cases}$$

$$\text{(iii)}\ x^2, \qquad \text{(iv)} \begin{cases} e^x & (-\pi < x < \pi), \\ \cosh\pi & (x = \pm\pi), \end{cases}$$

$$\text{(v)}\ \pi - |x|.$$

Specify for which points x this sum is $f(x)$ (that is, $g^{\circ}(x)$).

29.15 Exercise example. Carry through an analysis of the Fourier series of

$$\text{(i)}\ \chi_{[-\pi/2, \pi/2]}, \qquad \text{(ii)} \begin{cases} x^2 & (x \geqslant 0), \\ -x^2 & (x < 0), \end{cases} \qquad \text{(iii)}\ x\chi_{(0,\pi)}(x), \qquad \text{(iv)}\ x\cos x,$$

on $[-\pi, \pi)$, parallel to that in 29.3, 29.4, and 29.12.

29.16 Using Fourier series to evaluate sums. Example 29.12 shows that a by-product of applying 29.11 to a given function may be the value of one or more infinite sums, obtained by taking specific values of x. The severe limitation of this technique is that it may be hard, or impossible, to discover a function whose Fourier series yields some specified series. Compare 29.12 and Exercise 30.9.

The repertoire of series summable by Fourier techniques without over-laborious calculations is increased by Parseval's Theorem; see 30.13, 30.14.

29.17 Exercise example. Use 29.4 and 29.14 to prove

$$\text{(i)}\ \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}, \qquad \text{(ii)}\ \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n^2} = \frac{\pi^2}{12},$$

$$\text{(iii)}\ \sum_{m=1}^{\infty} (-1)^{m-1} \frac{1}{2m-1} = \frac{\pi}{4}, \qquad \text{(iv)}\ \sum_{n=1}^{\infty} \frac{1}{n^2+1} = \frac{\pi\coth\pi}{2} - \frac{1}{2}.$$

29.18 Pointwise convergence assessed. It is worth noting that the conditions on f in 29.9 and 29.11 imply the convergence of f's Fourier series at a given point. This may not be easy to prove directly by elementary convergence tests for series.

In 29.11 we gave a convenient criterion for the convergence of $s_n(x)$ to $\frac{1}{2}(f(x+) + f(x-))$, involving a fairly weak smoothness restriction on f. However 29.11 says nothing about the **speed** of convergence. In fact (see Exercise 29.4), there is a nice interplay between the smoothness of f and good behaviour of its Fourier series: the more derivatives f possesses, the faster its Fourier coefficients will tend to 0. The rate of convergence of $s_n(x)$ generally varies with x. In particular, in a neighbourhood of a jump discontinuity containing no other discontinuities the partial sums exhibit *Gibbs's phenomenon.* They always overshoot the limiting value, as illustrated in Fig. 29.5. Surprisingly, this overshoot of $s_n(x)$ does not tend to zero as $n \to \infty$; it can be estimated to be about 9% in the limit. Exercise 29.10 analyses the phenomenon for the function $\chi_{(0,\pi)} - \chi_{(-\pi,0)}$.

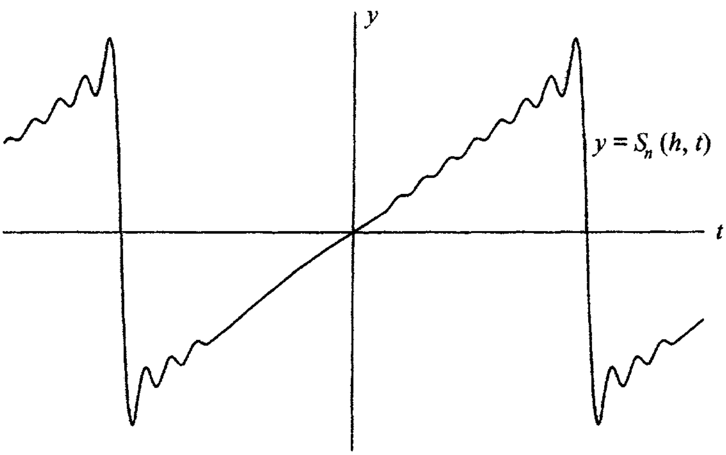


Figure 29.5

Not all functions in $\mathrm{L}^{\mathrm{O}}[-\pi, \pi]$ are piecewise differentiable. Indeed, differentiability is the exception rather than the rule. It can be shown, in a sense that can be made precise, that almost all continuous functions on a compact interval are not differentiable at any point (see [15] for example). Also, an integrable function g on $[-\pi, \pi)$ may have infinitely many discontinuities; no such function is piecewise differentiable.

Now assume merely that $f \in \mathrm{L}^{\mathrm{O}}[-\pi, \pi]$. Of course if the partial sums $s_n(x)$ are to converge to $\frac{1}{2}(f(x+) + f(x-))$ then the limits $f(x+)$ and $f(x-)$ must exist. This is true in particular when f is a monotonic function. More generally it is true if f is a function of bounded variation, that is, the difference of two increasing functions; see 24.8. It can easily be seen that every piecewise differentiable function is of bounded variation. **Jordan's test** asserts that $s_n(x) \to \frac{1}{2}(f(x+) + f(x-))$ whenever $f \in \mathrm{L}^{\circ}[-\pi, \pi]$ and f is of bounded variation. The proof can be found in [**1**], for example.

We conclude this chapter with some bad news about pointwise convergence:

- there exists a continuous function $f \in \mathrm{L}^{\circ}[-\pi, \pi]$ such that $\{s_n(x)\}$ diverges on an uncountable and dense subset of $\mathbb{R}$;
- there exists $f \in \mathrm{L}^{\circ}[-\pi, \pi]$ such that $\{s_n(x)\}$ diverges everywhere.

Not surprisingly these disquieting but intriguing facts are very hard to establish. We say a little more about them in Chapter 30. Meanwhile you are encouraged to tackle the exercises for this chapter alert to the fact that some Fourier series behave extremely badly. Don't forget to appeal explicitly to 29.11 or to some other suitable criterion before equating a given function with its Fourier series.

In the next chapter we desert pointwise convergence in favour of other modes of convergence, such as convergence in L^2 norm. It turns out very satisfying and theoretically far-reaching results can then be proved.

Exercises

29.1 By writing $\cos kx$ as the real part of e^{ikx} and $\sin kx$ as the imaginary part of e^{ikx} prove that

$$\text{(i)} \sum_{k=1}^{n} \sin kx = \sin \frac{1}{2}x \left(\frac{\sin \frac{1}{2}(n+1)x}{\sin \frac{1}{2}x} \right), \quad \text{(ii)} \sum_{k=0}^{n-1} \sin(k+\tfrac{1}{2})x = \frac{\sin^2 \frac{1}{2}nx}{\sin \frac{1}{2}x}.$$

((i) gives the formula for $D_n(x)$ used in 29.6; (ii) is used in Chapter 30 to handle the Fejér kernel $F_n := (D_0 + \cdots + D_{n-1})/n$.)

29.2 Let $0 < \alpha < \pi$ and define g on $[-\pi, \pi]$ by

$$g(x) = \begin{cases} \frac{1}{2}(\pi - \alpha)x & (0 \leqslant x \leqslant \alpha), \\ \frac{1}{2}(\pi - x)\alpha & (\alpha < x \leqslant \pi), \\ -g(-x) & (-\pi \leqslant x < 0). \end{cases}$$

Sketch the graph of g and of its periodic extension $f := g^{\circ}$. Prove that

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{n^2} \sin n\alpha \sin nx \quad (x \in \mathbb{R}).$$

29.3 (a) Prove that $-\log|2\sin\frac{1}{2}x|$ is integrable on $[-\pi, \pi]$.
(b) Prove that $\sum_{n=1}^{\infty} (\cos nx)/n$ is the Fourier series of $-\log|2\sin\frac{1}{2}x|$.

(c) Discuss the convergence of this Fourier series.

29.4 Let $f \in L^{\bigcirc}[-\pi,\pi]$ be an even function. Let f have a continuous derivative. Prove that $\lim_{n\to\infty} na_n = 0$.

29.5 Let

$$f(x) = \sum_{k=-\infty}^{\infty} e^{-|x+(2k-1)\pi|}.$$

(a) By first evaluating f on $[-\pi,\pi)$ show that the defining series is convergent, that f is 2π-periodic, and that f is piecewise differentiable.

(b) Let $\{c_n\}$ be the coeffcients in the complex exponential Fourier series of f. Prove that

$$c_n = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-|x|-inx}\,dx,$$

and hence evaluate c_n.

(c) Use the (analogue for (‡) of) 29.11 to deduce that

$$\frac{e^{2\pi}+1}{e^{2\pi}-1} = \frac{1}{\pi}\sum_{n=-\infty}^{\infty}(n^2+1)^{-1}.$$

29.6 Let $f\colon [0,\pi] \to \mathbb{R}$. Assume that $f(y+)$ exists for all $y \in [0,\pi)$, $f(y-)$ exists for all $y \in (0,\pi]$, and f has at most a finite number of discontinuities. Define

$$F(x) := \int_0^x f(y)\,dy \quad (x \in [0,\pi]).$$

(a) Appeal to the Indefinite Integral Theorem (5.4) to show that F is piecewise differentiable.

(b) Use FTT to prove that $\beta_n = a_n/n$ for $n = 1, 2, \ldots$, where

$$a_n := \frac{2}{\pi}\int_0^{\pi} f(x)\cos nx\,dx \quad \text{and} \quad \beta_n := \frac{2}{\pi}\int_0^{\pi} F(x)\sin nx\,dx.$$

(c) Use 29.11 to show that $F(x) - \frac{1}{2}a_0x$ is equal to its Fourier series at each $x \in [0,\pi]$ and deduce that

$$F(x) = \frac{1}{2}a_0x + \sum_{n=1}^{\infty}\frac{1}{n}a_n\sin nx \quad (x \in [0,\pi]).$$

29.7 By taking $f = \chi_{[0,\pi/2]} - \chi_{(\pi/2,\pi]}$ in the preceding exercise evaluate

$$1 - \frac{1}{3^2} - \frac{1}{5^2} + \frac{1}{7^2} + \frac{1}{9^2} - \frac{1}{11^2} - \frac{1}{13^2} + \cdots.$$

29.8 Let $t \in \mathbb{R}$ and let $f(x) = \cos xt$ $(x \in [-\pi,\pi))$.

(a) Calculate the Fourier coefficients of f.

(b) Prove that

$$\cos xt = \frac{\sin \pi t}{\pi t} + \frac{2t \sin \pi t}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n \cos nx}{t^2 - n^2} \qquad (x, t \in \mathbb{R}).$$

(c) Deduce that

$$\pi \cot \pi t - \frac{1}{t} = \sum_{n=1}^{\infty} \frac{2t}{t^2 - n^2} \qquad (t \in \mathbb{R}).$$

(d) Show that term-by-term integration with respect to t over $[0, y]$ $(0 < y < 1)$ of the series in (c) is legitimate.

(e) Deduce that

$$\frac{\sin \pi y}{\pi y} = \lim_{N \to \infty} \prod_{n=1}^{N} \left(1 - \frac{y^2}{n^2}\right).$$

29.9 Define inductively the sequence $\{B_r(x)\}_{r \geqslant 0}$ on $[0, 2\pi)$ by $B_0(x) = 1$ and

$$B_r'(x) = B_{r-1}(x) \quad \text{and} \quad \int_0^{2\pi} B_r(x)\,dx = 0 \qquad (r \geqslant 1).$$

(a) Show that, for $r \geqslant 2$,

$$\int_0^{2\pi} B_{r-1}(x) \cos nx\,dx = n \int_0^{2\pi} B_r(x) \sin nx\,dx,$$
$$\int_0^{2\pi} B_{r-1}(x) \sin nx\,dx = -n \int_0^{2\pi} B_r(x) \cos nx\,dx.$$

(b) Compute the Fourier series of B_r for each r, and discuss to what limit it converges at each $x \in [0, 2\pi]$.

(c) Show that as $r \to \infty$, $(-1)^{r-1} B_{2r}(x) \to 2\cos x$ for all $x \in [0, 2\pi]$.

[On $[0, 2\pi)$, $B_r(x)$ is a polynomial, the rth *Bernouilli polynomial*. For these functions it is appropriate to work on $[0, 2\pi)$ rather than on $[-\pi, \pi)$; the theory of pointwise convergence adapts in the obvious way.]

29.10 [This exercise illustrates Gibbs's phenomenon] Let $f = \chi_{(0,\pi)} - \chi_{(-\pi,0)}$. Fix x with $|x| \leqslant \pi/2$.

(a) Show that

$$\pi s_n(x) = \int_{-x}^{x} D_n(u)\,du - \int_{-x+\frac{\pi}{2}}^{x+\frac{\pi}{2}} D_n(u)\,du.$$

(b) Integrate by parts the second integral in (a) to verify that

$$\pi s_n(x) - \int_{-x}^{x} D_n(u)\,du = \mathbf{O}(n^{-1}).$$

(c) Use integration by parts to show that

$$\int_{-x}^{x} D_n(u)\,\mathrm{d}u - \int_{-x}^{x} \frac{\sin(n+\frac{1}{2})u}{u}\,\mathrm{d}u$$

is $\mathbf{O}(n^{-1})$. [Hint: use Exercise 2.11.]

(d) Put together (a)–(c) to show that

$$J_n(x) := \frac{2}{\pi}\int_0^{(n+1/2)x} \frac{\sin y}{y}\,\mathrm{d}y$$

approximates $s_n(x)$ with error at most Kn^{-1} (K a constant).

(e) Show that $J_n(x)$ achieves its maximum at $x(n+\frac{1}{2}) = \pi$, and that the maximum value is $2/\pi\int_0^{\pi} y^{-1}\sin y\,\mathrm{d}y$. [This integral can be shown to be approximately 1.0895. Thus near zero the partial sum $s_n(x)$ overshoots the true value by about 9% in the limit as $n \to \infty$.]

30 Fourier series: convergence reassessed

In Chapter 29 we studied pointwise convergence of Fourier series. When solving a differential equation by Fourier methods we want the solution to be a superposition of trigonometric functions at all points of a given range. Thus, tacitly, pointwise convergence is the natural choice in this situation. Sadly, for many functions f in $\mathrm{L}^{\circ}[-\pi,\pi]$ the partial sum sequence $\{s_n\}$ of the Fourier series of f does not behave well at individual points. We now try to get better convergence by averaging the partial sums to iron out the effect of the oscillations of the trigonometric functions from which s_n is constructed. Accordingly we replace the nth partial sum, s_n, by the arithmetic mean, σ_n, of the first n partial sums, or look at norm convergence in L^1 or L^2. We show that

- $\{\sigma_n\}$ converges for any $f \in \mathrm{L}[-\pi,\pi]$, and converges uniformly to f if f is continuous and satisfies $f(-\pi) = f(\pi)$ (Fejér's Theorem, 30.4);
- if $f \in \mathrm{L}^2[-\pi,\pi]$ then $\|s_n - f\|_2 \to 0$ (30.11).

30.1 Continuous periodic functions. We shall work principally on $[-\pi,\pi]$. However there is one respect in which it is still helpful to consider the extension of f from $[-\pi,\pi)$ to a 2π-periodic function on $\mathbb{R}$. Within $\mathrm{L}^{\circ}[-\pi,\pi]$ lies

$$\mathrm{C}^{\circ}[-\pi,\pi] := \{\, g\colon \mathbb{R} \to \mathbb{C} \mid g \text{ is continuous, has period } 2\pi, \text{ and } g(-\pi) = g(\pi)\,\}.$$

The continuous functions on $\mathbb{R}$ of period 2π are exactly the (well-defined) periodic extensions of functions f which are continuous on $[-\pi,\pi]$ and such that $f(-\pi) = f(\pi)$. We remark that $t \mapsto \mathrm{e}^{\mathrm{i}t}$ maps $\mathbb{R}$ (or just $[-\pi,\pi)$) onto the unit circle in the complex plane, $T := \{z \in \mathbb{C} \mid |z| = 1\}$, and that there is a one-to-one correspondence between $\mathrm{C}^{\circ}[-\pi,\pi]$ and $\mathrm{C}(T)$, given by $g \leftrightarrow g^*$ where $g(t) = g^*(\mathrm{e}^{\mathrm{i}t})$ for all t. From a theoretical point of view, it is more natural to work with $\mathrm{C}(T)$ than with $\mathrm{C}^{\circ}[-\pi,\pi]$. However, to maintain the links with the classical Fourier series of functions on $\mathbb{R}$ we opt for $\mathrm{C}^{\circ}[-\pi,\pi]$. We equip $\mathrm{C}^{\circ}[-\pi,\pi]$ with the norm

$$\|f\|_\infty := \sup\{|f(x)| \mid x \in [-\pi,\pi]\},$$

and recall that $\|f_n - f\|_\infty \to 0$ is equivalent to $f_n \xrightarrow{\mathrm{u}} f$ on $[-\pi,\pi]$ (8.5).

We need hte following density result in 30.11. Exercise 30.2 seeks the details of the proof.

30.2 Lemma. $C^{\circ}[-\pi,\pi]$ is a dense subset of $(L^2[-\pi,\pi], \|\cdot\|_2)$.

Sketch proof. Certainly $C^{\circ}[-\pi,\pi] \subseteq L^2[-\pi,\pi]$. Let $\varepsilon > 0$ and let $f \in C^{\circ}[-\pi,\pi]$. By considering real and imaginary parts we may assune f is real-valued. By 28.12 there exists $\varphi \in L^{\text{step}}$ with $\|f-\varphi\|_2 < \varepsilon$. Now choose $\psi \in L^{\text{step}}$ such that $\psi = 0$ outside some interval $[-\pi+\delta, \pi-\delta]$ $(\delta > 0)$ and $\|\varphi - \psi\|_2 < \varepsilon$. Finally, approximate ψ by $g \in C^{\circ}[-\pi,\pi]$ so that $\|\psi - g\|_2 < \varepsilon$ (take g to be piecewise linear, with $g(-\pi) = g(\pi)$, or use 11.13). Then $\|f-g\|_2 < 3\varepsilon$ by the triangle inequality. □

30.3 Cesàro means and the Fejér kernel. Given a complex sequence $\{\alpha_n\}$, define

$$\beta_n := \frac{1}{n}(\alpha_1 + \cdots + \alpha_n);$$

$\{\beta_n\}$ is the sequence of *Cesàro means* of $\{\alpha_n\}$. Consider $\alpha_n := (-1)^n$. Of course $\{\alpha_n\}$ is not convergent. However $\beta_{2n} = 0$ and $\beta_{2n-1} = -1/(2n-1)$, so $\{\beta_n\}$ converges to 0. So averaging may convert a non-convergent sequence into a convergent one. On the other hand, if $\alpha_n \to \alpha$ then it is easy to prove that $\beta_n \to \alpha$ too. When α_n is the nth partial sum of a series $\sum a_k$ and $\{\beta_n\}$ converges, then $\sum a_k$ is said to be *Cesàro summable* or *$(C,1)$-summable*. For a little more on this topic, which was exhaustively explored in the 19th century, see Exercise 30.1.

We now apply these ideas to Fourier series. Throughout the chapter we shall almost always express the Fourier series of $f \in L[-\pi,\pi]$ in the complex exponential form

$$(\ddagger) \qquad \sum_{n=-\infty}^{\infty} c_n e^{inx}, \quad \text{with } c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt.$$

For $(\ddagger)$, the Dirichlet kernel is

$$D_n(x) := \sum_{k=-n}^{n} e^{ikx} = \begin{cases} \dfrac{\sin(n+\frac{1}{2})x}{\sin\frac{1}{2}x} & (x \neq 2m\pi \quad (m \in \mathbb{Z})), \\ n+\frac{1}{2} & (x = 2m\pi \quad (m \in \mathbb{Z})). \end{cases}$$

For $n = 1, 2, \ldots$ we define

$$s_n := \sum_{k=-n}^{n} c_n e^{ikx} \quad \text{and} \quad \sigma_n := \frac{s_0 + \cdots + s_{n-1}}{n}.$$

By the formula for s_n in 29.6 and property (L) we have immediately

$$\sigma_n = \frac{1}{2\pi}\int_0^{\pi} (f(x+u) + f(x-u))F_n(u)\,du,$$

where $F_n := \dfrac{1}{n}(D_0 + \cdots + D_{n-1})$ is the *Fejér kernel*. We have, for $x \neq 2m\pi$,

$$F_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} \frac{\sin(k+\frac{1}{2})x}{\sin\frac{1}{2}x} = \frac{1}{n}\Big(\frac{\sin^2\frac{1}{2}nx}{\sin^2\frac{1}{2}x}\Big).$$

To carry out this summation it is easiest to regard $\sin(k+\frac{1}{2})x$ as the imaginary part of $e^{i(k+\frac{1}{2})x}$; see Exercise 29.1. The graph of F_n is as indicated in Fig. 30.1. We note the following properties:

(F1) F_n is continuous and 2π-periodic,
(F2) $F_n(u) = F_n(-u)$ for all u,
(F3) $\int_0^\pi F_n(u)\,du = 2\pi$,
(F4) $F_n(u) \geqslant 0$ for all u,
(F5) $F_n(u) \leqslant \dfrac{1}{2n\sin^2\delta/2}$ for $0 < \delta \leqslant u < \pi$.

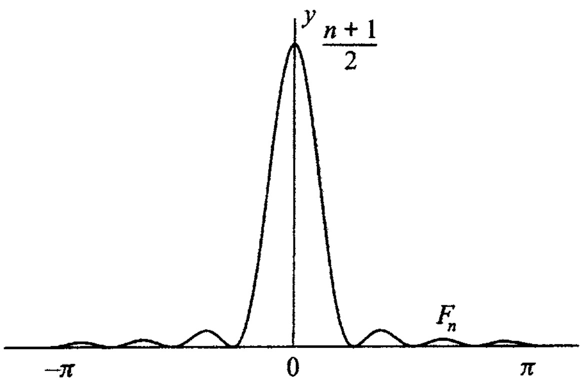


Figure 30.1

By moving from $\{s_n\}$ to $\{\sigma_n\}$ we can get stronger conclusions than in Chapter 29 under weaker conditions. In particular, continuity of f, which was of limited benefit to us before, is now sufficient to give highly satisfactory results.

30.4 Fejér's Theorem.

(a) Let $f \in L^{\circ}[-\pi,\pi]$ and assume that x is such that $f(x+)$ and $f(x-)$ exist. Then $\sigma_n(x) \to s(x) := \frac{1}{2}(f(x+) + f(x-))$.

