Harmonic Analysis:
From Fourier to Wavelets

María Cristina Pereyra

Lesley A. Ward

Department of Mathematics and Statistics, MSC03 2150, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA

E-mail address: crisp@math.unm.edu

School of Mathematics and Statistics, University of South Australia, Mawson Lakes SA 5095, Australia

E-mail address: Lesley.Ward@unisa.edu.au

2000 *Mathematics Subject Classification.* Primary 42-01, Secondary 42-02, 42A16, 42A20, 42A24, 42A38, 42A65, 42B20, 42B25, 42C40

To Tim, Nicolás, and Miguel

To Jorge and Alexander

Contents

List of Figures

List of Tables

Preface

Over two hundred years ago, Jean Baptiste Joseph Fourier began to work on the theory of heat and how it flows. His book *Théorie Analytique de la Chaleur* (The Analytic Theory of Heat) was published in 1822. In that work, he began the development of what is possibly the single most influential body of mathematical ideas, encompassing Fourier theory and the field now known as *harmonic analysis* that has grown from it. Since that time, the subject has been exceptionally significant both in its theoretical implications and in its enormous range of applicability throughout mathematics, science, and engineering.

On the theoretical side, the theory of Fourier series was a driving force in the development of mathematical analysis, the study of functions of a real variable. For instance, notions of convergence were created in order to deal with the subtleties of Fourier series. One could also argue that set theory, including the construction of the real numbers and the ideas of cardinality and countability, was developed because of Fourier theory. On the applied side, all the signal processing done today relies on Fourier theory. Everything from the technology of mobile phones to the way images are stored and transmitted over the Internet depends on the theory of Fourier series. Most recently the field of wavelets has arisen, uniting its roots in harmonic analysis with theoretical and applied developments in fields such as medical imaging and seismology.

In this book, we hope to convey the remarkable beauty and applicability of the ideas that have grown from Fourier theory. We present for an advanced undergraduate and beginning graduate student audience the basics of harmonic analysis, from Fourier's heat equation, and the decomposition of functions into sums of cosines and sines (frequency analysis), to dyadic harmonic analysis, and the decomposition of functions into a Haar basis (time localization). In between these two different ways of decomposing functions there is a whole world of time/frequency analysis (wavelets). We touch on aspects of that world, concentrating on the Fourier and Haar cases.

The book is organized as follows. In the first five chapters we lay out the classical theory of Fourier series. In Chapter 1 we introduce Fourier series for periodic functions and discuss physical motivations. In Chapter 2 we present a review of different modes of convergence and appropriate classes of functions. In Chapter 3 we discuss pointwise convergence for Fourier series and the interplay with differentiability. In Chapter 4 we introduce approximations of the identity, the Dirichlet, Fejér, and Poisson kernels, and summability methods for Fourier series. Finally in Chapter 5 we discuss inner-product vector spaces, and the completeness of the trigonometric basis in $L^2(\mathbb{T})$.

In Chapter 6, we examine discrete Fourier and Haar analysis, on finite-dimensional spaces. We present the Discrete Fourier Transform and its celebrated cousin the Fast Fourier Transform, an algorithm discovered by Gauss in the early 1800s and rediscovered by Cooley and Tukey in the 1960s. We compare them with the Discrete Haar Transform and the Fast Haar Transform algorithm.

In Chapters 7 and 8 we discuss the Fourier transform on the line. In Chapter 7 we introduce the time–frequency dictionary, convolutions, and approximations of the identity in what we call paradise (the Schwartz class). In Chapter 8, we go beyond paradise and discuss the Fourier transform for tempered distributions, construct a time–frequency dictionary in that setting, and discuss the delta distribution as well as the principal value distribution. We survey the mapping properties of the Fourier transform acting on L^p spaces, as well as a few canonical applications of the Fourier transform including the Shannon sampling theorem.