(b) Assume that $f \in C^{\circ}[-\pi,\pi]$. Then $\sigma_n \xrightarrow{u} f$ on $[-\pi,\pi]$.

Further, if $\{s_n(x)\}$ converges, then in (a) $s_n(x) \to s(x)$ and in (b) $s_n(x) \to f(x)$.

Proof. We can write

$$\begin{aligned}
&2\pi|\sigma_n(x) - s(x)| \\
&= \left|\int_0^\pi (f(x+u) + f(x-u) - 2s(x))F_n(u)\,\mathrm{d}u\right| && \text{(using (F4))} \\
&\leqslant \int_0^\delta |(f(x+u) + f(x-u) - 2s(x))|\,F_n(u)\,\mathrm{d}u \\
&\quad + \int_\delta^\pi |(f(x+u) + f(x-u) - 2s(x))|\,F_n(u)\,\mathrm{d}u && \text{(by (F4))} \\
&\leqslant \sup_{0<u<\delta} |(f(x+u) + f(x-u) - 2s(x))| \int_0^\delta F_n(u)\,\mathrm{d}u \\
&\quad + \frac{1}{2n\sin^2\delta/2}\int_\delta^\pi |(f(x+u) + f(x-u) - 2s(x))|\,\mathrm{d}u && \text{(by (F5))} \\
&\leqslant \pi \sup_{0<u<\delta} |(f(x+u) + f(x-u) - 2s(x))| + \frac{4\|f\|_1}{2n\sin^2\delta/2} && \text{(by (F3), (F4)).}
\end{aligned}$$

Take $\varepsilon > 0$ and choose δ such that $|(f(x+u) + f(x-u) - 2s(x))| < \varepsilon$ for $0 < u < \delta$; in general δ will depend on x. Then $|\sigma_n(x) - s(x)|$ can be made arbitrarily small by first choosing δ suitably small and then taking n sufficiently large. This proves (a).

For (b) first note that $s(x) = f(x)$ for every x if $f \in \mathrm{C}^{\mathrm{O}}[-\pi,\pi]$. To see that convergence is uniform we require to choose the δ in the preceding paragraph to be **independent of** x. This is possible because on the compact interval $[-\pi,\pi]$ the continuous function f is uniformly continuous; see 9.5.

Finally, if $\{s_n(x)\}$ converges then $\{\sigma_n(x)\}$ converges to the same limit. □

30.5 Commentary on Fejér's Theorem. Notice the difference between the behaviour of D_n and of F_n. The latter is non-negative, and dies away uniformly on $[\delta,\pi)$ as $n \to \infty$. These facts were used to get the estimate of $|\sigma_n(x) - s(x)|$ in 30.4. The Riemann–Lebesgue Lemma, which exploits the oscillatory nature of the sine function and was a crucial ingredient in the proofs in Chapter 29, is not needed in the proof of Fejér's Theorem.

Compare part (a) of 30.4 with the results on convergence of $\{s_n(x)\}$ in Chapter 29. Notice that in (a) we need no smoothness restrictions on f beyond the assumptions—necessary for the conclusion to be meaningful—that f is integrable and $f(x+)$ and $f(x-)$ exist. Further, the final assertion in 30.4 implies that

$$\sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{\mathrm{i}nx} = \frac{1}{2}(f(x+) + f(x-))$$

exactly when both sides make sense. This is pleasing, but not especially useful. Rather sophisticated tests may be needed to show directly that the Fourier series indeed converges. Of course, in our criteria for pointwise convergence in Chapter 29 convergence was guaranteed, as part of the conclusion.

We now turn to convergence in the L^1norm. The first result is closely related to Fejér's Theorem. Its importance comes from the uniqueness theorem which is a corollary.

30.6 L^1-convergence of Cesàro means. Let $f \in L[-\pi, \pi]$. Then $\|\sigma_n - f\|_1 \to 0$ as $n \to \infty$.

Proof. By the formulae in 30.3,

$$\begin{aligned}
2\pi\|\sigma_n - f\|_1 &= 2\pi\int_{-\pi}^{\pi} |\sigma_n(x) - f(x)|\,dx \\
&= \int_{-\pi}^{\pi}\left|\int_0^{\pi} (f(x+u) - f(x))F_n(u)\,du\right| dx \\
&\leqslant \int_{-\pi}^{\pi}\left\{\int_0^{\pi} |f(x+u) - f(x)|\,F_n(u)\,du\right\} dx \\
&= \int_0^{\pi}\left\{\int_{-\pi}^{\pi} |f(x+u) - f(x)|\,F_n(u)\,dx\right\} du \qquad \text{(by FTT)} \\
&= \int_0^{\pi} \|f_u - f\|_1 F_n(u)\,du.
\end{aligned}$$

We now use the fact that $u \mapsto \int |f_u - f|$ is continuous (23.5), so that $\|f_u - f\|_1$ can be made arbitrarily small by taking u sufficiently small. The proof then proceeds in the same manner as that in 30.4. □

30.7 Uniqueness theorem for Fourier series. Let $f, g \in L[-\pi, \pi]$ have Fourier coefficients $\{c_n\}$, $\{d_n\}$. Then $c_n = d_n$ for all $n \in \mathbb{Z}$ only if $f = g$ a.e.

Proof. If f and g have the same Fourier coefficients $\{c_n\}$, then $\|\sigma_n - f\|_1 \to 0$ and $\|\sigma_n - g\|_1 \to 0$. Hence $\|f - g\|_1$ must be zero (by the triangle inequality). Therefore $f = g$ a.e. □

Since $L^2[-\pi, \pi] \subseteq L^1[-\pi, \pi]$, we can form the Fourier series (‡) of $f \in L^2[-\pi, \pi]$. There are two ways to approach the basic theory of Fourier series in L^2. One treats this as part of the theory of orthonormal bases in complete inner product spaces (Hilbert spaces), and develops this general machinery first. The other investigates the special case of Fourier series first, and then sets it in the Hilbert space framework. We opt for the latter strategy, since it enables us to complete our discussion of Fourier series *per se* without encumbering it with abstractions, and at the same time provides a prototype for the general theory presented in the next two chapters.

Before we embark, a brief refresher on complex conjugates may be opportune. Let $z = a + ib$. The complex conjugate, $\overline{z}$, of z is $\overline{z} := a - ib$. It is easy to check that $z\overline{z} = |z|^2$, that $\overline{\overline{z}} = z$, and that z is real if and only if $z = \overline{z}$. Conjugates of complex-valued functions are defined pointwise: $\overline{f}(x) := \overline{f(x)}$ for all x in the domain of f.

At the root of the theory lies the following result, parallel to Exercise 6.2. For the proof, use 12.11.

30.8 Orthogonality relations for $\{e^{inx}\}_{n\in\mathbb{Z}}$. For $m, n \in \mathbb{Z}$,

$$\int_{-\pi}^{\pi} e^{imx}e^{-inx}\,dx = \begin{cases} 0 & (m \neq n), \\ 2\pi & (m = n). \end{cases}$$

Let $e_n(x) := e^{inx}/\sqrt{2\pi}$ $(n \in \mathbb{Z})$. Then

$$\langle e_n, e_m\rangle := \int_{-\pi}^{\pi} e^{imx}e^{-inx}\,dx = \begin{cases} 0 & (m \neq n), \\ 1 & (m = n). \end{cases}$$

[In the terminology of Chapter 31 this is the statement that $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal set with respect to the inner product

$$\langle f, g\rangle := \int_{-\pi}^{\pi} f(x)\overline{g(x)}\,dx$$

on $L^2[-\pi,\pi]$, which is linked to the L^2 norm,

$$\|f\|_2 := \left(\int_{-\pi}^{\pi} |f(x)|^2\,dx\right)^{1/2},$$

by the equation $\langle f, f\rangle = \|f\|_2^2$. It is this association of the norm with an inner product which gives the L^2 theory its distinctive flavour—quite different from the treatment of pointwise convergence.]

Among other things the following result shows that s_n, where as usual $s_n(x) = \sum_{k=-n}^{n} c_k e^{ikx}$, gives a better L^2-approximation to f than any other function $t_n(x) = \sum_{k=-n}^{n} d_k e^{ikx}$ $(d_k \in \mathbb{C})$ (such a function t_n is called a *(complex) trigonometric polynomial*).

30.9 Partial sums of Fourier series of L^2 functions. Let $f \in L^2[-\pi,\pi]$. Let $t_n(x) := \sum_{k=-n}^{n} d_k e^{ikx}$ $(d_k \in \mathbb{C})$. Then

(a) $\|f - s_n\|_2 \leqslant \|f - t_n\|_2$, with equality if and only if $t_n = s_n$;

(b) $\|f - s_n\|_2^2 = \|f\|_2^2 - 2\pi\sum_{k=-n}^{n} |c_k|^2$;

(c) $\sum |c_k|^2$ converges, and $2\pi\sum_{k=-\infty}^{\infty} |c_k|^2 \leqslant \|f\|^2$.

Proof. For $z \in \mathbb{C}$ we have $|z|^2 = z\overline{z}$. Thus, with j, k running from $-n$ to n,

$$\begin{aligned}
\|f - t_n\|_2^2 &= \int_{-\pi}^{\pi} (f - t_n)(\overline{f - t_n}) \\
&= \int_{-\pi}^{\pi} |f|^2 - \int_{-\pi}^{\pi} f\overline{t_n} - \int_{-\pi}^{\pi} t_n\overline{f} + \int_{-\pi}^{\pi} t_n\overline{t_n} \\
&= \int_{-\pi}^{\pi} |f|^2 - \sum_j \overline{d_j} \int_{-\pi}^{\pi} f(x)\mathrm{e}^{-\mathrm{i}jx}\,\mathrm{d}x - \sum_k d_k \int_{-\pi}^{\pi} f(x)\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x \\
&\qquad\qquad + \sum_{j,k} d_k\overline{d_j} \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}kx}\mathrm{e}^{-\mathrm{i}jx}\,\mathrm{d}x \\
&= \|f\|^2 - 2\pi \sum_k d_k\overline{c_k} - 2\pi\sum_k \overline{d_k}c_k + 2\pi \sum_k |d_k|^2 \\
&= \|f\|^2 - 2\pi \sum_k |c_k|^2 + 2\pi \sum_k |d_k - c_k|^2 .
\end{aligned}$$

All the required conclusions now tumble out. The term $\sum_k |d_k - c_k|^2$ is non-negative, and is zero if and only if $d_k = c_k$ for all k. This gives (a). Part (b) is immediate when we take $d_k = c_k$, and (c) follows because a series of non-negative terms converges if and only if its partial sums are bounded above. □

30.10 L^2**-convergence of trigonometric series.** Let $\{d_k\}$ be a complex sequence. A series $\sum_{k=-\infty}^{\infty} d_k\mathrm{e}^{\mathrm{i}kx}$ is called a *trigonometric series*. We say that this series converges if its partial sums converge in the L^2 sense, that is, if there exists $f \in \mathrm{L}^2[-\pi, \pi]$ such that

$$\|f - t_n\|_2 \to 0 \ \text{ as } n \to \infty, \quad \text{where } t_n(x) := \sum_{k=-n}^{n} d_k\mathrm{e}^{\mathrm{i}kx}.$$

By the Riesz–Fischer Theorem, 28.16, $\mathrm{L}^2[-\pi, \pi]$ is a complete normed space. Therefore $\{t_n\}$ converges if it is a Cauchy sequence.

Now suppose $\sum_{k=-\infty}^{\infty} |d_k|^2$ converges. For $n > m$ we have, using 30.8,

$$\|t_n - t_m\|^2 = \sum_{m<|k|\leqslant n} |d_k|^2 ,$$

and this tends to 0 as $m, n \to \infty$, by 30.9. Hence there exists $f \in \mathrm{L}^2[-\pi, \pi]$ such that $\sum_{k=-\infty}^{\infty} d_k\mathrm{e}^{\mathrm{i}kx}$ converges to f.

30.11 Theorem (L^2**-convergence of Fourier series**). Let $f \in \mathrm{L}^2[-\pi, \pi]$, with Fourier series $\sum_{k=-\infty}^{\infty} c_k\mathrm{e}^{\mathrm{i}kx}$. Then

(a) $\sum_{k=-\infty}^{\infty} |c_k|^2$ converges;

(b) $\sum_{k=-\infty}^{\infty} c_k\mathrm{e}^{\mathrm{i}kx}$ converges in $\mathrm{L}^2[-\pi, \pi]$;

(c) $\|s_n - f\|_2 \to 0$ as $n \to \infty$, where $s_n(x) := \sum_{k=-n}^{n} c_k e^{ikx}$;

(d) $2\pi\|f\|_2^2 = \sum_{k=-\infty}^{\infty} |c_k|^2$;

(e) If $\sum_{k=-\infty}^{\infty} c_k e^{ikx}$ is the Fourier series of $g \in L^2[-\pi,\pi]$, then $g = f$ (in the L^2 sense, that is, $f = g$ a.e.).

Proof. Part (a) comes from 30.9. Applying 30.10 with $d_k := c_k$ we obtain (b).

For (c) we first use 30.2 to choose $g \in C^0[-\pi,\pi]$ with $\|f - g\|_2 < \varepsilon$. We now apply Fejér's Theorem, 30.4, to g to obtain a trigonometric polynomial $t_m = \sum_{k=-m}^{m} d_k e^{ikx}$ (some Cesàro mean of g) such that $\|g - t_m\|_\infty < \varepsilon$. Then

$$\|g - t_m\|_2^2 = \int_{-\pi}^{\pi} |g - t|^2 \leqslant 2\pi\varepsilon^2.$$

By the triangle inequality, $\|f - t_m\|_2 < K\varepsilon$, where $K = \sqrt{2\pi} + 1$. Finally, we invoke the best approximation property, 30.9(a), to obtain $\|f - s_n\|_2 < K\varepsilon$ for all $n \geqslant m$. This suffices to prove (c). We derive (d) from 30.9(b).

The uniqueness property stated in (e) follows from 30.9 and (c). □

30.12 Remarks. It is worth examining closely the difference between the statements (b) and (c) in 30.11. The former says that the Fourier series of f converges to **some** function g in $L^2[-\pi,\pi]$; the latter makes the stronger assertion that it converges to f itself. Statements (b) and (d) together would imply (c)—but this does not help us since we needed (c) to prove (d).

What is at issue here is the completeness (as defined below in 32.5) of the orthonormal sequence $\{e_n\}_{n\in\mathbb{Z}}$. It is described by (b), (d), or (e), any one of which implies the others (see 32.4). There are several ways to arrive at the completeness property, of which that via Fejér's Theorem is one; there is no entirely elementary route.

In view of its practical usefulness we restate (d) of 30.11 separately. A mnemonic for the formulae (but not a proof!) is provided by formally taking (†) or (‡) as in 29.1, squaring it to get a double sum, integrating term-by-term twice over $[-\pi,\pi]$, and using the orthogonality relations (Exercise 6.2 or 30.8) to dispose of the unwanted terms.

30.13 Parseval's Theorem. Let $f \in L^2[-\pi,\pi]$. Then

for (‡), $$\|f\|_2^2 = 2\pi \sum_{k=-\infty}^{\infty} |c_k|^2,$$

for (†), $$\|f\|_2^2 = \frac{1}{2}\pi a_0^2 + \pi \sum_{n=1}^{\infty} (a_n^2 + b_n^2).$$

Proof. For (‡) use 30.9(c) and the last part of 30.11. Then for (†) use the relationships $2c_n = a_n - ib_n$ $(n \geqslant 0)$ and $2c_n = a_n + ib_n$ $(n < 0)$ (check these!). □

30.14 Example. All the functions whose Fourier series we presented in Chapter 29 were L^2 functions, so Parseval's Theorem applies to them. To take one example: let $f(x) = x^2$ on $[-\pi, \pi]$, so that $a_0 = 2\pi^2/3$, $a_n = 4(-1)^n/n^2$, and $b_n = 0$ $(n \geqslant 1)$. Thus Parseval's Theorem gives

$$\frac{2}{5}\pi^4 = \frac{1}{\pi}\int_{-\pi}^{\pi} \left|x^2\right|^2 \, dx = \frac{2}{9}\pi^4 + 16\sum_{n=1}^{\infty} \frac{1}{n^4}.$$

We deduce that

$$\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{16}\left(\frac{2}{5} - \frac{2}{9}\right) = \frac{\pi^4}{90}.$$

30.15 Exercise example. Use Fourier series given in Chapter 29 to prove that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \quad \text{and} \quad \sum_{m=1}^{\infty} \frac{1}{(2m-1)^4} = \frac{\pi^4}{96}.$$

30.16 Tailpiece: the good and bad behaviour of Fourier series. To put in context the elementary theory we have developed we set our main results alongside some much deeper statements. By the end of the 19th century the search for criteria for pointwise convergence, for Cesàro summability, and for other forms of summability, had led into a web of little by-ways. Shortly thereafter, it emerged that the Fourier series of a continuous function could diverge, at a single point or even at very many points. Fourier analysis was given a new, and clearer, focus by the development of the Lebesgue integral and by the study of the L^p-spaces, as indicated by the results we have presented. Also, forty years apart, two definitive results on pointwise convergence were published: in $L^1[-\pi, \pi]$ there exist functions whose Fourier series diverge everywhere (I. Kaplansky, 1926), but in the smaller space $L^2[-\pi, \pi]$ every function's Fourier series converges pointwise a.e. (L. Carleson, 1966).

	always true	may occur
$L[-\pi, \pi]$	$\lVert\sigma_n - f\rVert_1 \to 0$	$\{s_n(x)\}$ diverges a.e.
$L^2[-\pi, \pi]$	$\lVert s_n - f\rVert_2 \to 0$ $s_n \xrightarrow{\text{a.e.}} f$	
$C^0[-\pi, \pi]$	$\sigma_n \xrightarrow{\text{u}} f$	$\{s_n(x)\}$ diverges on an uncountable dense set

Table 30.1

We proved in 30.11 that any complex trigonometric series $\sum_{n=-\infty}^{\infty} d_n e^{inx}$ for which $\sum |d_n|^2$ converges is the Fourier series of some $f \in L^2[-\pi,\pi]$. It is natural to ask whether there is an analogous characterization of sequences arising as Fourier coefficients of functions in other classes. This turns out to be a difficult question. In particular it is not a simple matter to determine which convergent trigonometric series are Fourier series of functions in $L^{\circ}[-\pi,\pi]$. Exercise 30.9 gives the steps in a proof that $\sum_{n\geqslant 2} \sin nx/\log n$ is not a Fourier series. On the other hand, $\sum_{n\geqslant 2} \cos n/\log n$ can be shown to be a Fourier series.

For fuller discussions of the behaviour and characterization of Fourier series, see the specialized texts [**11**], [**4**], and [**5**].

Exercises

30.1 Calculate $s_n := a_1 + \cdots + a_n$ and $\sigma_n := (s_1 + \cdots + s_n)/n$ for the following choices of a_n, and decide whether $\lim s_n$ and $\lim \sigma_n$ exist:

$$\text{(i) } 2^{-n}, \quad \text{(ii) } (n(\text{mod } 3) - 1), \quad \text{(iii) } i^n.$$

30.2 Check the details of the proof in 30.2, taking care not to confuse the various norms that arise.

30.3 Let α be irrational. Prove that for any $f \in C^{\circ}[-\pi,\pi]$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(2\pi n\alpha).$$

[Hint: do it first for $f(x) = e^{ikx}$ $(k \in \mathbb{Z})$ and then appeal to Fejér's Theorem.]

30.4 Let $\{d_k\}_{k\in\mathbb{Z}}$ be a doubly infinite sequence of complex numbers. Define

$$\tau_n(x) := \frac{1}{n}\sum_{k=0}^{n-1}\Big(\sum_{m=-n}^{n} d_m e^{imx}\Big) \quad (n = 1, 2, \dots).$$

(a) Prove that

$$\tau_n(x) = \sum_{m=-n+1}^{n-1}\Big(1 - \frac{|m|}{n}\Big)e^{imx}$$

(b) Let $f \in C^{\circ}[-\pi,\pi]$ be such that $\tau_n \xrightarrow{u} f$ on $[-\pi,\pi]$. Prove that $d_n = c_n$, the nth Fourier coefficient of f. [This is a special case of 30.7.]

30.5 Let $f, g \in L^{\circ}[-\pi,\pi]$ have Fourier coefficients $\{c_n\}$, $\{d_n\}$ respectively. Consider the integral

$$\int_{-\pi}^{\pi}\Big\{\int_{-\pi}^{\pi} f(t)g(x-t)e^{-inx}\,dx\Big\}\,dt \quad (n \in \mathbb{Z}).$$

(a) Take $n = 0$ and apply first Tonelli's Theorem and then Fubini's Theorem to show that

$$(f * g)(x) = \int_{-\pi}^{\pi} f(t)g(x-t)\,\mathrm{d}t$$

is defined for almost all x and belongs to $\mathrm{L}^{\circ}[-\pi, \pi]$. [cf. 26.15 and 26.16.]

(b) Use FTT to prove that $f * g$ has Fourier coefficients $\{c_n d_n/(2\pi)\}$.

30.6 (a) Calculate the Fourier coefficients $\{c_n\}$ for e^x on $[-\pi, \pi)$.

(b) Deduce that

$$\frac{1}{2}\pi \coth \pi = \frac{1}{2} + \sum_{n=1}^{\infty} \frac{1}{1+n^2}.$$

(c) By considering the Fourier coefficients of $\mathrm{e}^x \pm \mathrm{e}^{-x}$ evaluate

$$\sum_{n=1}^{\infty} \frac{1}{(1+n^2)^2} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{n^2}{(1+n^2)^2}.$$

30.7 Let $\alpha \in \mathbb{R} \setminus \mathbb{Z}$. Compute the Fourier coefficients $\{c_n\}$ for $\sin \alpha x$ on $[-\pi, \pi)$ [computational hint: write $\sin \alpha x$ as $\frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}\alpha x} - \mathrm{e}^{-\mathrm{i}\alpha x})$]. Deduce that

$$\sum_{n=1}^{\infty} \frac{n^2}{\alpha^2 - n^2} = \frac{\pi(2\pi\alpha - \sin 2\pi\alpha)}{8\alpha \sin^2 \pi\alpha}.$$

30.8 Let f be such that $f, f' \in \mathrm{C}^{\circ}[-\pi, \pi]$ and suppose that $\int_{-\pi}^{\pi} f = 0$. Use Parseval's Theorem to prove that $\|f\|_2 \leqslant \|f'\|_2$, with equality if and only if there exist constants α, β such that $f(x) = \alpha\mathrm{e}^{\mathrm{i}x} + \beta\mathrm{e}^{-\mathrm{i}x}$ $(x \in [-\pi, \pi])$.

30.9 (a) Use Exercise 7.7 and Exercise 29.6 to prove that $\sum_{n\geqslant 2} \sin nx/\log n$ is not the Fourier series of any $f \in \mathrm{L}^{\circ}[-\pi, \pi]$.

(b) Show that $\sum_{n\geqslant 2} \sin nx/\log n$ converges for every x. [Hints: let $\alpha_n := \sin nx$ and $\beta_n := 1/\log n$ and $\mu_n := \alpha_2 + \cdots + \alpha_n$ $(n \geqslant 2)$ and $\mu_1 = 0$; then write

$$\alpha_2\beta_2 + \cdots + \alpha_n\beta_n = (\mu_2 - \mu_1)\beta_2 + \cdots + (\mu_n - \mu_{n-1})\beta_n = \\ \mu_2(\beta_2 - \beta_3) + \cdots + \mu_{n-1}(\beta_{n-1} - \beta_n) + \mu_n\beta_n,$$

and use the Telescoping Lemma, 2.19, and Exercise 29.1. [The same strategy can be used to prove that a real series $\sum c_n d_n$ converges if $\sum c_n$ converges and $d_n \nearrow d \in \mathbb{R}$ (**Abel's test**).]

(d) Deduce from 30.6 that $\sum_{n\geqslant 2} \sin nx/\log n$ is not integrable on $[-\pi, \pi]$.

31 L^2-spaces: orthogonal sequences

Among the L^p-spaces, L^2 is special. Like the Euclidean spaces $\mathbb{R}^k$, its norm comes from an inner product. This gives the notion of orthogonality which underlies Fourier theory in L^2, and gives a rich geometrical theory.

In this chapter we assume familiarity with rudimentary linear algebra. Where exactly the same arguments work in an arbitrary inner product space as in the special case of an L^2-space we work in the general setting. Our results are motivated by the geometry of the most familiar inner product spaces, namely the real Euclidean spaces $\mathbb{R}^k$ $(k = 1, 2, 3)$, with the usual scalar product as the inner product. They subsume the corresponding results in Chapter 30.

31.1 Inner product spaces. A complex inner product space is a vector space, X, over $\mathbb{C}$, equipped with a function $(u, v) \mapsto \langle u, v\rangle$ from $X \times X$ to $\mathbb{C}$ such that.

(IP1) for all $u \in X$, $\langle u, u\rangle \geqslant 0$, with equality if and only if $u = 0$;

(IP2) for all $u, v \in X$, $\langle v, u\rangle = \overline{\langle u, v\rangle}$;

(IP3) for all $u_1, u_2, v \in X$ and $\lambda \in \mathbb{C}$, $\langle u_1 + \lambda u_2, v\rangle = \langle u_1, v\rangle + \lambda\langle u_2, v\rangle$.

To define a real inner product space take X to be a vector space over $\mathbb{R}$, restrict the scalar λ to be real, and omit the complex conjugate sign in (IP2). Any inner product space has an associated norm: $\|u\| := \langle u, u\rangle^{1/2}$.

Throughout, I will denote an arbitrary interval in $\mathbb{R}$. We work with the normed space $L^2(I)$, which as usual we write as L^2 when $I = \mathbb{R}$. We treat the elements of $L^2(I)$ as though they were individual functions but remembering that we regard functions equal a.e. as indistinguishable. We make $L^2(I)$ into an inner product space by defining

$$\langle f, g\rangle := \int_I f(x)\overline{g(x)}\,dx.$$

31.2 Orthogonality. We say elements u, v in an inner product space X are *orthogonal* (and write $u \perp v = 0$) if $\langle u, v\rangle = 0$. A subset S of X is called *orthogonal* if $u \perp v = 0$ for all $u, v \in S$ with $u \neq v$, and is *orthonormal* if in addition $\langle u, u\rangle = 1$ for all $u \in S$.

We can consider all vectors orthogonal to a given vector or set of vectors: given $\varnothing \neq S \subseteq X$ we define $S^\perp := \{v \in X \mid u \perp v = 0 \text{ for all } u \in S\}$. It is an easy exercise to show that $S^\perp$ is always a subspace in the vector space sense (that is, is closed under addition and scalar multiplication).

For our purposes the most important orthonormal sets are infinite sequences, a key example being $\{e^{inx}/\sqrt{2\pi}\}_{n\in\mathbb{Z}}$ in $L^2[-\pi,\pi]$. We now present examples of infinite orthonormal sequences in $L^2(\mathbb{R})$, $L^2[-1,1]$, and $L^2(\mathbb{R}^+)$. We shall show later that these examples are less diverse than they may at first sight appear.

31.3 Example: the Hermite functions. Define, for $n = 0, 1, 2, \ldots$,

$$H_n(x) = (-1)^n e^{x^2} D^n(e^{-x^2}) \quad \text{and} \quad h_n(x) = e^{-x^2/2} H_n(x);$$

here D^n denotes the differentiation operator d/dx performed n times. It is easily shown that $H_n(x)$ is a polynomial of degree n. The exponential damping factor ensures that $h_n \in L^2(\mathbb{R})$. We claim that $\{h_n\}_{n\geqslant 0}$ is an orthogonal sequence. This can be checked by repeated integration by parts. Assume $m \geqslant n$. Then

$$\begin{aligned}
\langle h_n, h_m\rangle &= \int_{-\infty}^{\infty} e^{-x^2} H_m(x) H_n(x)\,dx \\
&= \int_{-\infty}^{\infty} (-1)^m D^m(e^{-x^2}) H_n(x)\,dx \\
&= (-1)^m \Big\{ \Big[D^{m-1}(e^{-x^2}) H_n(x)\Big]_{-\infty}^{\infty} - 2n\int_{-\infty}^{\infty} D^{m-1}(e^{-x^2}) H_{n-1}(x)\,dx \Big\} \\
&= \ldots \\
&= (-1)^{m+n} 2^n n! \int_{-\infty}^{\infty} H_0(x) D^{m-n} e^{-x^2}\,dx \\
&= 0 \quad \text{for } m > n,
\end{aligned}$$

by the FTC. Moreover, putting $m = n$,

$$\langle h_n, h_n\rangle = \int_{-\infty}^{\infty} e^{-x^2} H_n(x)^2\,dx = 2^n n! \int_{-\infty}^{\infty} e^{-x^2}\,dx = 2^n n! \sqrt{\pi}.$$

Therefore, $\|h_n\|_2 = (2^n n!\sqrt{\pi})^{1/2}$, and by dividing h_n by this normalization factor, we produce an orthonormal sequence $\{h_n/\|h_n\|_2\}$ in $L^2(\mathbb{R})$.

31.4 Exercise example: the Legendre polynomials. Define the Legendre polynomials, $P_0(x), P_1(x), \ldots$, by

$$P_n(x) = \frac{1}{2^n n!} D^n\left((x^2-1)^n\right).$$

(a) Prove that $P_n(x)$ is a polynomial of degree n with $P_n(0) = 1$ and $P_n(\pm 1) = 0$ ($n \geqslant 1$). Prove also that any polynomial of degree n is a linear combination of $P_0(x), \ldots, P_n(x)$.

(b) Integrate by parts to prove that

$$2^{m+n} m! n! \int_{-1}^{1} P_n(x) P_m(x)\,dx = \int_{-1}^{1} D^{m+n}\left((x^2-1)^n\right)dx.$$

(c) Deduce that $\{n+\frac{1}{2})^{1/2}P_n(x)\}_{n\geqslant 0}$ is an orthonormal sequence in $L^2[-1,1]$. [Hint: Exercise 6.5.]

31.5 Exercise example: the Laguerre functions. Define $\psi_n\colon \mathbb{R}^+ \to \mathbb{R}$ by

$$\psi_n(x) = \frac{1}{n!}e^{x/2}D^n\left(x^n e^{-x}\right) \quad (n = 0,1,2,\dots).$$

(a) Prove that $\psi_n(x) = e^{-x/2}L_n(x)$, where L_n is a polynomial of degree n with leading coefficient $(-1)^n$, and deduce that $\psi_n \in L^2(\mathbb{R}^+)$.
(b) Integrate by parts to prove that

$$\int_0^\infty x^k D^n\left(x^n e^{-x}\right)\,dx = \begin{cases} 0 & \text{if } k < n, \\ (-1)^n & \text{if } k = n. \end{cases}$$

(c) Use (a) and (b) to show that $\langle \psi_m, \psi_m\rangle = 0$ for $m \neq n$ and that

$$\langle \psi_n, \psi_n\rangle = \frac{(-1)^n}{(n!)^2}\int_0^\infty x^n D^n\left(x^n e^{-x}\right)\,dx = 1.$$

The next two examples concern orthogonal solutions of differential equations.

31.6 Example: Legendre's equation. For $n = 0,1,2,\dots$ let $y_n \in C^2[-1,1]$ be a solution of the differential equation

$$(1-x^2)y_n'' - 2xy_n' + n(n+1)y_n = 0.$$

Multiplying the nth equation by y_m and the mth by y_n and subtracting we get

$$(1-x^2)(y_n''y_m - y_m''y_n) - 2x(y_n''y_m - y_m''y_n) + (n(n+1) - m(m+1))y_my_n = 0,$$

that is,

$$D\left((1-x^2)(y_n'y_m - y_m'y_n)\right) = (m(m+1) - n(n+1))y_ny_m.$$

Now integrate over $[-1,1]$ and apply the FTC:

$$(m(m+1) - n(n+1))\int_{-1}^1 y_ny_m = \left[(1-x^2)(y_n'y_m - y_m'y_n)\right]_{-1}^1 = 0.$$

Therefore $\int_{-1}^1 y_ny_m = 0$. We have proved that $\{y_n\}$ is an orthogonal sequence in $L^2[-1,1]$.

31.7 Exercise example. For $n = 0,1,2,\dots$ let y_n be a solution of the differential equation

$$y_n'' = (x^2 - 2n - 1)y_n.$$

Assuming that $y_n \in L^2(\mathbb{R})$ prove that $\{y_n\}_{n\geqslant 0}$ is an orthogonal sequence.

We now develop some general theory pertaining to orthonormal sets, which we shall then apply to the examples above. Although we are principally interested in infinite orthonormal sets it is worth investigating finite orthonormal sets first, since in that case no convergence questions arise.

31.8 Pythagoras' Theorem. Let X be an inner product space. Then

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 \quad \text{if } u \perp v = 0 \text{ in } X.$$

'Theorem' is a rather grand name for a result which is so easily proved: to obtain the formula we merely expand $\langle u+v, u+v\rangle$ to get

$$\langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \overline{\langle u,v\rangle} + \langle v,v\rangle = \langle u,u\rangle + \langle v,v\rangle.$$

By induction, $\|v_1 + \cdots + v_m\|^2 = \|v_1\|^2 + \cdots + \|v_m\|^2$ whenever $v_i \perp v_j = 0$ for $i \neq j$.

31.9 Finite orthonormal sets. To motivate the results which follow consider $\mathbb{R}^3$ with the inner product as the usual scalar product. Take a plane, M, through 0 spanned by two mutually orthogonal unit vectors, u_1, u_2, and a point, x, not on M. Then the shortest route from x to M is along the (unique) line through x perpendicular to M. The point v on M at the foot of this perpendicular is closer to x than is any other point of M, and $x = u + v$, with $x - u = v \in M^{\perp}$. See Fig. 31.1.

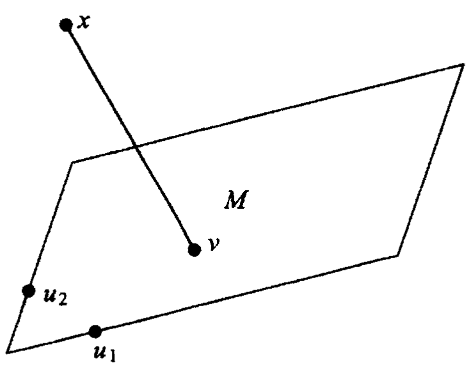


Figure 31.1

We now return to a general inner product space X.

31.10 Properties of finite orthonormal sets. Let $S := \{u_1, \ldots, u_m\}$ be a finite orthonormal set and let $u = \alpha_1 u_1 + \cdots + \alpha_m u_m$ ($\alpha_1, \ldots, \alpha_m$ scalars). Take the product with u_k. Using (IP3) we have $\alpha_k = \langle u, u_k \rangle$ ($k = 1, \ldots, m$). It follows that S is linearly independent (consider $u = 0$) and that every element u in the subspace M spanned by S is uniquely expressible as $u = \sum_{i=1}^{m} \langle u, u_i \rangle u_i$.

Now let $w \in X$ and let $u := \sum_{i=1}^{m} \langle w, u_i \rangle u_i$. Then

$$\langle w - u, u_j \rangle = \langle w, u_j \rangle - \sum_{i=1}^{m} \langle w, u_i \rangle \langle u_i, u_j \rangle = \langle w, u_j \rangle - \langle w, u_j \rangle = 0,$$

so that $w - u$ is orthogonal to each of $u_1, \ldots, u_m$ and so to every element of M. We can therefore write

$$w = u + (w - u), \quad \text{where } u := \sum_{i=1}^{m} \langle w, u_i \rangle u_i \in M \text{ and } (w - u) \in M^{\perp}.$$

By Pythagoras' Theorem,

$$\|w\|^2 = \|w - u\|^2 + \|u\|^2 \geqslant \|u\|^2 = \sum_{i=1}^{m} |\langle w, u_i \rangle|^2.$$

We have proved that $\sum_{i=1}^{m} |\langle w, u_i \rangle|^2 \leqslant \|w\|^2$ (**Bessel's inequality**). When $m = 1$ Bessel's inequality gives $|\langle w, u_1 \rangle| \leqslant \|w\|$ whenever $\|u_1\| = 1$. Let $z \in X$, $z \neq 0$. Then $\|z/\|z\|\| = 1$ so, using (IP3),

$$|\langle w, z \rangle| \leqslant \|w\| \|z\|.$$

This is trivially true if $z = 0$ too. This is the **Cauchy–Schwarz inequality** in its general form; 23.2 gave the result for the case $X = L^2$.

Now take $v = \sum_{i=1}^{m} \alpha_i u_i$ to be a general element of M. Expanding out the product $\langle w - v.w - v \rangle$ using (IP2) and (IP3),

$$\begin{aligned} \|w - v\|^2 &= \langle w - \sum_{i=1}^{m} \alpha_i u_i,\ w - \sum_{i=1}^{m} \alpha_i u_i \rangle \\ &= \|w\|^2 - \sum_{i=1}^{m} \alpha_i \overline{\langle w, u_i \rangle} - \sum_{i=1}^{m} \overline{\alpha_i} \langle w, u_i \rangle + \sum_{i=1}^{m} |\alpha_i|^2 \\ &= \|w\|^2 - \sum_{i=1}^{m} |\langle w, u_i \rangle|^2 + \sum_{i=1}^{m} |\alpha_i - \langle w, u_i \rangle|^2. \end{aligned}$$

This is the general version of the calculation carried out for $L^2[-\pi, \pi]$ in 30.9. From it we deduce that $\|w - v\|^2 = \|w - \sum_{i=1}^{m} \alpha_i u_i\|^2$ is a minimum when $\alpha_i = \langle w, u_i \rangle$ for $i = 1, \ldots, m$, which happens exactly when $w \perp (w - v)$.

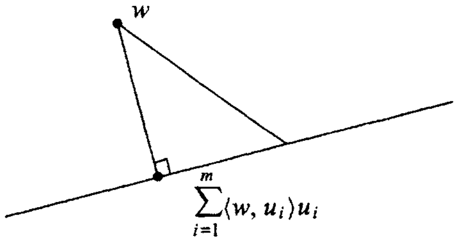


Figure 31.2

31.11 Example. Consider the orthogonal set $\{e^{ijx}\}_{j=1}^{m}$ in $L^2[-\pi,\pi]$. We shall find the best possible approximation in $L^2[-\pi,\pi]$ to the function x by functions of the form $\sum_{j=1}^{m} \alpha_j e^{ijx}$. By 31.10

$$\int_{-\pi}^{\pi} \left| x - \sum_{j=1}^{m} \alpha_j e^{ijx} \right|^2 dx$$

is minimized by choosing

$$\sqrt{2\pi}\alpha_j = \langle x, \frac{1}{\sqrt{2\pi}} e^{ijx} \rangle \quad (j = 1, \dots, m)$$

and the minimum value of the displayed integral is

$$\|x\|^2 - \sum_{j=1}^{m} |\alpha_j|^2 = \frac{2}{3}\pi^3 - 2\pi \sum_{j=1}^{m} j^{-2}.$$

The expression on the right-hand side diminishes as m increases, but is never smaller than $\pi^3/3$; we proved in 29.17 that $\sum_{n=1}^{\infty} n^{-2} = \pi^2/6$.

The minimum-distance result in 31.10 relies on the vectors $u_1, u_2, \dots, u_m$ being orthonormal. Fortunately there is a standard procedure for manufacturing orthonormal sets.

31.12 The Gram–Schmidt process. Suppose we have a sequence $\{v_1, v_2, \dots\}$ in an inner product space X such that $\{v_1, \dots, v_m\}$ is linearly independent for each $m \geqslant 1$. We prove by induction that there exists an orthonormal sequence $\{u_1, u_2, \dots\}$ such that for all $k \geqslant 1$

(i) u_k is a linear combination of $\{v_1, \dots, v_k\}$, and

(ii) $\{v_1, \dots, v_k\}$ and $\{u_1, \dots, u_k\}$ span the same subspace.

For the base step of the induction we need only to normalize v_1, by taking $u_1 := v_1/\|v_1\|$ (note that, as a member of a linearly independent set, $v_1 \neq 0$, so $\|v_1\| \neq 0$ by (IP1)). Now assume that we have constructed an orthonormal set $\{u_1, \dots, u_m\}$ so that (i) and (ii) hold for $k \leqslant m$. Define

$$z_{m+1} := v_{m+1} - (\langle v_{m+1}, u_1\rangle u_1 + \cdots + \langle v_{m+1}, u_m\rangle u_m)$$

and $u_{m+1} := z_{m+1}/\|z_{m+1}\|$. Because $\{v_1, \dots, v_{m+1}\}$ is linearly independent,

$$v_{m+1} \notin \langle v_1, \dots, v_m\rangle = \langle u_1, \dots, u_m\rangle,$$

so $z_{n+1} \neq 0$ and u_{m+1} is well defined. Also, by 31.10, z_{m+1} is orthogonal to each of $u_1, \dots, u_n$. Hence $\{u_1, \dots, u_{m+1}\}$ is orthonormal. Certainly $\langle v_1, \dots, v_{m+1}\rangle \subseteq \langle u_1, \dots, u_{m+1}\rangle$; because both spaces have dimension $m+1$ (remember that an orthonormal set is linearly independent), we have equality.

The uniqueness result below is sufficiently important to be called a theorem.

31.13 Uniqueness theorem for orthogonal polynomials. Assume $\{u_n\}_{n\geqslant 0}$ and $\{w_n\}_{n\geqslant 0}$ are sequences in an inner product space X which contains the polynomials and suppose that

(i) each sequence is orthogonal, and

(ii) each of $u_n(x)$ and $w_n(x)$ is a polynomial of degree n.

Then there exist non-zero constants A_n such that $w_n = A_n u_n$ for all n.

Proof. Fix n. Let $y_n(x) := w_n(x) - A_n u_n(n)$ where A_n is chosen so that $y_n(x)$ is a polynomial of degree $< n$ (that is, choose A_n to cancel the terms in x^n). Then y_n is orthogonal to both u_n and w_n. Therefore $y_n \perp y_n = 0$, so $y_n = 0$. □

31.14 Example. Let us apply the Gram–Schmidt process in $\mathrm{L}^2[-1,1]$ with $\{v_1, v_2, \dots\}$ as $\{1, x, x^2, \dots\}$. We arrive at an orthonormal sequence which, by the Uniqueness Theorem, is the sequence of Legendre polynomials apart from normalization constants. Suppose we wish to find how good an L^2 approximation exists to x^n by a polynomial $p(x)$ of degree at most $n-1$. By 31.4(a) we can express $p(x)$ in the form $\alpha_0 Q_0(x) + \cdots + \alpha_{n-1}Q_{n-1}(x)$, with $\{Q_n(x)\}$ the sequence of Legendre polynomials normalized by taking $Q_n(x) = C_n P_n(x)$ where $C_n = (n+\frac{1}{2})^{1/2}$. The best choice of α_j is $\langle x^n, Q_j(x)\rangle$. This is unappealing to compute. With a little cunning we can arrive at the error $\|x^n - p(x)\|$ more easily. We can choose K so that $x^n - KQ_n(x)$ is a polynomial of degree $\leqslant n-1$. In fact $K = 2^n n! C_n/(2n)! = 2^n n!(n+\frac{1}{2})^{1/2}/(2n)!$, the coefficient of x^n in $Q_n(x)$. Now, for some constants β_j,

$$\Big\|x^n - \sum_{j=0}^{n-1} \alpha_j Q_j(x))\Big\|^2 = \Big\|KQ_n(x) - \sum_{j=0}^{n-1} \beta_j Q_j(x)\Big\|^2 = K^2 + \sum_{j=0}^{n-1} |\beta_j|^2,$$

by Pythagoras' Theorem and the orthonormality of $\{Q_n\}$. This has minimum value K^2, attained when all β_j are zero. Therefore K gives the minimum value of $\|x^n - p(x)\|$ as $p(x)$ varies over polynomials of degree less than n.

31.15 Weighted polynomials and classical orthogonal sequences. Polynomials do not belong to $\mathrm{L}^2(I)$ when I is an unbounded interval. We can overcome this difficulty by incorporating appropriate exponential damping factors. We claim that the Hermite functions arise, up to normalization, from applying the Gram–Schmidt process to $\{x^n e^{-x^2/2}\}$ in $\mathrm{L}^2(\mathbb{R})$ and the Laguerre functions from $\{x^n e^{-x/2}\}$ in $\mathrm{L}^2(\mathbb{R}^+)$. We can bring examples such as these within the scope of the Uniqueness Theorem by working with a redefined inner product. Suppose $w(x)$ is a non-negative integrable function on an interval I. Then

$$\langle f, g\rangle_w := \int_I w(x) f(x)\overline{g(x)}\,\mathrm{d}x$$

defines an inner product on

$$\mathrm{L}^2_w(I) := \{\, f : \mathbb{R} \to \mathbb{C} \mid f \text{ is measurable and } wf^2 \in \mathrm{L}^2(I)\,\}.$$

Table 31.1 indicates how this framework subsumes a range of examples of practical importance. In each case the choice of weight function w guarantees that the space $\mathrm{L}^2_w(I)$ contains the polynomials. In the table, the final column shows the result, up to normalization, of applying the Gram–Schmidt process to $\{1, x, x^2, \dots\}$ and invoking the Uniqueness Theorem.