In Chapters 9, 10, and 11 we discuss wavelet bases, with emphasis on the Haar basis. In Chapter 9, we survey the windowed Fourier transform, the Gabor transforms, and the wavelet transform. We develop in detail the Haar basis, the geometry of dyadic intervals, and take some initial steps into the world of dyadic harmonic analysis. In Chapter 10, we discuss the general framework of multiresolution analysis for constructing other wavelets. We describe some canonical applications to image processing and compression and denoising, illustrated in a case study of the wavelet-based FBI fingerprint standard. We state and prove Mallat's theorem, and explain how to search for suitable multiresolution analyses. In Chapter 11, we discuss algorithms and connections to filter banks. We revisit the algorithm for the Haar basis and the multiresolution analysis that they induce. We describe the cascade algorithm, and how to implement the wavelet transform given multiresolution analysis, using filter banks to obtain the Fast Wavelet Transform. We describe some properties and design features of known wavelets, as well as the basics of image/signal denoising and compression.

To finish our journey, in Chapter 12 we present the Hilbert transform, the most important operator in harmonic analysis after the Fourier transform. We describe the Hilbert transform in three ways: as a Fourier multiplier, as a singular integral, and as an average of Haar shift operators. We discuss how the Hilbert transform acts on the function spaces L^p, as well as some tools for understanding the L^p spaces. In particular we discuss the Riesz–Thorin Interpolation Theorem, and as an application derive some of the most useful inequalities in analysis. Finally we explain the connections of the Hilbert transform with complex analysis and with Fourier analysis.

Each chapter ends with ideas for projects in harmonic analysis that students can work on rather independently, using the material in our book as a springboard. We have found that such projects help students to become deeply engaged in the subject matter, in part by giving them the opportunity to take ownership of a particular topic. We believe the projects will be useful both for individual students using our book for independent study, and for students using the book in a formal course.

The prerequisites for our book are advanced calculus and linear algebra. Some knowledge of real analysis would be helpful but is not required. We introduce concepts from Hilbert spaces, Banach spaces, and the theory of distributions as needed. Chapter 2 is an interlude about analysis on compact intervals. Appendix A reviews vector, normed and inner-product spaces, as well as some key concepts from analysis on the line, used in our later chapters. We view the book as an introduction to serious analysis and computational harmonic analysis through the lens of Fourier and wavelet analysis.

Examples, exercises, and figures appear throughout the text. The notations $A := B$ and $B =: A$ both mean that A is defined to be the quantity B. We use the symbol $\Box$ to mark the end of a proof, and the symbol $\Diamond$ to mark the end of an example, exercise, aside, remark, or definition.

Suggestions for instructors

The first author used drafts of our book twice as the text for a one-semester course on *Fourier Analysis and Wavelets* at the University of New Mexico, aimed at upper-level undergraduate and graduate students. She covered most of the material in Chapters 1 and 3–11, and used a selection of the student projects, omitting Chapter 12 for lack of time. The concepts and ideas in Chapter 2 were discussed as the need arose while lecturing on the other chapters, and students were encouraged to revisit that chapter as needed.

One can design other one-semester courses based on this book. The instructor could make such a course more theoretical (following the L^p stream, excluding Chapters 6, 10 and 11), or more computational (excluding the L^p stream and Chapter 12, and including Chapter 6 and parts of Chapters 10 and 11). In both situations Chapter 2 is a resource, not meant as lecture material. For a course exclusively on Fourier analysis, Chapters 1–8 have more than enough material. For an audience already familiar with Fourier series, one could start in Chapter 6 with a brief review. Then do the discrete Fourier and Haar transforms, for which only linear algebra is needed, and move on to Fourier integrals, Haar analysis and wavelets. Finally, treat the Hilbert transform, or instead supplement the course with more applications, perhaps inspired by the projects. We believe that the emphasis on the Haar basis and dyadic harmonic analysis make the book distinctive, and we would include that material.

The 20 projects vary in difficulty and sophistication. We have written them to be flexible and open-ended, and we encourage instructors to modify them and to create their own. Some of our projects are structured sequentially, with each part building on earlier parts, while in other projects the individual parts are independent. Some projects are simple in form but quite ambitious, asking students to absorb and report on recent research papers. Our projects are suitable for individuals or teams of students.

It works well to ask students both to give an oral presentation on the project and to write a report and/or a literature survey. The intended audience is another student at the same stage of their studies, but without knowledge of the specific project content. In this way, students develop skills in various types of mathematical communication, and students with differing strengths get a chance to shine. Instructors can reserve the last two or three weeks of lectures for student talks, if there are few enough students to make this practicable. We find this to be a worthwhile use of time.