I	$w(x)$	Orthogonal sequence
$[-1,1]$	1	Legendre polynomials
$(-1,1)$	$(1-x^2)^{-1/2}$	Chebychev polynomials
$\mathbb{R}$	e^{-x^2}	Hermite polynomials, $\{H_n(x)\}$
$\mathbb{R}^+$	e^{-x}	Laguerre polynomials, $\{L_n(x)\}$

Table 31.1

Exhaustive accounts can be found in older analysis texts such as [19] of the properties of these various orthogonal sequences. We give only selected illustrations of such properties and how to derive them. The following may be sought for a sequence $\{p_n\}$ in $\mathrm{L}^2(I)$ or $\mathrm{L}^2_w(I)$:

- orthogonality;
- a differential equation satisfied by p_n, for each n;
- recurrence relations involving the functions p_n and their derivatives;
- a closed form $g(x,t)$ for $\sum t^n p_n(x)$, so $g(x,t)$ serves as a generating function for $\{p_n\}$.

These are inter-related, and a given set of results may be reached by a variety of routes. For example, for the Hermite functions or Hermite polynomials, orthogonality can be approached 'head-on' as in 31.3. Alternatively we deduce orthogonality from a suitable set of differential equations as in Exercise 31.3.

The differential equations may be derived via recurrence relations (see Exercise 31.2(c)). A generating function may be an aid to finding such recurrence relations. Exercise 31.5, on the Chebychev polynomials, exploits the Uniqueness Theorem to get the associated differential equations.

A given orthogonal sequence of polynomials $\{p_n\}$ may be normalized in different ways to suit different applications. In the L^2 setting $\|p_n\|_2 = 1$ may be best, but an alternative normalization may give tidier recurrence relations, or be traditional.

For applications in numerical analysis only the first few terms in an orthogonal sequence are usually needed to give quite accurate results. This is good since coefficients in orthogonal polynomials are not given by simple formulae and explicit calculations quickly get unwieldy. For example, the first few Hermite polynomials are:

$$\begin{array}{lll} H_0(x) = 0, & H_1(x) = 2x, & H_2(x) = 4x^2 - 2, \\ H_3(x) = 8x^3 - 12x, & H_4(x) = 16x^4 - 48x^2 + 12, & H_5(x) = 32x^5 - 160x^3 + 120x. \end{array}$$

We return briefly to the topic of numerical integration, previously discussed in Chapter 7. We refer the interested reader to specialist texts for a detailed account.

31.16 Gaussian quadrature. In this method, which we illustrated by a simple example in 7.13, we seek to find an approximate value for $\int_I wf$, where I is a bounded or unbounded interval, w is a continuous strictly positive integrable function on I and is such that $L^2_w(I)$ contains the polynomials, and $f \in L^2_w(I)$. We aim to write

$$\int_I wf = \sum_{j=1}^{n} a_j f(x_j) + E_n,$$

where the error E_n is zero whenever f is a polynomial of degree $\leqslant 2n-1$. We require the interpolation points, $x_1, \dots, x_n$, and the constants, $a_1, \dots, a_n$, to be independent of f.

The assertions below legitimize the proposed method. Let $\{q_0(x), q_1(x), \dots\}$ be the orthonormal sequence in $L^2_w(I)$ resulting from the Gram–Schmidt process applied to the polynomials, with $q_n(x)$ being of degree n.

(a) $q_n(x)$ has n distinct zeros in I.

(b) $x_1, \dots, x_n$ are necessarily the zeros of $q_n(x)$ in I.

(c) With $x_1, \dots, x_n$ as in (b), there exist unique constants a_j such that $\int_I wf = \sum_{j=1}^{n} a_j f(x_j)$ for every polynomial of degree $\leqslant 2n-1$.

The proofs of these statements are not entirely routine, and owe more to algebra than to integration theory. Consequently we do not attempt fully to justify them, contenting ourselves with just a few comments.

To indicate the role played by orthogonality we prove (b). Let $f(x) = p_n(x)p(x)$, where $p_n(x) := (x - x_1)\dots(x - x_n)$ and $p(x)$ is an arbitrary polynomial of degree $< n$. Then f is of degree at most $2n - 1$, so

$$\int_I wp_np = \int_I wf = 0.$$

This means that p_n is orthogonal to all polynomials of degree $< n$. By the Uniqueness Theorem, p_n and q_n differ at most by a non-zero multiplicative constant and so have the same zeros.

Now assume that I is a compact interval. It can be shown that the error E_n can be expressed as an integral involving $f^{(2n)}$, provided $f \in \mathrm{C}^{2n}(I)$. Then E_n can be estimated by the Integral MVT.

Exercises

31.1 Let X be an inner product space. Prove the parallelogram law, *viz.*

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2) \quad (u, v \in X).$$

Find elements $f, g \in \mathrm{L}^1$ such that

$$\|f+g\|_1^2 + \|f-g\|_1^2 \neq 2(\|f\|_1^2 + \|g\|_1^2).$$

[So the norm on L^1 cannot arise from an inner product in the way that the norm on L^2 does.]

31.2 Consider the Hermite polynomials $\{H_n(x)\}$ defined in 31.3. Prove the following:

(a) $H_n'(x) = 2nH_n(x)$;
(b) $H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0 \ (n \geqslant 1)$;
(c) $H_n''(x) - 2xH_n'(x) + 2nH_n(x) = 0$.

31.3 Use the preceding exercise to show that $h_n(x) = \mathrm{e}^{-x^2/2}H_n(x)$ satisfies the differential equation

$$h_n'' = (x^2 - 2x - 1)h_n.$$

[31.7 then gives another proof of the orthogonality of $\{h_n\}$ in $\mathrm{L}^2(\mathbb{R})$.]

31.4 In $\mathrm{L}^2(\mathbb{R}^+)$, let $\{\psi_n\}$ be the sequence of Laguerre functions defined in 31.5.

(a) By induction, or by quoting Leibnitz' formula for the nth derivative, prove that

$$\psi_n(x) = \sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{n!}{k!}x^k\mathrm{e}^{-x/2}.$$

(b) Let $g(x,t) := \sum_{n=0}^{\infty} t^n\psi_n(x)$. Prove that

$$g(x,t) = (1-t)^{-1}\mathrm{e}^{-((1+t)x)/(2(1-t))} \quad (x \in [0,\infty),\ t \in (-1,1)).$$

(c) With the aid of 31.10 prove that

$$\left\|g(x,t) - \sum_{n=0}^{N} t^n \psi_n(t)\right\|_2 \to 0 \quad \text{as } N \to \infty.$$

31.5 Consider the space $X := L^2_w[-1,1]$ where $w(x) = (1-x^2)^{-1/2}$. Let $T_n(x) := 2^{-n+1}\cos(n\cos^{-1} x)$.

(a) Prove that $\{T_n(x)\}_{n\geqslant 0}$ is an orthogonal sequence in X.

(b) Show that it is possible to find constants a_r such that $t_n(x) = \sum_{r=0}^{n} a_r x^r$ is a solution of the differential equation

$$(1-x^2)t_n'' - xt_n' + n^2 t_n = 0.$$

(c) Let $\tau_n(x) := (1-x^2)^{-1/4} t_n(x)$. Show that τ_n satisfies

$$(1-x^2)\tau_n'' - 2x\tau_n' + \left(n^2 - \frac{1}{4} - \frac{1}{4(1-x^2)}\right)\tau_n = 0,$$

and hence show that $\{t_n\}_{n\geqslant 0}$ is an orthogonal sequence in X.

(d) Show that $T_n(x)$ is a polynomial of degree n and that t_n and T_n are equal up to a non-zero multiplicative constant.

(e) Show that $\{T_n\}$, after normalization, is an orthonormal sequence in X.

31.6 For $n = 1,2,3,\ldots$ define r_n on $[0,1]$ by $r_n(x) = \operatorname{sgn}(\sin(2^n \pi x))$.

(a) Sketch the graphs of r_1, r_2, r_3.

(b) Show that each $r_n \in L^{\text{step}}$ and that $\int r_n = 0$.

(c) Let $n > m$. Show that $r_m(x)r_n(x) = \pm r_n(x)$, and specify how the choice of sign is to be made.

(d) Deduce that $\{r_n\}$ is an orthonormal sequence in $L^2[0,1]$.

[The functions $\{r_n\}$ are the **Rademacher functions**, which have applications in probability theory and elsewhere.]

31.7 Find x_1, x_2, a_1, a_2 such that

$$\int_0^\infty e^{-x} f(x)\, dx = a_1 f(x_1) + a_2 f(x_2)$$

for all polynomials $f(x)$ of degree $\leqslant 3$.

31.8 Find a, b, and α such that

$$\int_{-1}^{1} (1-x^2)^{-1/2} f(x)\, dx = af(\alpha) + bf(-\alpha) + E$$

where $E = 0$ whenever $f(x)$ is a polynomial of degree $\leqslant 3$. Assuming that $f \in C^4[-1,1]$ with $|f^{(4)}| \leqslant M$, show, with the aid of the Integral MVT, that $|E| \leqslant 5\pi M/192$.

32 L^2-spaces as Hilbert spaces

In analysing an L^2-space we can profitably exploit the interplay of its inner product structure with the convergence notions associated with its norm. An inner product space X is called a *Hilbert space* if the associated norm makes X a Banach space (that is, Cauchy sequences in X converge). Theorem 28.16 tells us that L^2 is a Hilbert space, as is $L^2(I)$ for any interval I.

In Chapter 31 we presented examples of infinite orthonormal sets, but in our theory in 31.10 dealt with only finitely many vectors at a time. The exponential Fourier series of $f \in L[-\pi, \pi]$, as defined in Chapter 29, can be written in the form

$$\sum_{k=-\infty}^{\infty} \langle f, e_k \rangle e_k(x) \quad \text{where } e_k(x) := \frac{1}{\sqrt{2\pi}} e^{ikx}.$$

This suggests that we should investigate series of the form $\sum_{k=1}^{\infty} \langle w, u_k \rangle u_k$, where w is an element of an inner product space X and $\{u_k\}$ is an infinite orthonormal sequence in X. Of course now that we have an infinite series we must consider whether it converges or not.

32.1 Technicalities: inner products and limits. Let $\{x_n\}$ be a sequence in an inner product space X and suppose $x_n \to x$, that is, $\|x_n - x\| \to 0$. Let $y \in X$. The Cauchy–Schwarz inequality implies that

$$|\langle x - x_n, y \rangle|^2 \leqslant \|x - x_n\| \|y\| \to 0 \quad \text{as } n \to \infty,$$

so $\langle x_n, y \rangle \to \langle x, y \rangle$. Likewise $\langle y, x_n \rangle \to \langle y, x \rangle$. We have proved that the inner product is continuous with respect to each of its arguments. It follows easily that, for $S \subseteq X$, the orthogonal complement $S^\perp$ is norm-closed, in the sense that for any sequence $\{v_n\}$ in $S^\perp$ with $\|v_n - v\| \to 0$ we have $v \in S^\perp$.

Assume $\sum_{k=1}^{\infty} y_k$ is a convergent series, so there exists $y \in X$ such that $s_n := \sum_{k=1}^{n} y_k \to y$. For any $x \in X$,

$$\sum_{k=1}^{n} \langle x, y_k \rangle = \langle x, s_n \rangle \to \langle x, y \rangle.$$

Here we have used the linearity of the inner product, (IP3), and its continuity. We conclude that

$$\sum_{k=1}^{\infty} \langle x, y_k \rangle = \Big\langle x, \sum_{k=1}^{\infty} y_k \Big\rangle \quad \text{whenever } \sum_{k=1}^{\infty} y_k \text{ converges.}$$

32.2 Convergence of orthogonal series. Let X be a Hilbert space and let $\{u_k\}$ be an infinite orthonormal sequence. Then, for all $w \in X$,

(a) $\sum_{k=1}^{\infty} |\langle w, u_k\rangle|^2$ converges (in $\mathbb{R}$);

(b) $\sum_{k=1}^{\infty} \langle w, u_k\rangle u_k$ converges (in X) and $w - \sum_{k=1}^{\infty}\langle w, u_k\rangle u_k$ is orthogonal to u_j for all j.

Proof. By Bessel's inequality (see 31.10),

$$S_n := \sum_{k=1}^{n} |\langle w, u_k\rangle|^2 \leqslant \|w\|^2 \quad \text{for all } n.$$

The Monotonic Sequence Theorem, 2.15, implies that $\{S_n\}$ converges, which proves (a).

We now prove that the series $\sum_{k=1}^{\infty}\langle w, u_k\rangle u_k$ converges, by showing that its partial sums form a Cauchy sequence. Let $s_n := \sum_{k=1}^{n}\langle w, u_k\rangle u_k$ and let $n > m$. By Pythagoras' Theorem,

$$\|s_n - s_m\|^2 = \Big\| \sum_{k=m+1}^{n} \langle w, u_k\rangle u_k^2\Big\| = \sum_{k=m+1}^{n} |\langle w, u_k\rangle|^2 = S_n - S_m,$$

which tends to 0 as $m, n \to \infty$ since $\{S_n\}$ is converges, and hence a Cauchy sequence in $\mathbb{R}$. By (IP3) and 32.1,

$$\Big\langle w - \sum_{k=1}^{\infty} \langle w, u_k\rangle u_k, u_j\Big\rangle = \langle w, u_j\rangle - \sum_{k=1}^{\infty} \langle w, u_k\rangle\langle u_k, u_j\rangle = 0 \quad (j = 1, 2, \ldots). \quad \square.$$

32.3 The Projection Theorem. Let $S = \{u_k\}$ be an infinite orthonormal set in a Hilbert space X. Thanks to 32.2(b) we may define

$$Pw := \sum_{k=1}^{\infty} \langle w, u_k\rangle u_k.$$

It can be shown easily that $P\colon X \to X$ is a linear map such that $P(Pw) = Pw$ for all $w \in X$, and that the image $M := \{Pw \mid w \in X\}$ of P is a subspace containing S [in fact M is the smallest norm-closed subspace which contains S]. The map P is called the *projection of X onto M*. Further, each $w \in X$ can be written as $w = Pw + (w - Pw)$, with $Pw \in M$ and $w - Pw \in M^{\perp}$ (the latter by 32.2(b) and 32.1). By Pythagoras' Theorem,

$$\|w - y\|^2 = \|(Pw - y) + (w - Pw)\|^2 = \|Pw - y\|^2 + \|w - Pw\|^2 \geqslant \|w - Pw\|^2,$$

for any $y \in M$. We have shown that Pw is closer to w than any other point of M. Thus we have proved for an infinite orthonormal sequence a result analogous to that proved in 31.10 for a finite orthonormal set. [We have here instances of a fundamental result—the Projection Theorem—which asserts that for any norm-closed subspace M in a Hilbert space X we have $X = M \oplus M^{\perp}$.]

32.4 Theorem. Let $\{u_k\}$ be an orthonormal sequence in a Hilbert space X. Then the following are equivalent:

(ONB1) $w = \sum_{k=1}^{\infty} \langle w, u_k \rangle u_k$ for all $w \in X$;

(ONB2) $\langle v, w \rangle = \sum_{k=1}^{\infty} \langle v, u_k \rangle \langle u_k, w \rangle$ for all $v, w \in X$;

(ONB3) $\|w\|^2 = \sum_{k=1}^{\infty} |\langle w, u_k \rangle|^2$ for all $w \in X$ (**Parseval's formula**);

(ONB4) for all $w \in X$, $\langle w, u_k \rangle = 0$ for all k implies $w = 0$;

(ONB5) given $w \in X$ and $\varepsilon > 0$ there exist $m \geqslant 1$ and scalars $\alpha_1, \dots, \alpha_m$ such that $\|w - \sum_{j=1}^{m} \alpha_j u_j\| < \varepsilon$.

Proof. Assume (ONB1). This implies (ONB2) by 32.1. Trivially (ONB2) implies (ONB3) and (ONB3) implies (ONB4). Now assume (ONB4). By (b) in 32.2 there exists $u \in X$ such that $u = \sum_{k=1}^{\infty} \langle w, u_k \rangle u_k$, and $\langle (w - u), u_j \rangle = 0$ for all j. Therefore $u = w$ by (ONB4), so (ONB1) holds. Since (ONB1) means by definition that $\|w - \sum_{k=1}^{n} \langle w, u_k \rangle u_k\| \to 0$ as $n \to \infty$, we see that (ONB1) implies (ONB5). The reverse implication comes from 31.10. □

32.5 Orthonormal bases. An orthonormal sequence $\{u_k\}$ satisfying one and hence all of the conditions (ONB1)–(ONB5) is said to be a *complete orthonormal set* or an *orthonormal basis*. [An orthonormal basis is not a basis as the term is used in linear algebra, except when X is finite-dimensional. For a set S to be a basis of an arbitrary vector space X every element of X must by definition be a **finite** linear combination of elements of S.] To show that a given orthonormal set is complete we usually use either (ONB4) or (ONB5).

Because of its importance we record explicitly our earlier theorem on L^2-convergence of Fourier series in the present terminology.

32.6 Theorem.

(a) $\{e^{ikx}/\sqrt{2\pi}\}_{k \in \mathbb{Z}}$ is an orthonormal basis for the complex space $\mathrm{L}^2[-\pi, \pi]$.

(b) $\{1/\sqrt{2\pi}, \cos x/\sqrt{\pi}, \sin x/\sqrt{\pi}, \cos 2x/\sqrt{\pi}, \sin 2x/\sqrt{\pi}, \dots\}$ is an orthonormal basis for the real space $\mathrm{L}^2_{\mathbb{R}}[-\pi, \pi]$.

Completeness of an orthonormal set in an L^2-space is a Good Thing. It allows us to find Fourier-type expansions for functions in the space, and to exploit these in the same way as we exploit Fourier series in $\mathrm{L}^2[-\pi, \pi]$. We would therefore like to prove that the orthonormal sequences introduced in Chapter 31—the Legendre polynomials and the Hermite functions, etc.—are complete. Before discussing these sequences we make some comments to pull various ideas together.

32.7 Density and orthonormal bases. Let $S = \{u_k\}$ be an orthonormal sequence in a Hilbert space X. Let L be the subspace consisting of all **finite** linear combinations of elements of S. (Thus if $S = \{1, x, x^2, \dots\}$ in $L_{\mathbb{R}}[-1,1]$ then L is the space of polynomials and if $S = \{e^{ikx}\}$ in $L^2[-\pi,\pi]$ then L is the space of trigonometric polynomials.) Certainly $L \subseteq M$, where

$$M := \Big\{ \sum_{k=1}^{\infty} \langle x, u_k \rangle u_k \mid x \in X \Big\},$$

since M is a subspace of X containing each u_j. We claim that the following conditions are equivalent:

(a) L is dense in X;

(b) the orthonormal set S is complete;

(c) $M = X$.

Theorem 32.4 tells us that (b) and (c) are equivalent and imply (a). If (a) holds then, given $w \in X$ and $\varepsilon > 0$, we can find $v \in L$ such that $\|w - v\| < \varepsilon$. Arguing just as in 32.4 we deduce that S is complete.

We next consider the density of the polynomials in various spaces of functions, for various norms.

32.8 Polynomial approximation. Let $f \in C^{\circ}[-\pi,\pi]$ (as defined in 30.1) and let $\varepsilon > 0$ be given. We shall show that we can find a polynomial $q(x)$ such that $|f(x) - q(x)| < C\varepsilon$ for all $x \in [-\pi,\pi]$, where C is some constant independent of x. Fejér's Theorem tells us that f can be uniformly approximated by functions of the form $\sum_{k=-n}^{n} c_k e^{ikx}$. Each e^{ikx} has an expansion

$$e^{ikx} = \sum_{m=0}^{\infty} \frac{1}{m!}(ikx)^m,$$

and this series is uniformly convergent on $[-\pi,\pi]$ (use the M-test, 8.11). By chopping off the tail of the series we can find a polynomial $p_k(x)$ with complex coefficients such that $|e^{ikx} - p_k(x)| < \varepsilon$ for all x. We can then take $q(x) := \sum_{k=-n}^{n} c_k p_k(x)$.

The preceding result is an interim one. We now rescale to an arbitrary compact interval, at the same time removing the 'equal values at endpoints' condition demanded in $C^{\circ}[-\pi,\pi]$. Superscripts in the following proof indicate scaling factors: $F^d(x) := F(dx)$.

32.9 Weierstrass' approximation theorem. Let $[a,b]$ be a compact interval in $\mathbb{R}$ and let $f\colon [a,b] \to \mathbb{C}$ be continuous. Then, given $\varepsilon > 0$, there exists a polynomial p such that

$$\|f - p\|_{\infty} := \sup\{\, |f(x) - p(x)| \mid x \in [a,b] \,\} < \varepsilon.$$

Proof. Choose $N \in \mathbb{N}$ such that $-N\pi < a < b < N\pi$. Extend f to a continuous function F on an interval $[-N\pi, N\pi]$ so that $F(-N\pi) = F(N\pi)$. This can be done by making F linear on $[-N\pi, a]$ and on $[b, N\pi]$, as shown in Fig. 32.1.

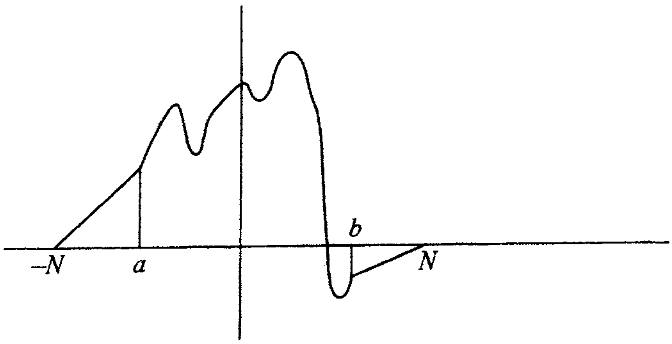


Figure 32.1

Rescaling, we have $F^{2N} \in C^{\circ}[-\pi, \pi]$, and so there exists by 32.8 a polynomial q such that

$$\sup\{|F^{2N}(x) - q(x)| \mid x \in [-\pi, \pi]\} < \varepsilon.$$

Then $p := q^{1/(2N)}$ approximates F uniformly to within ε on $[-N\pi, N\pi]$, and hence approximates f likewise on $[a, b]$. $\square$

32.10 Completeness of the Legendre polynomials in $L^2[-1,1]$**.** We want to prove that the set of polynomials is dense in $L^2[-1,1]$ (see 32.7). Weierstrass' Theorem tells us that the polynomials are dense in $C[-1,1]$, with respect to its norm $\|.\|_\infty$. This is neither the space we want nor the norm we want, but it is easy to patch things up. Take $f \in L^2[-1,1]$ and fix the customary $\varepsilon > 0$. Because $C[-1,1]$ is dense in $L^2[-1,1]$ (see 28.11), we can find a continuous function g with $\|f - g\|_2 < \varepsilon$. Now choose a polynomial q such that $\|g - q\|_\infty < \varepsilon/\sqrt{2}$. Then

$$\|f - q\|_2 \leqslant \|f - g\|_2 + \|g - q\|_2 < \varepsilon + \left(\int_{-1}^{1} |g - q|^2\right)^{1/2} < 2\varepsilon.$$

32.11 Establishing completeness: summary. We can adapt the argument of the preceding section to prove completeness of the polynomials in weighted L^2-spaces on bounded intervals. This leads to a proof, for example, of the completeness of the Chebychev polynomials in $L^2_w[-1,1]$, where the weight function is $w(x) = (1 - x^2)^{-1/2}$ (Exercise 32.4).

On unbounded intervals different techniques are needed. The usual approach is to show that (ONB5) holds. For example, for the normalized Hermite functions,

$\{h_n/\|h_n\|_2\}$, it suffices to show that $\int_{-\infty}^{\infty} f(x)h_n(x)\,dx = 0$ $(n = 0, 1, 2, \dots)$ implies $f = 0$, for every $f \in L^2(\mathbb{R})$. We cannot do this yet. We do it in the next chapter, with the aid of the Fourier transform. A method similar to that used for the Hermite functions works for the Laguerre functions, but based on the Laplace transform instead of the Fourier transform.

To show non-completeness we usually use the contrapositive of (ONB5). See Exercise 32.1 for an example.

Exercises

32.1 Let $\{r_n\}$ be the sequence of Rademacher functions introduced in Exercise 31.6. Show that $r_n(x) = -r_n(1-x)$ for all x, for each n and that $f \perp r_n = 0$ for all n when $f \in L^2[0,1]$ is such that $f(x) = f(1-x)$ for all x. Deduce that $\{r_n\}$ is not complete in $L^2[0,1]$.

32.2 Let g be a continuous real-valued function on $[0,1]$ and assume that

$$\int_0^1 x^n g(x)\,dx = 0, \quad (n = 0, 1, 2, \dots).$$

By approximating g uniformly by polynomials show that

$$\int_0^1 g(x)^2\,dx = 0$$

and deduce that $g \equiv 0$.

32.3 Recall the Gaussian quadrature method introduced in 31.16. Suppose that w is a continuous strictly positive function on the compact interval $[a,b]$ and that $f \in C[a,b]$. Assume that

$$\int_I wf = \sum_{j=1}^{n} a_j f(x_j) + E_n,$$

where $a_1, \dots, a_n$ are constants, $x_1, \dots, x_n$ are fixed points in $[a,b]$, and the error E_n is zero whenever f is a polynomial of degree $\leqslant 2n-1$. Use Weierstrass' Theorem to show that $E_n \to 0$ as $n \to \infty$.

32.4 [Continuation of Exercise 31.5] Prove that the inner product space $L^2_w[-1,1]$, where $w(x) = (1-x^2)^{-1/2}$, is a Hilbert space, in which the orthogonal sequence $\{T_n\}$ of Chebychev polynomials, when normalized, is an orthonormal basis.

33 Fourier transforms

This chapter introduces the theory of the Fourier transform, parallel to the treatment of Fourier series in Chapters 29 and 30. We deal very briefly with the 'pointwise' theory and its applications. This rests heavily on methods from complex analysis, and is discussed, along with the Laplace transform, in Chapter 9 of [**13**].

It is possible to regard the Fourier transform as a limiting case of a Fourier series, by scaling to smaller and smaller intervals. A discussion of this process can be found in [**4**] and [**11**]. We do not take this route but draw heavily on the parallels between Fourier series and transforms for motivation.

The reason that Fourier series are so useful is that we can capture a function from its Fourier coefficients, in different ways under different assumptions. Analogously we can 'invert' the Fourier transform, to recapture f from its transform $\hat{f}$. We have

- a 'pointwise' inversion theorem, applicable when f is integrable and piecewise differentiable (cf. 29.11);
- an 'L^1-convergence' theorem, applicable to any $f \in \mathrm{L}^1$ (cf. 30.6);
- an 'L^2-convergence' theorem, applicable to any $f \in \mathrm{L}^2$ (cf. 30.11); (here we must decide how the transform of f is to be defined, since $\hat{f}$ as in 33.1 may not make sense).

We discuss these in turn.

33.1 The Fourier transform. Given $f \in \mathrm{L}$, we define the Fourier transform $\hat{f}$ of f by

$$\hat{f}(y) := \int_{-\infty}^{\infty} f(x) \mathrm{e}^{-\mathrm{i}yx}\, \mathrm{d}x \quad (y \in \mathbb{R});$$

observe that the integral exists for each $y \in \mathbb{R}$: the exponential factor is measurable and of modulus 1, so we can invoke 23.2. We often want to combine a general function f with a concrete function such as an exponential or polynomial. To avoid clumsy notation we shall write $x^m f$ for the function g for which $g(x) = x^m f(x)$, and so on.

The transform $\hat{f}$ will be a complex-valued function, even when f is real-valued. Hence we work almost entirely with complex-valued functions in this chapter, and sometimes use notions and results previously given just for the real case. The complex versions come from looking at real and imaginary parts.

Many of the results we give are best interpreted in the normed space L^1, rather than in the class L of integrable functions. All this means in practice is that in the L^1 setting, functions from L which are equal a.e. are regarded as equal; see 28.3.

33.2 Examples. The first three examples below are easily handled using the (extended) FTC, for complex-valued functions.

(1) Let $f(x) = e^{-|x|}$. Then $\hat{f}(y) = 2(1+y^2)^{-1}$ (Exercise 16.2).

(2) Let $f = \chi_{[-1,1]}$. Then

$$\hat{f}(y) = \int_{-1}^{1} e^{-iyx}\,dx = \left[\frac{-1}{iy}e^{-iyx}\right]_{-1}^{1} = \frac{2\sin y}{y}.$$

(3) Let $f(x) = e^{-x}\chi_{[0,\infty)}(x)$. Then

$$\hat{f}(y) = \int_{0}^{\infty} e^{-x-iyx}\,dx = (1+iy)^{-1} = \frac{1}{1+y^2} - i\frac{y}{1+y^2}.$$

(4) Consider $f(x) = (1-|x|) \vee 0$, the *tent function* shown in Fig. 33.1. Then $f \in L$, and a routine computation gives

$$\hat{f}(y) = \int_{-1}^{0} (1+x)\,e^{-ixy}\,dx + \int_{0}^{1} (1-x)\,e^{-ixy}\,dx = \left(\frac{\sin y/2}{y/2}\right)^2.$$

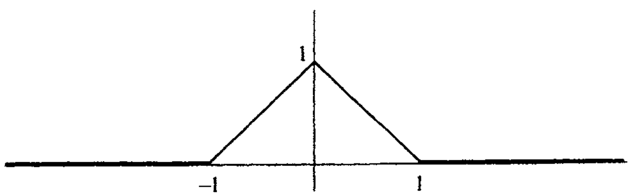


Figure 33.1

(5) Let $f(x) = e^{-\frac{1}{2}x^2}$. We claim that $\hat{f}(y) = \sqrt{2\pi}e^{-\frac{1}{2}y^2}$. We can derive this formula by differentiating under the integral (which is justified by 33.7 below) and then integrating by parts:

$$\begin{aligned}
\frac{d}{dy}\hat{f}(y) &= \int_{-\infty}^{\infty} \frac{\partial}{\partial y}\left(e^{-ixy}e^{-\frac{1}{2}x^2}\right)dx \\
&= -i\int_{-\infty}^{\infty} xe^{-ixy}e^{-\frac{1}{2}x^2}\,dx \\
&= i\left[e^{-ixy}e^{-\frac{1}{2}x^2}\right]_{-\infty}^{\infty} - i\int_{-\infty}^{\infty} ye^{-ixy}e^{-\frac{1}{2}x^2}\,dx \\
&= -y\hat{f}(y).
\end{aligned}$$

Solving this differential equation for $\hat{f}$, we obtain $\hat{f}(y) = Ce^{-\frac{1}{2}y^2}$, where the constant, C, equals $\hat{f}(0)$. This is just $\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx$, namely $\sqrt{2\pi}$.

For an alternative derivation using complex analysis methods see [13],

(6) By (5) (with $(t-u)/\sqrt{\lambda}$ as y) and the substitution $v = \sqrt{\lambda}x$,

$$\sqrt{2\pi}e^{-\frac{1}{2\lambda}(t-u)^2} = \int_{-\infty}^{\infty} \sqrt{\lambda}e^{-\frac{1}{2}x^2}e^{i(t-u)x/\sqrt{\lambda}}\,dx = \int_{-\infty}^{\infty} e^{-\frac{1}{2\lambda}v^2}e^{iuv}e^{-itv}\,dv.$$

This formula, which we use in 33.15, tells us that the transform of $e^{-\frac{1}{2\lambda}x^2}e^{iux}$ takes the value $\sqrt{2\pi}e^{-\frac{1}{2\lambda}(t-u)^2}$ at t (u and λ being treated as constants).

In the manner of the last example, further transforms can be found, or recognized, from the formulæ below. The proofs are elementary.

33.3 New hats from old.

(a) Let $f, g \in \mathrm{L}$, $\lambda \in \mathbb{C}$. Then $(\widehat{f+\lambda g}) = \hat{f} + \lambda\hat{g}$.

(b) Let $f \in \mathrm{L}$ and $a \in \mathbb{R}$. Then

(i) $\widehat{f_a} = e^{iay}\hat{f}$;

(ii) $\widehat{f^a} = \dfrac{1}{a}\hat{f}^{1/a}$ $(a \neq 0)$;

(iii) $(e^{iax}f)\hat{} = \hat{f}_a$;

(iv) $\widehat{f^*} = \overline{\hat{f}}$, where $f^*(x) = \overline{f(-x)}$ and $\overline{z}$ denotes as usual the complex conjugate of the complex number z.

33.4 Exercise example.

(a) Find the Fourier transform of $\chi_{[a,b]}$ $(-\infty < a < b < \infty)$.

(b) Find the Fourier transform of $e^{-\alpha x^2}$ $(\alpha > 0)$.

(c) Let $f(x) = (\sin(x/2)/(x/2))^2$. Use the formula $\sin^2 u = \mathrm{Im}(e^{2iu})$ and Exercise 23.5 to prove that $2\pi\hat{f}(y) = (1 - |y|) \vee 0$.

33.5 Distasteful facts. Consider again the examples in 33.2. In (1), (4), (5), and (6), f and $\hat{f}$ both belong to L. In the other cases $\hat{f}$ is less well behaved. In (2), $\hat{f}$ is an old acquaintance from Chapter 13: it is not integrable, though its improper integral does exist. Example (3) exhibits worse behaviour still. In this case $\hat{f}$ does not even have an improper integral.

Although transforms of functions in L generally do not belong to L, they behave well in other respects.

33.6 Lemma. Let $f \in \mathrm{L}$. Then $\hat{f}$ is continuous and $|\hat{f}(y)| \to 0$ as $|y| \to \infty$.

Proof. Continuity of $\hat{f}$ is immediate from the Continuous DCT, with $|f|$ as the dominating function. The second assertion is a form of the Riemann–Lebesgue Lemma. $\square$

Under stronger conditions the transform behaves even better.

33.7 The transform of a derivative and the derivative of a transform.

(a) Let $f \in \mathrm{L}$. Then $\hat{f}$ is differentiable if $xf \in \mathrm{L}$ and then $(\hat{f})' = -\mathrm{i}\widehat{(xf)}$.

(b) Assume that f and f' belong to L and that f' is continuous. Then $\widehat{f'} = \mathrm{i}x\hat{f}$.

Proof. For (a) we must justify differentiation under the integral sign; see 20.9 for the details.

For (b) we use (extended) integration by parts:

$$\int_{-\infty}^{\infty} f'(t)\mathrm{e}^{-\mathrm{i}tx}\,\mathrm{d}t = \left[f(t)\mathrm{e}^{-\mathrm{i}tx}\right]_{-\infty}^{\infty} - (-\mathrm{i}x)\int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}tx}\,\mathrm{d}t = \mathrm{i}x\hat{f}(x).$$

(In the integrated term $f(t)\mathrm{e}^{-\mathrm{i}tx}$, the exponential factor has modulus 1 while $f(t) \to 0$ as $t \to \pm\infty$, since our assumptions are such that 23.8 applies.) $\square$

33.8 Example: the Hermite functions. In 31.3 we considered the Hermite functions given by

$$h_n(x) := (-1)^n \mathrm{e}^{\frac{1}{2}x^2}\mathrm{D}^n\mathrm{e}^{-x^2} \quad (n \geqslant 0).$$

These lie in $\mathcal{S}$. From the definition,

$$h_n'(x) - xh_n(x) = xh_n(x) + (-1)^n\mathrm{e}^{\frac{1}{2}x^2}\mathrm{D}^{n+1}\mathrm{e}^{-x^2} - xh_n(x) = -h_{n+1}(x).$$

Now take the transform of both sides, and use 33.7:

$$\mathrm{i}x\widehat{h_n}(x) - \mathrm{i}\widehat{h_n} = -\widehat{h_{n+1}}(x).$$

We deduce that both $\{(-\mathrm{i})^n h_n\}$ and $\{\widehat{h_n}\}$ satisfy the same recurrence relation, *viz.* $y_n - \mathrm{i}xy_n = y_{n+1}$. Since $\widehat{h_0} = \sqrt{2\pi}h_0$, we conclude that

$$\widehat{h_n} = \sqrt{2\pi}(-\mathrm{i})^n h_n \quad (n = 0, 1, 2, \dots).$$

We now give our first inversion theorem, for which we present a head-on proof similar in style to that of 29.11. The final formula shows how we should ideally like f and $\hat{f}$ to be be related.

33.9 A simple Fourier inversion theorem. Assume $f \in \mathrm{L}$ is such that f is piecewise differentiable. Then

$$\frac{f(x+)+f(x-)}{2}=\frac{1}{2\pi}\lim_{R\to\infty}\int_{-R}^{R}\hat{f}(y)\mathrm{e}^{\mathrm{i}xy}\,\mathrm{d}y.$$

If further $\hat{f} \in \mathrm{L}$ and f is continuous at x, then

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(y)\mathrm{e}^{\mathrm{i}xy}\,\mathrm{d}y.$$

Proof. We let

$$g(v):=\frac{f(x+v)+f(x-v)-(f(x+)+f(x-))}{2},$$

which reduces to $(f(x+v)+f(x-v)-2f(x))/2$ in case f is continuous at x. By hypothesis, g has a right-hand derivative at 0, so there exist positive constants η, M such that $|g(v)| \leqslant Mv$ for $v \in [0,\eta]$.

We shall make use of the fact that, for any $\delta > 0$,

$$\lim_{R\to\infty}\int_0^{\delta}\frac{\sin Rv}{v}\,\mathrm{d}v=\lim_{R\to\infty}\int_0^{R\delta}\frac{\sin u}{u}\,\mathrm{d}u=\frac{\pi}{2},$$

by Exercise 23.5. Now

$$\begin{aligned}\int_{-R}^{R}\hat{f}(y)\mathrm{e}^{\mathrm{i}xy}\,\mathrm{d}y&=\int_{-R}^{R}\Big\{\int_{-\infty}^{\infty}f(u)\mathrm{e}^{\mathrm{i}(x-u)y}\,\mathrm{d}u\Big\}\,\mathrm{d}y &&\text{(by definition of } \hat{f}\text{)}\\&=\int_{-\infty}^{\infty}\Big\{\int_{-R}^{R}f(u)\mathrm{e}^{\mathrm{i}(x-u)y}\,\mathrm{d}y\Big\}\,\mathrm{d}u &&\text{(by FTT)}\\&=\int_{-\infty}^{\infty}f(u)\Big\{\int_{-R}^{R}\mathrm{e}^{\mathrm{i}(x-u)y}\,\mathrm{d}y\Big\}\,\mathrm{d}u\\&=2\int_{-\infty}^{\infty}f(x+v)\frac{\sin Rv}{v}\,\mathrm{d}v\\&=2\int_{0}^{\infty}(f(x+v)+f(x-v))\frac{\sin Rv}{v}\,\mathrm{d}v.\end{aligned}$$

Therefore

$$\int_{-\infty}^{\infty}\hat{f}(y)\mathrm{e}^{\mathrm{i}xy}\,\mathrm{d}y-\frac{f(x+)+f(x-)}{2}=2\lim_{R\to\infty}\int_0^{\infty}g(v)\frac{\sin Rv}{v}\,\mathrm{d}v.$$

On $[\delta,\infty)$, the function $g(v)/v$ is integrable, and so, by the Riemann–Lebesgue Lemma,

$$\int_{-\infty}^{\infty}\hat{f}(y)\mathrm{e}^{\mathrm{i}xy}\,\mathrm{d}y-\frac{f(x+)+f(x-)}{2}=2\lim_{R\to\infty}\int_0^{\delta}g(v)\frac{\sin Rv}{v}\,\mathrm{d}v.$$

Finally, for $\delta \leqslant \eta$,

$$\left|\int_0^\delta g(v)\frac{\sin Rv}{v}\,\mathrm{d}v\right| \leqslant \int_0^\delta M\,\mathrm{d}v = M\delta,$$

and so can be made arbitrarily small by taking δ sufficiently small, independently of R.

Finally, if $\hat{f} \in \mathrm{L}$ then the DCT allows us to replace $\lim_{R\to\infty}\int_{-R}^{R}$ by $\int_{-\infty}^{\infty}$ (by 16.1). □

33.10 Example. Let $f(x) = \mathrm{e}^{-ax}\chi_{[0,\infty)}(x)$ $(a > 0)$, so f is integrable and piecewise differentiable, and $\hat{f}(y) = 1/(a + \mathrm{i}y)$. Apply the inversion theorem in 33.9:

$$\lim_{R\to\infty}\int_{-R}^{R}\frac{(a-\mathrm{i}y)}{(a^2+y^2)}\mathrm{e}^{\mathrm{i}xy}\,\mathrm{d}y = \begin{cases} 0 & (x<0), \\ \pi & (x=0), \\ 2\pi\mathrm{e}^{-ax} & (x>0). \end{cases}$$

Taking the real part and noting that the integrand is an even function we obtain

$$\lim_{R\to\infty}\int_0^R \frac{a\cos yx + y\sin yx}{a^2+y^2}\,\mathrm{d}y = \pi\mathrm{e}^{-ax} \quad (x>0).$$

We have to be satisfied with an improper integral here: $y\sin yx(a^2+y^2)^{-1}$ is not integrable (see 18.4(3)).

We now turn to the Fourier transform on L, or, by identifying functions equal a.e., on the normed space L^1.

Remember that the product fg of integrable functions f and g in general is not integrable. Although we cannot use the pointwise product on L, we do have the convolution product $f * g$ available, and this behaves admirably under the Fourier transform, as we showed in 26.16.

33.11 Convolution Theorem (recap). Given $f, g \in \mathrm{L}^1$,

$$(f * g)(x) := \int_{-\infty}^{\infty} f(t)g(x-t)\,\mathrm{d}t$$

exists for almost all x and defines an element $f * g$ of L^1, which is such that $(\widehat{f*g}) = \hat{f}\hat{g}$. Further, $*\colon \mathrm{L}^1 \times \mathrm{L}^1 \to \mathrm{L}^1$ is commutative and associative.

33.12 Example. Let $f := \chi_{[-1/2,1/2]}$. Here $\hat{f}(y) = \sin(y/2)/(y/2)$ (cf. 33.2(2)). Hence $\sin^2(y/2)/(y/2)^2$ is the transform of $f * f$. Since $f(t)f(x-t) = 1$ if and only if $-1/2 \leqslant t \leqslant 1/2$ and $-1/2 \leqslant x - t \leqslant 1/2$ we have $f(t)f(x-t) = \chi_{[-1/2,1/2]\cap[x-1/2,x+1/2]}(t)$. Therefore $(f * f)(x) = 0$ for $x \notin [-1, 1]$ because the intervals do not overlap and $(f * f)(x) = (1 - |x|)$ otherwise. The Convolution Theorem then gives another derivation of the formula for the transform of the tent function $(1 - |x|) \vee 0$.

Those whose interest in primarily in the Fourier transform as a practical tool may wish to skip to 33.17, taking on trust the useful uniqueness theorem stated there. The next section should be compared with the discussion of the Fejér kernel in 30.3; for an even closer analogy see 33.16.

33.13 Exploiting convolutions. The idea behind our inversion procedure for the transform on L^1 is simple. We know that $\hat{f} \notin \mathrm{L}^1$ in general. However we do know that $\hat{f} \in \mathrm{C}_0(\mathbb{R})$, which implies that it is bounded and measurable. Hence $\hat{f}(y)\mathrm{e}^{-y^2/(2\lambda)} \in \mathrm{L}^1$ for any $\lambda > 0$; the exponential factor fades out to leave $\hat{f}(y)$ as $\lambda \to \infty$. Define

$$k_\lambda(x) := \sqrt{\frac{\lambda}{2\pi}}\mathrm{e}^{-\frac{1}{2\lambda}x^2} \quad (\lambda > 0).$$

For each λ, k_λ is a rapidly decaying non-negative integrable function scaled so that $\int k_\lambda = 1$. Then, for any $\delta > 0$,

$$\begin{aligned} |(k_\lambda * f)(x) - f(x)| &= \left|\int k_\lambda(t)\,(f(x-t) - f(x))\,\mathrm{d}t\right| \\ &\leqslant \int_{|t|<\delta} k_\lambda(t)\,|f(x-t) - f(x)|\,\mathrm{d}t + \int_{|t|\geqslant\delta} k_\lambda(t)\,|f(x-t) - f(x)|\,\mathrm{d}t. \end{aligned}$$

We should like $k_\lambda * f$ to be a good approximation to f. Pointwise there is no reason to expect this, since f may not be continuous. However things work well for the L^1 norm, thanks to 23.5, which says that $\|f_a - f\|_1 \to 0$ as $a \to 0$. We shall prove, by a technical argument which draws on many earlier results, that $\|k_\lambda * f - f\|_1 \to 0$ as $\lambda \to \infty$.