It is fruitful to set up some project milestones throughout the course, in the form of a series of target dates for such tasks as preparing outlines of the oral presentation and of the report, drafting summaries of the first few items in a literature survey, rehearsing a presentation, completing a draft of a report, and so on. Early planning and communication here will save much stress later. In the second week of classes, we like to have an initial ten-minute conversation with each student, discussing a few preliminary sentences they have written on their early ideas for the content and structure of the project they plan to do. Such a meeting enables timely intervention if the proposed scope of the project is not realistic, for instance. A little later in the semester, students will benefit from a brief discussion with the instructor on the content of their project and their next steps, once they have got their teeth into the ideas.

Students may find it helpful to use the mathematical typesetting package LaTeX for their reports, and software such as Beamer to create slides for their oral presentations. Working on a project provides good motivation for learning such professional tools. Here is a natural opportunity for instructors to give formal or informal training in the use of such tools and in mathematical writing and speaking. We recommend Higham's book [**Hig**] and the references it contains as an excellent place to start. Instructors contemplating the task of designing and planning semester-long student projects will find much food for thought in Bean's book [**Bea**].

Acknowledgements

Our book has grown out of the lecture course we taught at the Institute for Advanced Study IAS/Princeton Program for Women in Mathematics on *Analysis and Partial Differential Equations*, May 17–28, 2004. We thank Manuela de Castro and Stephanie Molnar (now Salomone), our teaching assistants during the Program, for their invaluable help, and all the students for their enthusiasm and lively participation. We thank Sun-Yung Alice Chang and Karen Uhlenbeck for the opportunity to participate in this remarkable program.

The book also includes material from a course on wavelets developed and taught by the second author, twice at Harvey Mudd College, and again at the IAS/Park City Mathematics Institute Summer Session on *Harmonic Analysis and Partial Differential Equations*, June 29–July 19, 2003. We thank her teaching assistant at PCMI, Stephanie Molnar.

Early drafts of the book were written while the second author was a member of the Department of Mathematics, Harvey Mudd College, Claremont, California.

The mathematical typesetting software LaTeX and the version control system Subversion were invaluable, as was Barbara Beeton's expert assistance with LaTeX. The MathWorks helped us with Matlab through their book program for authors. We created the figures using LaTeX and Matlab. Early versions of some figures were kindly provided by Martin Mohlenkamp. We thank John Polking for his help in improving many of the Matlab figures, and John Molinder and Jean Moraes for recording and plotting the voice signal in Figure 1.2. We gratefully acknowledge Chris Brislawn's permission to reproduce figures and other material from his website about the FBI Fingerprint Image Compression Standard [**Bri02**].

In writing this book we have drawn on many sources, too numerous to list in full here. An important debt is to E.M. Stein and R. Shakarchi's book [**SS03**], which we used in our Princeton lectures as the main text for classical Fourier theory. We were particularly inspired by T. Körner's book [**Kör**].

Students who took the wavelets courses at the University of New Mexico proofread the text, suggesting numerous improvements. We have incorporated several of their ideas for projects. Harvey Mudd College students Neville Khambatta and Shane Markstrum, and Claremont Graduate University student Ashish Bhan, typed LaTeX notes and created Matlab figures for the second author's wavelets course. In the final stages, Matthew Dahlgren and David Weirich, graduate students at the University of New Mexico, did a careful and thoughtful reading, giving us detailed comments from the point of view of an informed reader working through the book without a lecture course. They also contributed to the subject index. Lisa Schultz, an honours student at the University of South Australia, completed both indexes and helped with the final touches. Over the years colleagues and students have given suggestions on draft versions, including Prof. Wilfredo Urbina, University of New Mexico graduate students Adam Ringler and Elizabeth Kappilof, and Brown University graduate student Constance Liaw.

We thank the following students who wrote project reports, gave presentations, and in some cases came up with research topics: Mike Klopfer (Project 10.6), Adam Ringler (Project 6.9), Kourosh Raeen (Project 3.4), and Bernadette Mendoza-Spencer (Project 12.9.1).