Fix $\varepsilon > 0$ and choose $\delta > 0$ so that $\int |f(x-t) - f(x)|\,\mathrm{d}x < \varepsilon$ for $|t| < \delta$. Now, from above,

$$\begin{aligned} \|k_\lambda * f - f\|_1 &\leqslant \int \left\{\int_{|t|<\delta} k_\lambda(t)\,|f(x-t) - f(x)|\,\mathrm{d}t\right\} \mathrm{d}x \\ &\quad + \int \left\{\int_{|t|\geqslant\delta} k_\lambda(t)\,|f(x-t) - f(x)|\,\mathrm{d}t\right\} \mathrm{d}x \\ &= \int_{|t|<\delta} \left\{\int k_\lambda(t)\,|f(x-t) - f(x)|\,\mathrm{d}x\right\} \mathrm{d}t \\ &\quad + \int_{|t|\geqslant\delta} \left\{\int k_\lambda(t)\,|f(x-t) - f(x)|\,\mathrm{d}x\right\} \mathrm{d}t \quad \text{(by FTT)} \\ &\leqslant \int_{|t|<\delta} k_\lambda(t) \left\{\int |f(x+t) - f(x)|\,\mathrm{d}x\right\} \mathrm{d}t \\ &\quad + 2\|f\|_1 \int_{|t|\geqslant\delta} \sqrt{\lambda/(2\pi)}\mathrm{e}^{-\frac{1}{2\lambda}t^2}\,\mathrm{d}t \qquad \text{(by (M) and (T))} \end{aligned}$$

which can be made less than 2ε by taking λ sufficiently large, by 16.1.

We now have the following theorem. By showing how $f \in \mathrm{L}^1$ is determined by $\hat{f}$, it provides an inversion theorem for the Fourier transform on L^1.

33.14 Theorem. Let $f \in \mathrm{L}^1$. Then

$$\lim_{\lambda\to\infty} \|k_\lambda * f - f\|_1 = 0, \quad \text{where } k_\lambda(x) := \sqrt{\frac{\lambda}{2\pi}}\mathrm{e}^{-\frac{1}{2\lambda}x^2} \quad (\lambda > 0).$$

Further, $2\pi(k_\lambda * f)(t) = \int \hat{f}(y)\mathrm{e}^{-\frac{1}{2\lambda}y^2}\mathrm{e}^{\mathrm{i}yt}\,\mathrm{d}y$.

Proof. Only the final statement remains to be proved. Here is the calculation:

$$\begin{aligned}
\frac{1}{2\pi}\int \hat{f}(y)\mathrm{e}^{-\frac{1}{2\lambda}y^2}\mathrm{e}^{\mathrm{i}yt}\,\mathrm{d}y &= \frac{1}{2\pi}\int \left\{\int f(x)\mathrm{e}^{-\frac{1}{2\lambda}y^2}\mathrm{e}^{-\mathrm{i}yx}\mathrm{e}^{\mathrm{i}yt}\,\mathrm{d}x\right\}\mathrm{d}y \\
&= \frac{1}{2\pi}\int \left\{\int f(x)\mathrm{e}^{-\frac{1}{2\lambda}y^2}\mathrm{e}^{-\mathrm{i}yx}\mathrm{e}^{\mathrm{i}yt}\,\mathrm{d}y\right\}\mathrm{d}x \quad &&\text{(by FTT)} \\
&= \frac{1}{2\pi}\int f(x)\left\{\int \mathrm{e}^{-\frac{1}{2\lambda}y^2}\mathrm{e}^{-\mathrm{i}yx}\mathrm{e}^{\mathrm{i}yt}\,\mathrm{d}y\right\}\mathrm{d}x \\
&= \sqrt{\lambda/(2\pi)}\int f(x)\mathrm{e}^{-\frac{1}{2\lambda}(t-x)^2}\,\mathrm{d}x &&\text{(from 33.2(6))} \\
&= (f * k_\lambda)(x) &&\text{(as } k_\lambda \text{ is even),} \\
&= (k_\lambda * f)(x).
\end{aligned}$$

as required. □

33.15 Summability kernels. In the arguments above we used the following properties of $\{k_\lambda\}$: for each λ,

(K1) k_λ is a non-negative integrable function,

(K2) $\int k_\lambda = 1$,

(K3) for each $\delta > 0$, $\int_{|x|>\delta} k_\lambda(x)\,\mathrm{d}x \to 0$ as $\lambda \to \infty$.

Any family $\{k_\lambda\}_{\lambda>0}$ satisfying (K1)–(K3) is said to be a *summability kernel* or an *approximate identity* for L^1. An analysis of the L^1 Fourier transform parallel to that in 33.14 can be carried through using any summability kernel.

33.16 Exercise example. Let $T(x) = (1 - |x|) \vee 0$ (as in 33.2) and, for $\lambda > 0$, let

$$k_\lambda(x) := \lambda g(\lambda x) \quad \text{where } g(x) := \frac{1}{2\pi}\left(\frac{\sin\frac{1}{2}x}{\frac{1}{2}x}\right)^2,$$

so that $\hat{T} = 2\pi g$.

(a) Prove that
$$\frac{1}{2\pi}\int_{-\lambda}^{\lambda}\Big(1-\frac{|t|}{\lambda}\Big)\hat{f}(t)e^{itx}\,dt=(k_\lambda * f)(x).$$

(b) Check that $\{k_\lambda\}$ is a summability kernel. Deduce that $\|k_\lambda * f - f\|_1 \to 0$ as $\lambda \to \infty$.

The theorem below is immediate from 33.14.

33.17 Uniqueness theorem for the L^1 Fourier transform. Let $f, g \in L$ be such that $\hat{f} = \hat{g}$. Then $f = g$ a.e.

As an application of this we derive a result which is worthwhile in its own right and which underpins our treatment of the L^2 Fourier transform later on.

33.18 Completeness of the normalized Hermite functions. We prove, as we claimed in Chapter 32, that $\{z_n\}$, where $z_n := h_n/\|h_n\|_2$, is an orthonormal basis for L^2. To this end we assume that $f \in L^2$ is such that $\langle f, z_n\rangle = 0$ for all $n \geqslant 0$. Since $z_n(x)$ is a polynomial of degree n multiplied by $e^{-x^2/2}$ we have
$$0=\int_{-\infty}^{\infty} f(x)x^k e^{-x^2/2}\,dx \quad (k=0,1,2,\dots).$$

Consider the Fourier transform of the L^1 function $g(x) := f(x)e^{-x^2/2}$:
$$\begin{aligned}\hat{g}(y)=\int_{-\infty}^{\infty} f(x)e^{-x^2/2}e^{-iyx}\,dx &= \int_{-\infty}^{\infty} f(x)e^{-x^2/2}\sum_{k=1}^{\infty}\frac{1}{k!}(-iyx)^k\,dx\\ &=\sum_{k=1}^{\infty}\frac{1}{k!}(-iy)^k\int_{-\infty}^{\infty} f(x)x^k e^{-x^2/2}\,dx=0,\end{aligned}$$

provided we can justify the interchange of $\sum$ and $\int$. For this we call on 17.5, with $u_k(x) = f(x)e^{-x^2/2}(-iyx)^k/k!$. Then
$$\sum\int |u_k| = \sum \frac{|y|^k}{k!}\int_{-\infty}^{\infty}\Big|f(x)x^k\Big|e^{-x^2/2}\,dx,$$

which is finite (use the Cauchy–Schwarz inequality). So $\hat{g}(y) = 0$ for all y. By 33.17, $\hat{g} = 0$ implies $g = 0$ a.e., so $f = 0$ in L^2. We have verified condition (ONB5) for completeness.

Theorem 33.14 told us about the Fourier transform of any function f for which $\hat{f}$ is defined. A different approach is to explore the Fourier transform on a set of very well-behaved functions which is dense in L^1 (and in L^2). We shall see that stronger smoothness assumptions on f or $\hat{f}$ further improve the behaviour of the transform.

Iterating 33.7, we obtain the following corollary. It reveals that there is a nice interplay between the rate of decay of $\hat{f}$ at $\pm\infty$ and the smoothness of f, and also between the smoothness of $\hat{f}$ and the rate of decay of f at $\pm\infty$. Obviously the faster $f(x) \to 0$ as $x \to \pm\infty$, the better the chance that $x^m f(x) \to 0$ too. We shall use the maximal such m as our measure of rate of decay of f. We naturally take the maximum n such that nth derivative $\mathrm{D}^n f$ exists as our measure of the smoothness of f; here by convention $\mathrm{D}^0 f = f$.

33.19 Smoothness vs. decay.

(a) Assume that f is such that $x^k f \in \mathrm{L}$ for $k = 0, \ldots, m$. Then $\mathrm{D}^m \hat{f}$ exists and
$$\mathrm{D}^m \hat{f} = (-\mathrm{i})^m (\widehat{x^m f}).$$

(b) Assume that, for each $k = 0, \ldots, m$, $\mathrm{D}^k f$ exists, is continuous and belongs to L. Then
$$\widehat{\mathrm{D}^m f} = (\mathrm{i}x)^m \mathrm{D}^m \hat{f}.$$

The foregoing results suggest that it may be worthwhile to investigate functions which are as well-behaved as possible as regards both smoothness and decay.

33.20 Schwartz space. We define Schwartz space, $\mathcal{S}$, to consist of those functions $f\colon \mathbb{R} \to \mathbb{C}$ such that

(i) $f \in \mathrm{C}^\infty(\mathbb{R})$, that is, the derivatives $\mathrm{D}^m f$ exist for all $m \geqslant 1$, and

(ii) $\lim_{x\to\pm\infty} x^n \mathrm{D}^m f(x) = 0$ for all $m, n \geqslant 0$.

The space $\mathcal{S}$ has good closure properties:

(a) $f, g \in \mathcal{S}$ and $\lambda \in \mathbb{C}$ imply $f + \lambda g, fg \in \mathcal{S}$,

(b) $f \in \mathcal{S}$ implies $x^n \mathrm{D}^m f \in \mathcal{S}$ for any $m, n \geqslant 0$ (proof by induction).

Every member of $\mathcal{S}$ is continuous and bounded on $\mathbb{R}$, and dominated by the integrable function x^{-2} near $\pm\infty$. The Comparison Theorem yields that $f \in \mathcal{S}$ implies that f is integrable. We deduce from 33.7 that $f \in \mathcal{S}$ implies that $\hat{f}$ exists and belongs to $\mathcal{S}$, because
$$x^n \mathrm{D}^m \hat{f} = (-1)^m \mathrm{i}^{m+n} \mathrm{D}^n (\widehat{x^m f}).$$

Compare this with the unsatisfactory behaviour of $\hat{}$ on the bigger space L^1 to which we drew attention in 33.5.

Any trigonometric function, polynomial, or exponential of a polynomial is infinitely differentiable. For any polynomial q we have $\lim_{x\to\pm\infty} q(x)\mathrm{e}^{-x^2} = 0$. It follows easily that any function of the form $p(x)\mathrm{e}^{-\alpha x^2 + \mathrm{i}\beta x}$ (p a polynomial, $\alpha, \beta \in \mathbb{R}$, $\alpha > 0$) belongs to $\mathcal{S}$.

Any function in $C^\infty(\mathbb{R})$ which is zero outside a bounded interval will lie in $\mathcal{S}$. However we cannot say that $f \in C^\infty(\mathbb{R})$ implies $f\chi_{[a,b]} \in \mathcal{S}$ always, because the latter function may not be continuous at a and b.

We have previously used $e^{-|x|}$ as an exponential damping function alternative to e^{-x^2}. This is debarred to us if we wish to work in $\mathcal{S}$ because $e^{-|x|}$ is not differentiable at 0 and so not in $\mathcal{S}$.

Our simple inversion theorem, 33.9 looks very tidy when restricted to $\mathcal{S}$, especially when we share out the factor of 2π by defining $\mathcal{F}f = 1/\sqrt{2\pi}f$.

33.21 Fourier inversion theorem for Schwartz space. Define $\mathcal{F}\colon \mathcal{S} \to \mathcal{S}$ by $\mathcal{F}f := \hat{f}/\sqrt{2\pi}$. Then f and $\mathcal{F}f$ are related by

$$(\mathcal{F}f)(y) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x)e^{-iyx}\,dx \quad \text{and} \quad f(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} (\mathcal{F}f)(y)e^{iyx}\,dy,$$

that is, $2\pi(\mathcal{F}^2 f)(x) = f(-x)$ for all x. Moreover, $f = (2\pi)^2(\mathcal{F}^4 f)$, so that f is uniquely determined by $\mathcal{F}f$.

33.22 Density of $\mathcal{S}$ in L^1 and L^2. $\mathcal{S}$ sits as a subspace inside each of L^1 and L^2, and more generally inside any L^p ($p \geqslant 1$). Because C^∞ functions of compact support are dense in either of these spaces, so is $\mathcal{S}$. A different proof that $\mathcal{S}$ is dense in L^2 comes from the fact that the normalized Hermite functions, which lie in $\mathcal{S}$, are a complete orthonormal set in L^2.

The last remark leads immediately to a viable approach to the L^2 Fourier transform.

33.23 The Fourier transform on L^2. Let $\{u_n\}_{n\geqslant 0}$ be the complete orthonormal sequence of normalized Hermite functions:

$$u_n(x) := (-1)^n e^{x^2/2} D(x^n e^{-x^2})/(2^n n!\sqrt{\pi})^{1/2}$$

(recall 31.3 and 33.18). From 33.8 we know that $\mathcal{F}u_n = (-i)^n u_n$. For $f \in L^2$ we can write

$$f = \sum_{n=0}^{\infty} \langle f, u_n\rangle u_n.$$

As we are working in L^2 this means $\|f - \sum_{n=0}^{N}\langle f, u_n\rangle u_n\|_2 \to 0$ as $N \to \infty$. Now define

$$\mathcal{F}f := \sum_{n=0}^{\infty} (-i)^n \langle f, u_n\rangle u_n.$$

(We shall shortly reconcile this with our previous usage of $\mathcal{F}$ as the rescaled $\hat{}$ on $\mathcal{S}$.)

We claim:

($\mathcal{F}1$) $\mathcal{F}f \in \mathrm{L}^2$, for all $f \in \mathrm{L}^2$;

($\mathcal{F}2$) for all $f \in \mathrm{L}^2$, $\mathcal{F}^2 f = g$, where $g(x) = \overline{f(-x)}$, and $\mathcal{F}^4 f = f$.

($\mathcal{F}3$) $\|\mathcal{F}f\|_2 = \|f\|_2$ for all $f \in \mathrm{L}^2$;

($\mathcal{F}4$) $\int f\overline{g} = \int \mathcal{F}f\overline{\mathcal{F}g}$, for all $f, g \in \mathrm{L}^2$ (the **Plancherel formula**);

($\mathcal{F}5$) the map $\mathcal{F}\colon \mathrm{L}^2 \to \mathrm{L}^2$ is linear, one-to-one and onto, and such that

$$\mathcal{F}(\lim f_m) = \lim \mathcal{F}f_m$$

whenever $\{f_m\}$ is a convergent sequence in L^2.

Use 28.16 for ($\mathcal{F}1$); ($\mathcal{F}2$) is then immediate. ($\mathcal{F}3$) and ($\mathcal{F}4$) come from (ONB3) and (ONB4). Linearity is straightforward. The fact that $\mathcal{F}$ is one-to-one and onto follow from ($\mathcal{F}2$). For the final assertion in ($\mathcal{F}5$)—continuity of $\mathcal{F}$—note that

$$\|\mathcal{F}f - \mathcal{F}f_m\|_2 = \|\mathcal{F}(f - f_m)\|_2 = \|f - f_m\|_2, \quad \text{where } f := \lim f_m.$$

[For the cognoscenti: $\mathcal{F}$ is a unitary operator on L^2; its eigenvalues are $\pm 1, \pm \mathrm{i}$.]

Finally we reconcile our L^1 transform, $\hat{\ }$, and our L^2 transform, $\mathcal{F}$, by showing that $1/\sqrt{2\pi}\hat{f} = \mathcal{F}f$ for $f \in \mathrm{L}^1 \cap \mathrm{L}^2$, the largest space on which both sides are defined. The equation is true by construction on the Hermite functions, and hence also by linearity on all finite linear combinations of the Hermite functions. By taking limits we see that the equality extends to $\mathrm{L}^1 \cap \mathrm{L}^2$, in which these linear combinations are dense; see 32.7. In particular the L^2 transform of $f \in \mathcal{S}$ is $1/\sqrt{2\pi}\hat{f}$, so that our premature introduction of the symbol $\mathcal{F}$ in 33.23 is vindicated.

Exercises

33.1 Let $f, g \in \mathrm{L}^1$.

(a) Prove that $f\hat{g}$ and $\hat{f}g$ belong to L^1.

(b) Use FTT to prove that $\int f\hat{g} = \int \hat{f}g$ (the **multiplication formula**).

[The calculation in 33.14 of $k_\lambda * f$ in terms of $\hat{f}$ is an instance of this result. Compare the multiplication formula with the Plancherel formula in 33.23. There the transforms are on the same side of the equation.]

33.2 Let $f \in \mathrm{L}$ and let $g(x) = \mathrm{e}^{-x^2}$. Assume that $f * f = g * g$. Show that $f = g$ a.e. or $f = -g$ a.e.

33.3 Prove that $\{g \in \mathrm{L}^1 \mid \hat{g} \text{ has compact support}\}$ is dense in L^1. [Hint: use Theorem 33.14.]

33.4 Let $f \in \mathrm{L}$ and let F be its indefinite integral. For $a < b$ and $R > 0$, let

$$I(R) := \frac{1}{2\pi}\int_{-R}^{R} \frac{\mathrm{e}^{-\mathrm{i}ay} - \mathrm{e}^{-\mathrm{i}by}}{-\mathrm{i}y}\hat{f}(y)\,\mathrm{d}y$$

(a) Use FTT to prove that

$$I(R) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\Big\{\int_{-R}^{R} \frac{e^{i(x-a)y} - e^{i(x-b)y}}{-iy}\,dy\Big\}\,dx.$$

(b) Use the fact that $\int_0^t x^{-1}\sin x\,dx \to \pi/2$ as $t \to \infty$ to show that

$$\lim_{R\to\infty}\int_{-R}^{R} \frac{e^{i(x-a)y} - e^{i(x-b)y}}{iy}\,dy = \begin{cases} 0 & (x < a \text{ or } x > b),\\ \pi & (x = a \text{ or } x = b),\\ 2\pi & (a < x < b).\end{cases}$$

(c) Use the Continuous DCT to prove that

$$F(b) - F(a) = \frac{1}{2\pi}\int_{-R}^{R} \frac{e^{-iby} - e^{-iay}}{iy}\hat{f}(y)\,dy.$$

(d) Deduce that F is uniquely determined by $\hat{f}$ [and hence that f is determined a.e. by $\hat{f}$, by 24.7].

33.5 Lemma 33.6 shows that $\hat{}$ maps L^1 into $C_0(\mathbb{R})$, meaning the continuous functions $g\colon \mathbb{R} \to \mathbb{C}$ such that $|g(y)| \to 0$ as $|y| \to \infty$. This exercise shows that $\hat{}$ does not map L^1 **onto** $C_0(\mathbb{R})$. [It can be shown that the image is dense for the supremum norm on $C_0(\mathbb{R})$; the proof uses the Weierstrass–Stone Theorem, a generalization of Weierstrass' approximation theorem.] Define g by

$$g(x) = \begin{cases} \dfrac{1}{\log x} & (x \geqslant e),\\ x/e & (0 \leqslant x < e),\\ -g(-x) & (x > 0).\end{cases}$$

Assume for a contradiction that $g = \hat{f}$ where $f \in L$.

(a) Check that $g \in C_0(\mathbb{R})$.

(b) Show that $\int_e^n g(y)/y\,dy \to \infty$ as $n \to \infty$ (so that $g(x)/x$ is not Lebesgue integrable on $[e, \infty)$).

(c) Show that

$$g(y) = \int_0^{\infty} f_1(x)\sin xy\,dx \quad \text{where } f_1(x) := (f(x) - f(-x))/2.$$

(d) Use FTT to show that

$$\int_e^n \frac{g(y)}{y}\,dy = \int_0^{\infty}\Big\{\int_{ex}^{nx} \frac{\sin u}{u}\,du\Big\} f_1(x)\,dx.$$

Use the DCT to show that the right-hand side tends to a finite limit as $n \to \infty$.

33.6 Let $f \in \mathcal{S}$ and let α, β be positive constants with $\alpha\beta = 2\pi$. Prove that

$$\alpha \sum_{k=-\infty}^{\infty} f(\alpha k) = \sum_{k=-\infty}^{\infty} \hat{f}(\beta k)$$

and hence show that

$$\sqrt{\alpha}\Big(\frac{1}{2} + \sum_{k=1}^{\infty} \mathrm{e}^{-\frac{1}{2}\alpha^2 k^2}\Big) = \sqrt{\beta}\Big(\frac{1}{2} + \sum_{k=1}^{\infty} \mathrm{e}^{-\frac{1}{2}\beta^2 k^2}\Big).$$

33.7 (a) Use the Plancherel formula to compute

$$\int_0^\infty \frac{\sin^2 x}{x^2(1+x^2)}\, \mathrm{d}x.$$

(b) Let $f \in \mathrm{L}^2$ and assume that

$$\int_{-\infty}^{\infty} \mathrm{e}^{-|x-y|} f(y)\, \mathrm{d}x = 0 \quad (x \in \mathbb{R}).$$

Prove that $f(y) = 0$ for almost all y.

33.8 (a) Let $f, g \in \mathrm{L}^2$. Prove that

$$(f * g)(x) := \int_{-\infty}^{\infty} f(t) g(x-t)\, \mathrm{d}t$$

exists for almost all $x \in \mathbb{R}$. Prove also that $f * g = \hat{k}$ for some $k \in \mathrm{L}^1$.

(b) Prove that if $k \in \mathrm{L}^1$ then there exists $f, g \in \mathrm{L}^2$ such that $\hat{k} = f * g$.

(c) Compute k in the case that $f(t) = g(t) = (\sin t)/t$ $(t \in \mathbb{R})$.

33.9 Let $f \in \mathrm{L}^2$ and let

$$F(x) = \int_{-\infty}^{\infty} f(x-t) \frac{\sin t}{t}\, \mathrm{d}t.$$

Show that this defines a function $F \in \mathrm{L}^2$ and that $\|F\|_2 \leqslant \pi \|f\|_2$.

33.10 [This exercise gives an alternative approach to the Fourier transform on L^2.]

(a) For any bounded interval I define $\mathcal{F}\chi_I = \hat{\chi}_I / \sqrt{2\pi}$. Prove that

$$\|\mathcal{F}\chi_I\|_2 = \|\chi_I\|_2 = \sqrt{\ell(I)}.$$

[Use Exercise 23.5(d).]

(b) Prove that if I and J are disjoint bounded intervals then

$$\langle \mathcal{F}\chi_I, \mathcal{F}\chi_J \rangle = \langle \chi_I, \chi_J \rangle = 0.$$

[Hint: consider $\chi_I \pm \chi_J$.]

(c) Show that $\mathcal{F}$ may be extended to L^{step} so that $\|\mathcal{F}\varphi\|_2 = \|\varphi\|_2$ for all $\varphi \in L^{\text{step}}$.

(d) Show that $\mathcal{F}$ extends to a norm-preserving linear map from L^2 to L^2.

34 Integration in probability theory

No account of integration theory would be complete without some discussion of its role in probability theory. This is a very large subject and we seek to do no more in this chapter than to build a bridge to specialized texts, such as **[10]**, which we recommend for further reading. We address only readers with some prior familiarity with probabilistic ideas, at the level, for example, of **[7]**—space precludes us introducing with proper motivation the concepts of random variable, independence, and so on. Like Chapter 24 this is a spectators' chapter, for which we set no exercises.

34.1 Measures in general. In Chapter 22 we introduced the (Lebesgue) measure, m. This extends the notion of length from intervals to a much richer class of subsets of $\mathbb{R}$ —the (Lebesgue) measurable sets. This class $\mathcal{M}$ is a *σ-algebra* of subsets of $\mathbb{R}$: it contains $\varnothing$ and $\mathbb{R}$, is closed under countable unions, countable intersections, and complements. On $\mathcal{M}$, m has the properties given in 22.5, and is in particular non-negative and countably additive. It is also *complete*, in the sense that any subset of a measurable set of measure zero is measurable (and necessarily itself of measure zero).

The class of Borel sets, $\mathcal{B}$, is the smallest σ-algebra of subsets of $\mathbb{R}$ containing the open intervals, or equivalently containing all intervals. It is properly contained in $\mathcal{M}$ (see Exercise 22.3). The difference in $\mathcal{B}$ and $\mathcal{M}$ comes solely through null sets (see Exercise 22.10). In technical terms $\mathcal{M}$ is the smallest complete σ-algebra containing $\mathcal{B}$, that is, the *completion* of $\mathcal{B}$. The measure m on $\mathcal{B}$ is called Borel measure, and a function $f\colon \mathbb{R} \to \mathbb{R}$ is Borel measurable if $f^{-1}(A)$ is a Borel set for each open set (or, equivalently, each Borel set) A in $\mathbb{R}$); cf. Exercise 22.4.

We can abstract the above. Take a σ-algebra $\mathcal{X}$ of subsets of a set S (less restrictive assumptions are possible here, but we wish to avoid a proliferation of definitions). A *measure* is defined to be a countably additive function $\mu\colon \mathcal{E} \to \mathbb{R}^+ \cup \{\infty\}$. We then say $(S, \mathcal{X}, \mu)$ is a *measure space*. In this context a function $f\colon S \to \mathbb{R}$ is defined to be measurable if $f^{-1}(A) \in \mathcal{X}$ for every open set A.

Working in the opposite direction from the one we took, a theory of integration can be developed in which $(\mathbb{R}, \mathcal{M}, m)$ is replaced by a measure space $(X, \mathcal{E}, \mu)$ and step functions are replaced, as they may be in the Lebesgue theory too, by *simple functions*. These are of the form $\sum_{r=1}^{n} c_r \chi_{A_r}$, where $c_1, \dots, c_n \in \mathbb{R}$ and $A_1, \dots, A_n$ are disjoint sets in $\mathcal{E}$ of finite measure. Exploring these ideas, and relaxing the assumptions to a minimum, is a substantial undertaking. It warrants a book to itself (and many such books have been written), and is different in flavour from the Lebesgue theory as we have presented it.

Here are some examples of measures.

(1) Lebesgue measure on $\mathbb{R}^2$ is defined on the measurable subsets of $\mathbb{R}^2$ by setting $m(S) = \int \chi_S(x,y)\,\mathrm{d}(x,y)$, as expected. As in the 1-dimensional case, the Lebesgue measurable sets can be obtained as the completion of the class of Borel measurable subsets, namely the σ-algebra generated by the 2-dimensional (open) intervals.

(2) Let S be a finite or countably infinite set with elements $x_1, x_2, \ldots$, let $\mathcal{X}$ be the family of all subsets of X, and define $\mu(A) := \sum_k \chi_A(x_k)$. Then $(S, \mathcal{X}, \mu)$ is a measure space, in which μ simply counts the number of points in A, and is accordingly called counting measure.

(3) Let α be an increasing function on $\mathbb{R}$ which is right-continuous, that is, $\alpha(x) = \alpha(x+)$ for every x. Define the associated *Stieltjes measure*, μ_α, as follows. For $-\infty < a < b < \infty$ let

$$\mu_\alpha((a,b]) := \alpha(b) - \alpha(a) \quad \text{and} \quad \mu_\alpha(\{a\}) := \alpha(a) - \alpha(a-).$$

Since an arbitrary bounded interval can be expressed as a countable disjoint union of intervals of the above types, it is not too difficult to show that μ_α extends to a well-defined measure on $\mathcal{B}$. As one special case we can take $\alpha(x) = x$, when μ_α is just the Lebesgue measure m, restricted to $\mathcal{B}$. We can also recognize the counting measure in (2) as having the form μ_α, by choosing α to have a jump discontinuity at each natural number and to be constant between discontinuities.

Of principal interest to us here are those measure spaces $(S, \mathcal{X}, \mu)$ in which $\mathcal{X}$ is a σ-algebra, and the measure μ is a *probability measure*, so that $\mu(X) = 1$, and $\mu(X \setminus A) = 1 - \mu(A)$.

34.2 Probability spaces and random variables. Suppose we wish to analyse, by observation or experiment, some quantity exhibiting non-deterministic behaviour, such as the result of throwing a die, the length of a supermarket queue, or the lifetime of an electric light bulb.

The starting point for the mathematical modelling of such a situation is a probability space. Such a space has three components:

- a set Ω (to be viewed as the possible results of measurements of some quantity exhibiting random behaviour),
- a complete σ-algebra $\mathcal{E}$ of subsets of Ω, called *events*, and
- a probability measure $\mathbb{P}\colon \mathcal{E} \to [0,1]$, where $\mathbb{P}(A)$ should represent the likelihood of an outcome lying in the set A.

The assumption that $\mathcal{E}$ is complete allows us to talk about events which occur with probability zero without having to worry whether these really are events. Taking complements, any set U such that $V \subseteq U \subseteq \Omega$, where $V \in \mathcal{E}$ and $\mathbb{P}(V) = 1$, is itself an event, which occurs *almost surely* (a.s.).

In certain cases the set of possible outcomes is finite, as with the throwing of a die. In others, the quantity being measured may take values in an interval of $\mathbb{R}$, as when the observation is the time spent in a supermarket queue. Either way, the quantity to be measured is modelled by a *random variable*. By this we mean a measurable function $X\colon \Omega \to \mathbb{R}$. [To avoid technical distractions, we are considering only random variables which are 'surely finite'; in general one allows $X\colon \Omega \to \mathbb{R} \cup \{\pm\infty\}$.] For any Borel set A in $\mathbb{R}$, the set $\{\omega \in \Omega \mid X(\omega) \in A\}$ belongs to $\mathcal{E}$ and we interpret $\mathbb{P}(\{x \in \Omega \mid X(x) \in A\})$ (henceforth abbreviated to $\mathbb{P}(X \in A)$) as the probability that the observed value of X lies in A.

The modelling of non-deterministic phenomena requires the notion of independence, of a pair (or more generally a family) of events or of random variables. This stems from the necessity to capture the idea of a sequence of independently performed measurements sampling the value of some random quantity.

Given a sequence of events $\{E_n\}$ it is often of interest to know how many of them occur, and in particular the probability of the event $E := \bigcap_n \bigcup_{k\geqslant n} E_k$ that infinitely many of them occur. Consider the Borel–Cantelli Lemmas which were presented as a purely measure-theoretic exercise in Exercise 22.8. Regard the sets $\{E_k\}$ specified there as a sequence of events in a probability space, and note that the proof of (b) requires the independence of $\{E_n\}$ rather than the stronger assumption of pairwise disjointness. We can now interpret the Borel–Cantelli Lemmas: if $\sum \mathbb{P}(E_k)$ converges, then E has probability 0, whereas if $\sum \mathbb{P}(E_k)$ diverges then $\mathbb{P}(E) = 1$ provided the events $\{E_k\}$ are independent. This dichotomy is an instance of a zero–one law. This example illustrates that some of the technical apparatus of probability theory is measure-theoretic. However the mathematical theory of probability is not just measure theory on a space of total measure 1. The reason is that the ramifications of the concept of independence extend well beyond pure measure theory.

34.3 Distribution functions and density functions. Let X be a random variable on a fixed probability space $(\Omega, \mathcal{E}, \mathbb{P})$. The *(cumulative) distribution function* of X is defined to be F (written F_X when X needs to be explicit), where

$$F(x) := \mathbb{P}(X \leqslant x).$$

Thus $F(x)$ is the probability assigned to $\{\omega \in \Omega \mid X(\omega) \leqslant x\}$. It follows easily from properties of $\mathbb{P}$ that $F\colon \mathbb{R} \to [0,1]$ is an increasing function which is right-continuous. Further, $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

At this point in elementary accounts of probability two different types of random variable are identified, and thereafter handled in parallel but separate ways.

First, a random variable X is said to be *discrete* if its values lie in a finite or countable set. For example, X might have a binomial distribution $B(1,p)$ $(0 < p < 1)$ or a Poisson distribution with parameter $\lambda > 0$. In the former case X takes values in $\{0,1\}$, with $\mathbb{P}(X = 0) = p$ and $\mathbb{P}(X = 1) = 1 - p$. In the latter,

$$\mathbb{P}(X = k) = \frac{1}{k!}\lambda^k \mathrm{e}^{-\lambda} \quad (k = 0, 1, 2, \dots).$$

Let X be a discrete random variable whose values are $a_1 < a_2 < \dots$. Then

$$F(x) = \sum_{a_k \leqslant x} \mathbb{P}(X = a_k),$$

so that F is constant in each interval $[a_k, a_{k+1})$, and has a jump discontinuity at each a_k.

The second class of random variables of major importance consists of those which are dubbed continuous. A *(probability) density function* is a non-negative (Lebesgue) integrable function f such that $\int_{\mathbb{R}} f = 1$. The random variable X with distribution function F is *continuous* if there is a density function f (also written f_X) such that

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t;$$

thus F is just the indefinite integral of f. Every such distribution function is continuous (see 19.4), in fact absolutely continuous (see 24.10). As examples of density functions we have the following.

(a) The most important distribution in probability and statistics is undoubtedly the normal distribution, N(0,1), on $\mathbb{R}$. Its density function is $\mathrm{e}^{-\frac{1}{2}x^2}/\sqrt{2\pi}$. The associated distribution function cannot be specified in a convenient closed form.

(b) On $\mathbb{R}^+$, the exponential distribution with parameter $\lambda > 0$ has density function $\lambda \mathrm{e}^{-\lambda x}$.

Let X be a continuous random variable with density function f. Certainly for $-\infty \leqslant a \leqslant b \leqslant \infty$,

$$\mathbb{P}(a < X \leqslant b) = \int_a^b f(t)\,\mathrm{d}t.$$

Further, by Exercise 22.5 and the MCT,

$$\mathbb{P}(X = b) = \mathbb{P}(\bigcap\,(b - n^{-1} < X \leqslant b)) = \lim_{n\to\infty} \int_{b-\frac{1}{n}}^{b} f(t)\,\mathrm{d}t = 0.$$

It follows that, for any interval $\langle a, b\rangle$ with endpoints a, b,

$$\mathbb{P}(X \in \langle a, b\rangle) = \int_a^b f(t)\,\mathrm{d}t.$$

More generally it can be shown that, for any Borel set B,

$$\mathbb{P}(X \in B) = \int_B f(t)\,\mathrm{d}t.$$

The fact that $\mathbb{P}(X = a) = 0$ for every a when X has a density function highlights the fact that many important distribution functions do not have associated density functions.

34.4 Discrete and continuous random variables united: the Lebesgue–Stieltjes integral. We should like to bring all random variables under a common mathematical umbrella. Encouragement is provided by our remarks on Stieltjes measure in 34.1. Every distribution function F has an associated Stieltjes measure μ_F defined on $\mathcal{B}$. Starting from this measure we can construct an integral—known as the Lebesgue–Stieltjes integral. This may be done via simple functions, as mentioned in 34.1. Alternatively it is possible to adapt the method we gave for the Lebesgue integral, as is shown in **[17]**. The development proceeds in the expected way, but care is needed with technicalities: intervals with the same endpoints are no longer indistinguishable and singleton sets are—quite deliberately—not always of zero measure. The major theorems—MCT, DCT, and the theorems of Fubini and Tonelli—all have Lebesgue–Stieltjes versions; see **[10]**. Observe the similarity of form between Theorem 17.5, Tonelli's Theorem, 26.7, and the well-known result that $\sum_n \sum_m a_{mn} = \sum_m \sum_n a_{mn}$ for any complex double sequence for which $\sum_n \sum_m |a_{mn}|$ or $\sum_m \sum_n |a_{mn}|$ converges. It turns out that all these are instances of FTT, in products of suitable measure spaces.

Let X be a random variable with distribution F and write μ_F as $\mathbb{P}$. We write the integral of g as $\int g(x)\,\mathrm{d}F(x)$, when this exists. The motivation for the notation comes from the continuous case, where F has a density function f and, a.e., $F'(x) = f(x)$, or '$f(x)\mathrm{d}x = \mathrm{d}F(x)$'. In particular,

$$\mathbb{P}(X \in B) = \int \chi_B(x)\,\mathrm{d}F(x) \quad (B \in \mathcal{B}).$$

Further, the *expectation* (or *mean*) of X is defined to be

$$\mathbb{E}\,X := \int x\,\mathrm{d}F(x),$$

if this integral exists. In case X has a density function f the expectation becomes

$$\mathbb{E}\,X = \int x f(x)\,\mathrm{d}x,$$

while if X is a discrete random variable with values in $\mathbb{N}$ we have

$$\mathbb{E}\,X = \sum_{k=1}^{\infty} k\mathbb{P}(X = k).$$

If X is a random variable then so is X^p for $p = 2, 3, 4, \ldots$. Then $\mathbb{E}\,X^p$ is the pth *mean* of X, and $\mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2$ is its *variance*, $\mathrm{var}(X)$. Of course, $\mathbb{E}\,X^p$ is given by an integral, reducing in the discrete case to a sum. This integral is not guaranteed to exist, for all or indeed for any values of p, though a value in $\mathbb{R}^+ \cup \{\infty\}$ can be assigned to it if X is non-negative (cf. 21.11). Many theorems in probability accordingly include such statements as 'provided $\mathbb{E}\,X^2$ is finite' or 'provided the expectations exist'.

More generally we may take any Borel measurable function $g\colon \mathbb{R} \to \mathbb{R}$ and form a new random variable $g(X) := g \circ X$. This has distribution function

$$\mathbb{P}(g(X) \leqslant x) = \int_{g^{-1}((-\infty,x])} \mathrm{d}F(x)$$

and the formula for its expectation is

$$\mathbb{E}\, g(X) = \int g(x)\, \mathrm{d}F(x).$$

These ideas can be extended to complex-valued functions, allowing us to consider expressions such as $\mathbb{E}\, \mathrm{e}^{\mathrm{i}X}$.

Certainly random variables which are neither discrete nor continuous arise quite naturally in applied probability. One is led to ask whether a general random variable is made up from discrete and continuous parts. The answer to this question lies in the following result discovered by Lebesgue. Let $F\colon \mathbb{R} \to \mathbb{R}$ be increasing and right-continuous. Then F can be written as $F_1 + F_2 + F_3$, where

(i) F_1 is a jump function (we do not define this formally, but the distribution function of a discrete random variable provides a prototype),

(ii) F_2 is absolutely continuous, and

(iii) F_3 is singular, meaning that it is continuous and increasing, with $F_3' = 0$ a.e.

(see **[10]** for the details).

34.5 Joint distributions. Multi-dimensional integrals enter naturally into probability theory. We consider only continuous random variables; the general case is treated using Stieltjes integrals (see **[10]**). Consider two random variables X and Y on the same probability space. These have a *joint density function* $f_{X,Y}(x, y)$ if

$$F_{X,Y}(x, y) := \mathbb{P}(X \leqslant x,\, Y \leqslant y) = \int_{\{(x,y) \mid X \leqslant x,\, Y \leqslant y\}} f_{X,Y}(x, y)\, \mathrm{d}(x, y),$$

and then, for any Borel set $B \subseteq \mathbb{R}^2$,

$$\mathbb{P}((X, Y) \in B) = \int_B f_{X,Y}(x, y)\, \mathrm{d}(x, y).$$

Observe that this is a 2-dimensional integral rather than a repeated integral, and that Fubini's Theorem is needed to relate it to the associated repeated integrals. This has to be done in particular to obtain the marginal density, f_X, as

$$f_X(x) = \int f_{X,Y}(x, y)\, \mathrm{d}y,$$

and likewise for f_Y.

For a Borel measurable function $g\colon \mathbb{R}^2 \to \mathbb{R}$, the function $g(X, Y)$ is again a random variable. Under suitable technical conditions the change of variables formula

$$\int_{T(U)} g = \int_U (g \circ T)\, |\mathbf{J}_T|$$

holds (see 27.4 and 27.9). From this it is easy to show that, again under suitable conditions,

$$f_{U,V} = (f_{X,Y} \circ T)\, |\mathbf{J}_T|\, \chi_{\operatorname{im} T},$$

where $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ and $(U,V) = T(X,Y)$. The preceding results are extremely useful for finding the distributions of compound random variables. Suppose that X and Y have known densities. Then, for example, the density function of XY is the marginal density f_U, where $(U,V) = T(X,Y)$, with $T(x,y) := (xy, x)$.

At this point in a treatment of probability it would be customary to introduce the ideas relating to conditioning. We do not do this here: both probabilistically and measure theoretically it would take us too far afield.

34.6 Characteristic functions. Let X be a random variable with distribution function F_X. The characteristic function of X is

$$\varphi_X(t) := \mathbb{E}\,\mathrm{e}^{\mathrm{i}tX}\ (= \mathbb{E}\,\cos tX + \mathrm{i}\mathbb{E}\,\sin tX).$$

In case X has a density function f,

$$\varphi_X(t) = \int_{\mathbb{R}} f(x)\mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}x,$$

which is, up to a sign, just the Fourier transform of f, so that the theory developed in Chapter 33 is available. Characteristic functions for some fundamental continuous distributions—the normal distribution and the exponential distribution for example—can be read off from the examples given in Chapter 33. Further, there is a tractable calculus for characteristic functions, whereby $\varphi_{g(X)}$ can be related to φ_X for a range of functions g, in the manner of 33.3 and 33.19. In addition, under suitable technical restrictions, the pth mean of X can be obtained from φ_X:

$$\mathrm{i}^k\mathbb{E}\,X^k = {\varphi_X}^{(k)}(0).$$

Assume that X and Y are random variables on the same probability space. It can be proved that

$$\mathbb{E}\,\mathrm{e}^{\mathrm{i}(tX+sY)} = \mathbb{E}\,\mathrm{e}^{\mathrm{i}tX}\mathbb{E}\,\mathrm{e}^{\mathrm{i}sY} \quad (s,t \in \mathbb{R})$$

is a necessary and sufficient confition for X and Y to be independent. In particular, if X and Y are independent, then

$$\mathbb{E}\,\mathrm{e}^{\mathrm{i}t(X+Y)} = \mathbb{E}\,\mathrm{e}^{\mathrm{i}tX}\mathbb{E}\,\mathrm{e}^{\mathrm{i}tY} \quad (t \in \mathbb{R}).$$

Underlying this is a Stieltjes version of the Convolution Theorem for the Fourier transform, 26.16, and, if X and Y are continuous random variables defined on $\mathbb{R}$, then the Fourier convolution $f_X * f_Y$ is the density function of $X+Y$; if the domain is $\mathbb{R}^+$ rather than $\mathbb{R}$, the Laplace convolution, as defined in Exercise 26.9, serves instead.

Assuming the random variable $X\colon \mathbb{R} \to \mathbb{R}$ is continuous, the Fourier inversion theorem, as formulated in Exercise 33.4, implies that φ_X determines F_X. This result extends to Lebesgue–Stieltjes integrals (see **[10]**), so that φ_X determines X even when X does not have a density function. Therefore a random variable is, as the terminology suggests, characterized by its characteristic function. Thus it is sometimes possible, through calculation of φ_X, to reveal that a random variable X with an apparently unfamiliar distribution is, for example, normally distributed. Further, the correspondence $\varphi_X \mapsto F_X$ is, in a suitable sense, continuous. This result (the **Continuity Theorem**) is clearly important for the analysis of limits of sequences of random variables, or of infinite sums.

34.7 Convergence of random variables. In Chapter 28 we explored various ways in which a sequence of functions might be said to converge, and hinted at how the different modes of convergence are, or are not, related. A good account of such issues in a measure theory setting can be found in [**3**]. In the probabilistic context one wishes to consider a sequence $\{Y_n\}$ of random variables on some given probability space, $(\Omega, \mathcal{E}, \mathbb{P})$, converging to some random variable Y. The modes of convergence in play here include

- convergence almost surely (that is, almost everywhere with respect to the measure $\mathbb{P}$);
- convergence in pth mean, for some $p \in \{1, 2, \dots\}$, meaning $\mathbb{E}\,|Y_n - Y|^p \to 0$, assuming these expectations exist (convergence in L^p norm);
- convergence in probability, meaning that $\mathbb{P}(|X_n - X| > 0) \to 0$ for each fixed $\varepsilon > 0$ (called convergence in measure when the probability space is replaced by a general measure space), and
- convergence in distribution, meaning $\mathbb{P}(Y_n \leqslant x) \to \mathbb{P}(Y \leqslant x)$ at each continuity point x of F_X, where F_Y is the distribution function given by $F_Y(x) := \mathbb{P}(Y \leqslant x)$.