Abbreviated versions of Projects 3.4, 4.8 and 9.7 were tested as guided one-day projects with 15–18 upper-level undergraduates and first-year graduate students, in three editions of the one-week minicourse on *Fourier Analysis and Wavelets* taught by the first author as part of the NSF-sponsored *Mentoring Through Critical Transition Points* (MCTP) Summer Program, held at the University of New Mexico in July 2008, June 2009 and July 2010.

We are grateful for the helpful and constructive suggestions of the anonymous reviewers.

Special thanks to our editor Ed Dunne and the AMS editorial staff, who have guided us from our brief initial draft to the completion of this book.

The generous contributions of all these people helped us to make this a better book, and we thank them all. All remaining errors and infelicities are entirely our responsibility.

Finally we would like to thank our families for their patience and love while this book has been in the making. During these years a baby was born, one of us moved to Australia, and energetic toddlers grew into thriving 9- and 11-year-old children. Parents, our constant supporters and cheerleaders, now more than ever need our support and love. This writing project is over, but life goes on.

María Cristina Pereyra, University of New Mexico
Lesley A. Ward, University of South Australia
December, 2011

Chapter 1

Fourier series: some motivation

In this book we discuss three types of Fourier analysis[1]: first, *Fourier series*, in which the input is a periodic function on $\mathbb{R}$, and the output is a two-sided series where the summation is over $n \in \mathbb{Z}$ (Chapters 1–5); second, *finite Fourier analysis*, where the input is a vector of length N with complex entries, and the output is another vector in $\mathbb{C}^N$ (Chapter 6); and third, the *Fourier transform*, where the input is a function on $\mathbb{R}$, and the output is another function on $\mathbb{R}$ (Chapters 7–8). For the Fourier transform we treat separately the case of the extremely well behaved Schwartz functions and that of the less well behaved distributions. We build on this foundation to discuss the windowed Fourier transform, Gabor transforms, various wavelet transforms, the Hilbert transform, and applications (Chapters 9–12).

In this first chapter we take our initial steps into the theory of Fourier series. We present a simple example in signal processing and compression (Section 1.1). We lay down the basic questions regarding expansions in trigonometric series, and compare to expansions in Taylor series (Section 1.2). We introduce and motivate the mathematical

[1]Named after French mathematician and physicist Jean Baptiste Joseph Fourier (1768–1830). For a sketch of Fourier's fascinating life, including his contributions to Egyptology, see [**Kör**, Chapters 92, 93].

definition of Fourier coefficients and series, and discuss periodic functions (Section 1.3). We describe the simple physical problem of heat diffusion on a rod that led Fourier to assert that "all periodic functions" could be expanded in a series of sines and cosines (Section 1.4).

1.1. An example: Amanda calls her mother

The central idea of Fourier analysis is to break a function into a combination of simpler functions. We think of these simpler functions as *building blocks*. We will be interested in reconstructing the original function from the building blocks. Here is a colorful analogy: *a prism or raindrop can break a ray of (white) light into all colors of the rainbow.* The analogy is quite apt: the different colors correspond to different *wavelengths/frequencies* of light. We can reconstruct the white light from the composition of all the different wavelengths. We will see that our simpler functions can correspond to pure frequencies. We will consider sine and cosine functions of various frequencies as our first example of these building blocks. When played aloud, a given sine or cosine function produces a *pure tone*, or note, or harmonic, at a single frequency. The term *harmonic analysis* evokes this idea of separation of a sound, or in our terms a function, into pure tones.

Expressing a function as a combination of building blocks is also called *decomposing* the function. In Chapters 9 and 10 we will study so-called *time–frequency decompositions*, in which each building block encodes information about time as well as about frequency, very much like musical notation does. We begin with an example.

Example 1.1. (*Toy Model of Voice Signal*) Suppose Amanda is in Baltimore, and she calls her mother in Vancouver, saying 'Hi Mom, it's Amanda, and I can't wait to tell you what happened today'. What happens to the sound? As Amanda speaks, she creates waves of pressure in the air, which travel toward the phone receiver. The sound has duration, about five seconds in this example, and intensity or loudness, which varies over time. It also has many other qualities that make it sound like speech rather than music, say. The sound of Amanda's voice becomes a *signal*, which travels along the phone wire or via satellite, and at the other end is converted back into a recognizable voice.