Neither of convergence in pth mean and a.s. convergence implies the other, but both are implied by uniform convergence. Each implies convergence in probability, which in turn implies convergence in distribution. These implications are discussed in detail in probability texts.

Sequences of special importance are those of the form $\{S_n\}$, where $S_n := (X_1 + \cdots + X_n)/n$ and $X_1, X_2, \dots$ are independent and identically distributed, with common expectation $\mu := \mathbb{E}\,X_n$. For instance, the random variables $X_1, X_2, \dots$ might be the outcomes of a seqnence of independent trials. The fundamental results known as the Weak Law of Large Numbers and the Strong Law of Large Numbers assert that, under suitable conditions and in a suitable manner, S_n converges to μ. For the Weak Law the mode of convergence is mean-square convergence and for the Strong Law it is a.s. convergence.

Appendix I: historical remarks

Henri Lebesgue (1875–1941) did his fundamental work on integration in the early years of the 20th century. His major contributions appear in his thesis "Integrale, longueur, aire", which was subsequently published in the journal *Ann. Mat. Pura Appl.* in 1904. He gave accounts of his theory in lecture notes running to several hundred pages, and published, along with other research papers, between 1901 and 1920. Lebesgue's achievement was remarkable: *inter alia*

- he formulated appropriate definitions of measurable sets, measurable functions, and integrable functions (all equivalent to those in use today),
- he established the convergence theorems for his integral,
- he revealed fully the relationship between integration and differentiation (almost everything in Chapter 24 is attributable to Lebesgue, as is the example of a non-integrable derivative given in 13.7), and
- he contributed to the theory of Fourier and trigonometric series, where the Lebesgue spaces can be seen in retrospect to provide the right setting for many valuable and deep results.

Integral calculus dates back to the work of Newton and Leibnitz in the 17th century, and the Fundamental Theorem of Calculus was known and freely used, empirically, in the following two centuries. Fourier published in 1822 his famous work using the series that bear his name in the solution of the heat equation. The background against which Lebesgue did his work was that of a period of intense activity in rigorous mathematical analysis in the latter half of the 19th century, including a detailed theoretical examination of Fourier series. This is associated with the names of Cauchy, Riemann, Weierstrass, and Jordan, among many others.

Various systematic approaches to the **theory** of integration had been propounded prior to 1900. Of these, Riemann's is nowadays the best known, and Riemann's work was significant in that it addressed the question of characterizing those functions which are (Riemann) integrable. The notion of measure had also been explored, most motably by Borel, under whom Lebesgue was a student in Paris. it would be misleading to convey the impression that during his productive years Lebesgue was the only important contributor to the understanding of the structure of the real line and of integrals of real-valued functions. Other investigators, including Borel, W.H. Young, and F. Riesz, gave alternative accounts, and some important theorems concerning the Lebesgue integral are credited to near contemporaries of Lebesgue. By 1920, Fubini and Tonelli had (independently) found their eponymous theorems, the important theorem in 17.5 had been obtained by Beppo Levi, and the Riesz–Fischer Theorem had been established.

In spirit, the approach in this book owes much to Lebesgue himself. Working initially with bounded functions on bounded intervals, Lebesgue set down a list of properties that the integral should possess, essentially equivalent to our Basic Properties with a form of the MCT added. From this starting point, he explored how the integral could be arrived at, and how the class of integrable functions could be characterized. Viewing the integral $\int_a^b f(x)\,\mathrm{d}x$ of a bounded non-negative function f as the area under its graph, Lebesgue split the plane into narrow horizontal slices, and considered the contribution to the desired area falling in each of these strips, in the manner we indicated in 22.6. For strips of width ε, the contribution from the strip corresponding to function values $f(x)$ in the range $[k\varepsilon, (k+1)\varepsilon)$ is given approximately by

$$\varepsilon\, m(S_k), \quad \text{where } S_k := \{\, x \in [a,b] \mid k\varepsilon \leqslant f(x) < (k+1)\varepsilon \,\},$$

and m is a function assigning a length, or 'measure', to suitable subsets of $\mathbb{R}$. This shows that the construction of the integral is reducible to the 'measure problem': which sets are 'measurable' and how should their 'measure' be defined? Lebesgue proposed the following solution to this problem. The notion of Borel measure was already available, so that $m(E)$ had an established meaning when E is a Borel set, and in particular when E is either open or closed. So Lebesgue defined the inner and outer measures of an arbitrary set E in $\mathbb{R}$ by

$$\begin{aligned} m_{\mathrm{i}}(E) &:= \sup\{\, m(F) \mid F \subseteq E,\ F \text{ closed} \,\}, \\ m_{\mathrm{o}}(E) &:= \inf\{\, m(G) \mid G \supseteq E,\ G \text{ open} \,\} \end{aligned}$$

(cf. Exercise 22.10 and notice echoes in the definition of the L^{C} integral (5.1 and Exercise 5.1)). Then E is designated to be measurable if

$$m_{\mathrm{i}}(E) = m_{\mathrm{o}}(E),$$

and when this is so $m(E)$ is given this common value. Every Borel set is measurable in Lebesgue's sense, and the Lebesgue and Borel measures of a Borel set coincide, and null sets are measurable, with measure zero. Lebesgue's insight in defining length (and also area) made possible a new geometry of what are now called 'fractal' sets, which were investigated by Fatou, Julia, and Hausdorff. Their analytic results, published around 1919, later became the basis of modern chaos theory, as well as the source of popular images like that of the Mandelbrot set. A good modern treatment of integration via inner and outer measures can be found in [**14**].

With his notion of measure in place, Lebesgue could define measurable functions, as in 22.7. This led in turn to the definition of the integral of a bounded measurable function on a bounded interval as the limiting case of the horizontal slicing approximation as $\varepsilon \to 0$. Lebesgue treated integrable functions with unbounded domain or image by exploiting a truncation process (the Truncation Lemma is essentially due to Lebesgue) and also gave other equivalent definitions of integrability.

Some comparison between the Riemann and Lebesgue approaches may help to put their relative merits in perspective. Let $[a,b]$ be a compact interval and let $f\colon [a,b] \to \mathbb{R}$ be bounded. Take any partition $\mathcal{P}$ with points $a = x_0 < x_1 < \cdots < x_N = b$, and let

$$m_i = \inf\{\, f(x) \mid x_i < x < x_{i+1} \,\} \quad \text{and} \quad N_i = \sup\{\, f(x) \mid x_i < x < x_{i+1} \,\}.$$

Then

$$L(f,\mathcal{P}) := \sum_{i=0}^{N-1} M_i(x_{i+1} - x_i) \quad \text{and} \quad U(f,\mathcal{P}) := \sum_{i=0}^{N-1} m_i(x_{i+1} - x_i)$$

are called, respectively, the lower and upper Riemann sums associated with $\mathcal{P}$. These can, of course, be recognized as integrals of step functions approximating f from above and below (cf. 9.2). The finer the partition $\mathcal{P}$, that is, the smaller the value of $\max_i |x_{i+1} - x_i|$, the better the approximations provided by $L(f,\mathcal{P})$ and $U(f,\mathcal{P})$ is likely to be. Let $\mathfrak{P}$ denote the family of all partitions on $[a,b]$. The function f is defined to be Riemann integrable if

$$\sup\{\, L(f,\mathcal{P}) \mid \mathcal{P} \in \mathfrak{P} \,\} = \inf\{\, U(f,\mathcal{P}) \mid \mathcal{P} \in \mathfrak{P} \,\}.$$

Let us consider a classic example. Take S to be the set of rational numbers in $[0,1]$, enumerated as $q_1, q_2, \ldots$. Let $f_n := \chi_{\{q_1,\ldots,q_n\}}$. Then it is easy to see that the Riemann integral of f_n exists and equals 0, On the other hand, χ_S is not Lebesgue integrable: because both S and $[0,1] \setminus S$ are dense in $[0,1]$, the upper and lower integrals of χ_S are respectively 0 and 1. Hence $\mathbb{Q} \cap [0,1]$ fails to be Riemann integrable, yet is Lebesgue integrable because it equals the zero function except on a null set. It is discontinuous everywhere. This discussion highlights the fact that Lebesgue's methods accommodate functions with discontinuities much better than Riemann's do. Riemann himself characterized the functions integrable in his sense in terms of the extent of oscillatory behaviour in neighbourhoods of discontinuity points. The problem of understanding sets arising as sets of discontinuities was a spur to the development of a theory of measure. It was proved subsequently by Lebesgue that a bounded function on $[a,b]$ is Riemann integrable if and only if its set of discontinuity points is null. This shows that the functions regularly arising in elementary applications, most of which are piecewise differentiable, are Riemann integrable on compact intervals. On these grounds it is often argued that the Riemann integral is adequate for elementary applications. We dispute this.

For a good theory of integration one requires that the class of functions designated as integrable should have good closure properties under operations one wants to perform. In 14.1 we indicated the great variety of ways in which limiting processes arise in connection with integrals. Therefore it is not surprising that one should seek a theory in which there are good theorems asserting that if $\{f_n\}$ is a sequence of integrable functions satisfying suitable conditions (monotonicity, ...) and such that $\{f_n\}$ converges to f in some suitable sense (pointwise a.e., in a certain norm, ...), then $\lim f_n$ is integrable and $\int \lim f_n = \lim \int f_n$. For a sequence $f_n \nearrow f$ on $[0,1]$ with each f_n Riemann integrable it does not follow that f is Riemann integrable: take, as above, $f_n := \chi_{\{q_1,\ldots,q_n\}}$ and $f := \chi_{\mathbb{Q}} \cap [0,1]$.

Thus a monotone convergence theorem is not available within Riemann's framework. Furthermore, the class of Riemann integrable functions on an interval $[a, b]$ is defective in another way: it does not give a complete space for the L^1 norm.

We may view the lack of good convergence theorems in an integration theory in the following way. If the class of integrable functions is too restricted then it may be necessary to impose conditions which are very stringent and/or very awkward in order to guarantee that $\lim f_n$ is integrable. This is just what happens for the Riemann integral, and matters become worse when integrals on non-compact intervals are encompassed, as improper integrals defined as limits.

A key feature of our approach to the Lebesgue integral harks back to ideas of P.J. Daniell (around 1917). Daniell advocated a 'building block' strategy, whereby an integral is first defined on a class $\mathcal{E}$ of 'elementary' functions, and then extended to a wider class of functions by taking limits. The way we did this—by taking $\mathcal{E}$ to be the step functions and then taking limits of increasing sequences—is appealing for two reasons. First, $\mathrm{L}^{\mathrm{step}}$ is easy to motivate as a candidate for $\mathcal{E}$, using 'integral as area'. Second, the transition from $\mathrm{L}^{\mathrm{step}}$ to $\mathrm{L}^{\mathrm{inc}}$ is exactly what we must do if the MCT is to work. However we should mention some variations on this theme. Instead of relying on the Monotonic Sequence Theorem we may use instead the Cauchy Convergence Principle. Thus, starting again with $\mathcal{E} = \mathrm{L}^{\mathrm{step}}$ we may define f to be Lebesgue integrable if there exists a sequence $\{\varphi_n\}$ of step functions such that $\varphi_n \xrightarrow{\text{a.e.}} f$ and $\{\varphi_n\}$ is *mean-fundamental* in the sense that $\{\|\varphi_n - \varphi_m\|_1 \to 0$ as $m, n \to \infty$. [What is being done here is to choose $\mathrm{L}_{\mathbb{R}}$ to be the completion, as a metric space, of $\mathrm{L}^{\mathrm{step}}$ equipped with the metric $d(\varphi, \psi) := \int |\varphi - \psi|$, where functions equal a.e. are identified.]

Those oriented towards measure theory frequently develop integration theory by first considering measures, and proceed to integrals by using as building blocks the simple functions. For Lebesgue integration, a simple function is defined as a finite linear combination of characteristic functions of bounded Lebesgue measurable sets. See, for example, **[3]** or **[15]** for accounts of integration theory taking this approach.

This book has been almost exclusively concerned with integrals on measurable subsets of $\mathbb{R}$ and $\mathbb{R}^k$. Much of the thrust of mathematical analysis in the 20th century has been towards the investigation of abstract spaces. For example, functional analysis, concerned with general normed spaces, Banach spaces, and Hilbert spaces, grew out of, *inter alia*, the study of particular integral operators. Similarly, the theory of general measure spaces and integration in an abstract setting owes much to the pioneering work of Lebesgue and his contemporaries. There are many important developments of these ideas featuring in many branches of mathematics and its applications. Fascinating as these may be, we cannot even hint at them here: they do not belong in an "Introduction to integration".

Appendix II: reference

The ideal reader of this book would have instant recall of the formulae and statements below, and would have no need of this appendix. Unfortunately it seems unlikely that it will be the least used part of the book.

Useful limits.

(a) As $x \to \infty$, $x^k e^{-x} \to 0$ for any $k \in \mathbb{R}$.

(b) As $x \to \infty$, $x^{-k} \log x \to 0$ for any $k > 0$.

(c) As $x \to 0+$, $x^k \log x \to 0$ for any $k > 0$.

(d) As $x \to 0$, $\dfrac{\sin x}{x} \to 1$.

(e) As $n \to \infty$, $\left(1 + \dfrac{x}{n}\right)^n \to e^x$ for any $x \in \mathbb{R}$.

Useful inequalities.

(a) $|\sin x| \leqslant 1$ on $\mathbb{R}$, with $|\sin x| \leqslant |x|$ providing a tighter bound for $|x| \leqslant 1$.

(b) $\begin{cases} e^{-x^2} \leqslant e^{-|x|} & \text{if } |x| \geqslant 1, \\ e^{-|x|} \leqslant e^{-x^2} & \text{if } |x| \leqslant 1. \end{cases}$

(c) $\sinh x \leqslant e^x \leqslant \cosh x$ for all $x \in \mathbb{R}$.

(d) $\log x \leqslant x$ for $x \geqslant 1$.

(e) $x^p e^{-x} \leqslant 1$ if x is sufficiently large.

(f) Given any $p > 0$, there exists $k > 0$ such that $x^{-p} \log x \leqslant 1$ if $x > k$.

(g) Given any $p > 0$, there exists $\delta > 0$ such that $x^p \log x \leqslant 1$ for $0 < x < \delta$.

Too-easily-forgotten trigonometric formulae.

$$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$$

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$$

$$\sin A \pm \sin B = 2 \sin\left(\tfrac{1}{2}(A \pm B)\right) \sin\left(\tfrac{1}{2}(A \mp B)\right)$$

$$\cos A + \cos B = 2 \cos\left(\tfrac{1}{2}(A + B)\right) \cos\left(\tfrac{1}{2}(A - B)\right)$$

$$\cos A - \cos B = -2 \sin\left(\tfrac{1}{2}(A + B)\right) \sin\left(\tfrac{1}{2}(A - B)\right)$$

$$2 \sin^2 \tfrac{1}{2}A = 1 - \cos A$$

$$2 \cos^2 \tfrac{1}{2}A = 1 + \cos A$$

Standard derivatives and antiderivatives. In the table below, f is the derivative of F, and F is an antiderivative of f. In all cases log is assumed to be applied to a strictly positive argument.

$F(x)$	$f(x)$	$F(x)$	$f(x)$
$x^\alpha \ (\alpha \neq 0)$	$\alpha x^{\alpha-1}$	$\log x$	$\dfrac{1}{x}$
e^x	e^x	$\log(\log x)$	$\dfrac{1}{x\log x}$
$a^x \ (:= e^{x\log a})$	$a^x \log a$	$x\log x - x$	$\log x$
$\cos x$	$-\sin x$	$\cosh x$	$\sinh x$
$\sin x$	$\cos x$	$\sinh x$	$\cosh x$
$\tan x$	$\sec^2 x$	$\tanh x$	$\operatorname{sech}^2 x$
$\cot x$	$-\operatorname{cosec}^2 x$	$\coth x$	$-\operatorname{cosech}^2 x$
$\sec x$	$\tan x \sec x$	$\operatorname{sech} x$	$-\tanh x \operatorname{sech} x$
$\operatorname{cosec} x$	$-\cot x \operatorname{cosec} x$	$\operatorname{cosech} x$	$-\coth x \operatorname{cosech} x$
$\cos^{-1}\left(\dfrac{x}{a}\right)$	$-\dfrac{1}{\sqrt{a^2-x^2}}$	$\cosh^{-1}\left(\dfrac{x}{a}\right)$	$\dfrac{1}{\sqrt{x^2-a^2}}$
$\sin^{-1}\left(\dfrac{x}{a}\right)$	$\dfrac{1}{\sqrt{a^2-x^2}}$	$\sinh^{-1}\left(\dfrac{x}{a}\right)$	$\dfrac{1}{\sqrt{a^2+x^2}}$
$\tan^{-1}\left(\dfrac{x}{a}\right)$	$\dfrac{a}{a^2+x^2}$	$\tanh^{-1}\left(\dfrac{x}{a}\right)$ or $\dfrac{1}{2a}\log\left(\dfrac{a+x}{a-x}\right)$	$\dfrac{a}{a^2-x^2}$
$\cot^{-1}\left(\dfrac{x}{a}\right)$	$-\dfrac{a}{x^2+a^2}$	$\coth^{-1}\left(\dfrac{x}{a}\right)$ or $\dfrac{1}{2a}\log\left(\dfrac{x+a}{x-a}\right)$	$-\dfrac{a}{x^2-a^2}$

Useful integrable functions outside L^{step} and L^{C}.

f	I	Condition for $f \in \mathrm{L}(I)$	$\int f \chi_I$	Ref.
x^p	$(0,1]$	$p > -1$	$\dfrac{1}{p+1}$	15.1
x^p	$[1,\infty)$	$p < -1$	$\dfrac{1}{p+1}$	15.1
$\dfrac{1}{1+x^2}$	$\mathbb{R}$		π	15.4
e^{-x}	$(0,\infty)$		1	15.3
$\mathrm{e}^{-\lvert x\rvert}$	$\mathbb{R}$		2	15.3
e^{-x^2}	$\mathbb{R}$		$\sqrt{\pi}$	15.3, 20.10
$x^p\mathrm{e}^{-ax}$	$[0,\infty)$	$p > -1,\ a > 0$	$\dfrac{(p+1)!}{a^{p+1}}$ $(p \in \mathbb{N})$	15.9, 16.5
$\mathrm{e}^{-ax}\sin bx$	$[0,\infty)$	$a > 0$	$\dfrac{b}{a^2+b^2}$	15.9, 16.5
$\mathrm{e}^{-ax}\cos bx$	$[0,\infty)$	$a > 0$	$\dfrac{a}{a^2+b^2}$	15.9, 16.5
$x^q(\log x)^p$	$(0,1]$	$q > -1,\ p > 0$	$(-1)^q\dfrac{p!}{(q+1)^{p+1}}$ $(p,q \in \mathbb{N})$	18.4, Ex. 16.5
$x^p(\log x)^q$	$[1,\infty)$	$q < -1,\ p > 0$	$\dfrac{p!}{\lvert q+1\rvert^{p+1}}$ $(q < -1, p \in \mathbb{N})$	Ex. 16.5

Series expansions. The following expansions are valid for $x \in \mathbb{R}$ unless otherwise indicated.

$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$	$\log(1-x) = -\sum_{k=1}^{\infty} \frac{x^k}{k} \quad (\lvert x\rvert < 1)$
$\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}$	$\sinh x = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}$
$\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$	$\cosh x = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}$
$(1-x)^{-1} = \begin{cases} \sum_{k=0}^{\infty} x^k & (\lvert x\rvert < 1) \\ \sum_{k=0}^{\infty} x^{-k-1} & (\lvert x\rvert > 1) \end{cases}$	$(1-x)^{-2} = \sum_{k=1}^{\infty} k x^{k-1} \quad (\lvert x\rvert < 1)$
$(1-x)^\alpha = \sum_{k=0}^{\infty} \alpha(\alpha-1)\dots(\alpha-k+1)\frac{x^k}{k!}$ $(\lvert x\rvert < 1,\ \alpha \in \mathbb{R})$	$\tan^{-1} x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)}$ $(\lvert x\rvert \leqslant 1)$

Bibliography

1. T.M. Apostol, *Mathematical analysis*, 2nd edition, Addison–Wesley, Reading, Mass., 1974.
2. E. Asplund and L. Bungart, *A first course in integration*, Holt, Rinehart and Winston, New York, 1966.
3. R.G. Bartle, *The elements of integration*, John Wiley and Sons, New York, 1966.
4. H. Dym and H.P. McKean, *Fourier series and integrals*, Academic Press, New York, 1972.
5. R.E. Edwards, *Fourier series*, Holt, Rinehart and Winston, New York, 1967.
6. H.B. Enderton, *Elements of set theory*, Academic Press, New York, 1977.
7. G.R. Grimmett and D.R. Stirzaker, *Probability and random processes*, 2nd edition, Oxford University Press, Oxford, 1992.
8. F.B. Hildebrand, *Methods of applied mathematics*, 2nd edition, Prentice–Hall, Englewood Cliffs, N.J., 1965.
9. H. Hochstadt, *Integral equations*, Wiley–Interscience, New York, 1973.
10. J.F.C. Kingman and S.J. Taylor, *Introduction to measure and probability*, Cambridge University Press, Cambridge, 1966.
11. T.W. Körner, *Fourier analysis*, Cambridge University Press, Cambridge, 1988.
12. E. Kreyszig, *Advanced engineering mathematics*, 7th edition, Wiley, New York, 1993.
13. H.A. Priestley, *Introduction to complex analysis*, revised edition, Oxford University Press, Oxford, 1990.
14. H.L. Royden, *Real analysis*, Macmillan, New York, 1963.
15. W. Rudin, *Real and complex analysis*, 2nd edition, McGraw–Hill, New York, 1974.
16. W.A. Sutherland, *Introduction to metric and topological spaces*, Oxford University Press, Oxford, 1975.
17. B. Sz.-Nagy, *Introduction to real functions and orthogonal expansions*, Oxford University Press, Oxford, 1965.
18. A.J. Weir, *Lebesgue integration and measure*, Cambridge University Press, Cambridge, 1973.
19. E.T. Whittaker and G.N. Watson, *A course of modern analysis*, 4th edition, Cambridge University Press, Cambridge, 1927.

Notation index

Acronyms, etc.

Subject index

Page numbers given in boldface refer to definitions and statements of theorems and those in italic to exercises, with the latter overriding the former where definitions are given in exercises.