Let's try our idea of breaking Amanda's voice signal into simpler building blocks. Suppose the signal looks like the function $f(t)$ plotted in the top half of Figure 1.1.

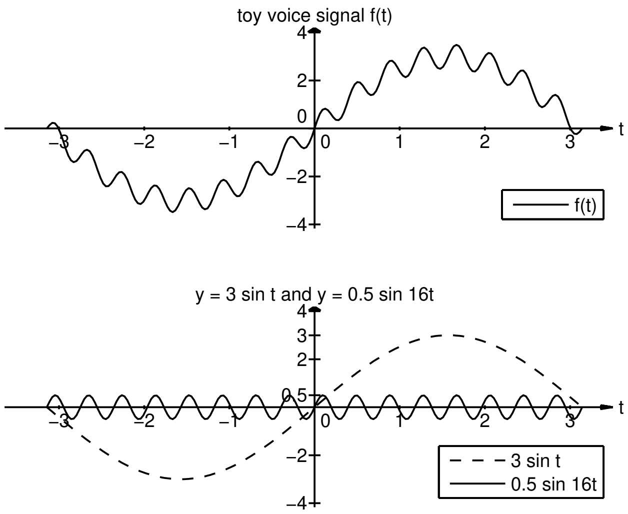

Figure 1.1. Toy voice signal $f(t)$ (upper plot), and the building blocks of the toy voice signal (lower plot).

In Figure 1.1 the horizontal axis represents time t in seconds, and the vertical axis represents the intensity of the sound, so that when $y = f(t)$ is near zero the sound is soft, and when $y = f(t)$ is large (positive or negative) the sound is loud.

In this particular signal, there are two sorts of wiggling going on; see the lower half of Figure 1.1. We recognize the large, slow wiggle as a multiple of $\sin t$. The smaller wiggle is oscillating much faster; counting the oscillations, we recognize it as a multiple of $\sin 16t$. Here $\sin t$ and $\sin 16t$ are our first examples of building blocks. They are functions of *frequency* 1 and 16 respectively. This means they complete 1 and 16 full oscillations as t runs through 2π units of time.

Next we need to know how much of each building block is present in our signal. The maximum amplitudes of the large and small wiggles are 3 and 0.5 respectively, giving us the terms $3\sin t$ and $0.5\sin 16t$. We add these together to build the original signal:

$$f(t) = 3\sin t + 0.5\sin 16t.$$

We have written our signal $f(t)$ as a sum of constant multiples of the two building blocks. The right-hand side is our first example of a *Fourier decomposition*; it is the Fourier decomposition of our signal $f(t)$.

Let's get back to the phone company. When Amanda calls her mother, the signal goes from her phone to Vancouver. How will the signal be encoded? For instance, the phone company could take a collection of say 100 equally spaced times, record the strength of the signal at each time, and send the resulting 200 numbers to Vancouver, where they know how to put them back together. (By the way, can you find a more efficient way of encoding the timing information?) But for our signal, we can use just *four* numbers, and achieve an even more accurate result. We simply send the frequencies 1 and 16, and the corresponding strengths or *amplitudes* 3 and 0.5. Once you know the code, namely that the numbers represent sine wave frequencies and amplitudes, these four numbers are all you need to know to rebuild our signal exactly.

Note that for a typical, more complicated signal, one would need more than just two sine functions as building blocks. Some signals would require infinitely many sine functions, and some would require cosine functions instead or as well. The collection

$$\{\sin(nt) : n \in \mathbb{N}\} \quad \cup \quad \{\cos(nt) : n \in \mathbb{N} \cup \{0\}\}$$

of all the sine and cosine functions whose frequencies are positive integers, together with the constant function with value 1, is an example of a *basis*. We will return to the idea of a basis later; informally, it means a given collection of building blocks that is able to express every function in some given class. Fourier series use sines and cosines;

other types of functions can be used as building blocks, notably in the wavelet series we will see later (Chapters 9 and 10).

Now let's be more ambitious. Are we willing to sacrifice a bit of the quality of Amanda's signal in order to send it more cheaply? Maybe. For instance, in our signal the strongest component is the big wiggle. What if we send only the single frequency 1 and its corresponding amplitude 3? In Vancouver, only the big wiggle will be reconstructed, and the small fast wiggle will be lost from our signal. But Amanda's mother knows the sound of Amanda's voice, and the imperfect reconstructed signal may still be recognizable. This is our first example of *compression* of a signal.

To sum up: We *analyzed* our signal $f(t)$, determining which building blocks were present and with what strength. We *compressed* the signal, discarding some of the frequencies and their amplitudes. We *transmitted* the remaining frequencies and amplitudes. (At this stage we could also have *stored* the signal.) At the other end they *reconstructed* the compressed signal.

Again, in practice one wants the reconstructed signal to be similar to the original signal. There are many interesting questions about which building blocks, and how many, can be thrown away while retaining a recognizable signal. This is part of the subject of signal processing, in electrical engineering.

In the interests of realism, Figure 1.2 shows a plot of an actual voice signal. Its complicated oscillations would require many more sine and cosine building blocks than we needed for our toy signal. $\Diamond$

In another analogy, we can think of the decomposition as a recipe. The building blocks, distinguished by their frequencies, correspond to the different ingredients in your recipe. You also need to know *how much* sugar, flour, and so on you have to add to the mix to get your cake f; that information is encoded in the amplitudes or coefficients.

The function f in Example 1.1 is especially simple. The intuition that sines and cosines were sufficient to describe many different functions was gained from the experience of the pioneers of Fourier analysis with physical problems (for instance heat diffusion and the vibrating string) in the early eighteenth century; see Section 1.4. This

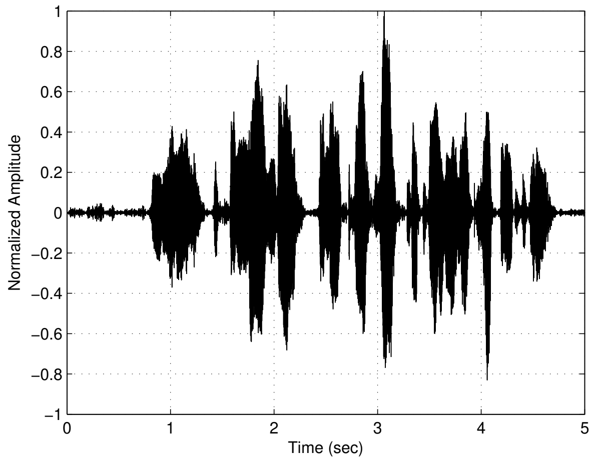

Figure 1.2. Plot of a voice saying 'Hi Mom, it's Amanda, and I can't wait to tell you what happened today'. The horizontal axis represents time, in seconds. The vertical axis represents the amplitude (volume) of the signal. The units of amplitude are decibels, but here we have normalized the amplitude by dividing by its maximum value. Thus the normalized amplitude shown is dimensionless and has maximum value one.

intuition leads to the idea of expressing a periodic function $f(\theta)$ as an infinite linear combination of sines and cosines, also known as a *trigonometric series*:

$$f(\theta) \sim \sum_{n=0}^{\infty} \left[b_n \sin(n\theta) + c_n \cos(n\theta) \right]. \tag{1.1}$$

We use the symbol $\sim$ to indicate that the right-hand side is the trigonometric series *associated with* f. We don't use the symbol $=$ here since in some cases the left- and right-hand sides of (1.1) are not equal, as discussed below.

We can rewrite the right-hand side of (1.1) as a linear combination of exponential functions. To obtain this, we use *Euler's Formula*[2]:

[2]This formula is named after the Swiss mathematician Leonhard Euler (pronounced "oiler"), (1707–1783). We assume the reader is familiar with basic complex number operations and notation. For example if $v = a + ib$, $w = c + id$, then $v + w = (a + c) + i(b + d)$ and $vw = (ac - bd) + i(bc + ad)$. The algebra is done

$e^{i\theta} = \cos\theta + i\sin\theta$, and the corresponding formulas for sine and cosine in terms of exponentials, applied to $n\theta$ for each n,

$$\cos\theta = (e^{i\theta} + e^{-i\theta})/2, \qquad \sin\theta = (e^{i\theta} - e^{-i\theta})/(2i).$$

We will use the following version throughout:

$$f(\theta) \sim \sum_{n=-\infty}^{\infty} a_n e^{in\theta}. \tag{1.2}$$

The right-hand side is called the *Fourier series* of f. We also say that f can be *expanded* in a Fourier series. The coefficient a_n is called the n^{th} *Fourier coefficient* of f, and is often denoted by $\widehat{f}(n)$ to emphasize the dependence on the function f. The coefficients $\{a_n\}$ are determined by the coefficients $\{b_n\}$ and $\{c_n\}$ in the sine/cosine expansion.

Exercise 1.2. Write a_n in terms of b_n and c_n. ♢

1.2. The main questions

Our brief sketch immediately suggests several questions:

How can we find the Fourier coefficients a_n from the corresponding function f?

Given the sequence $\{a_n\}_{n\in\mathbb{Z}}$, how can we reconstruct f?

In what sense does the Fourier series converge? Pointwise? Uniformly? Or in some other sense? If it does converge in some sense, how fast does it converge?

When the Fourier series does converge, is its limit equal to the original function f?

Which functions can we express with a trigonometric series

$$\sum_{n=-\infty}^{\infty} a_n e^{in\theta}\,?$$

For example, must f be continuous, or Riemann integrable?

as it would be for real numbers, with the extra fact that the imaginary unit i has the property that $i^2 = -1$. The *absolute value* of a complex number $v = a+ib$ is defined to be $|v| = \sqrt{a^2+b^2}$. See **[Tao06b**, Section 15.6] for a quick review of complex numbers.

> What other building blocks could we use instead of sines and cosines, or equivalently exponentials?

We develop answers to these questions in the following pages, but before doing that let us compare to the more familiar problem of power series expansions. An infinitely-often differentiable function f can be expanded into a *Taylor series*[3] centered at $x = 0$:

$$\sum_{n=0}^{\infty} c_n x^n. \tag{1.3}$$

The numbers c_n are called the *Taylor coefficients* and are given by

$$c_n = f^{(n)}(0)/n!.$$

When evaluated at $x = 0$, this Taylor series always converges to $f(0) = c_0$. (For simplicity we have used the series centered at $x = 0$; a slight variation familiar from calculus texts gives a Taylor series centered at a general point x_0.)

In the 18th century mathematicians discovered that the traditional functions of calculus ($\sin x$, $\cos x$, $\ln(1+x)$, $\sqrt{1+x}$, e^x, etc.) can be expanded in Taylor series, and that for these examples the Taylor series converges to the value $f(x)$ of the function for each x in an open interval containing $x = 0$, and sometimes at the endpoints of this interval as well. They were very good at manipulating Taylor series and calculating with them. That led them to believe that the same would be true for all functions, which at the time meant for the infinitely-often differentiable functions. That dream was shattered by Cauchy's[4] discovery in 1821 of a counterexample.

Example 1.3. (*Cauchy's Counterexample*) Consider the function

$$f(x) = \begin{cases} e^{-1/x^2}, & \text{if } x \neq 0; \\ 0, & \text{if } x = 0. \end{cases}$$

This function is infinitely often differentiable at every x, and $f^{(n)}(0) = 0$ for all $n \geq 0$. Thus its Taylor series is identically equal to zero. Therefore the Taylor series will converge to $f(x)$ only for $x = 0$. ◊

[3] Named after the English mathematician Brook Taylor (1685–1731). Sometimes the Taylor series centered at $x = 0$ is called the Maclaurin series, named after the Scottish mathematician Colin Maclaurin (1698–1746).

[4] The French mathematician Augustin-Louis Cauchy (1789–1857).

Exercise 1.4. Verify that Cauchy's function is infinitely differentiable. Concentrate on what happens at $x = 0$. $\Diamond$

The *Taylor polynomial* P_N *of order* N of a function f that can be differentiated at least N times is given by the formula

$$P_N(f,0)(x) = f(0) + f'(0)\,x + \frac{f''(0)}{2!}\,x^2 + \cdots + \frac{f^{(N)}(0)}{N!}\,x^N. \tag{1.4}$$

Exercise 1.5. Verify that if f is a polynomial of order less than or equal to N, then f coincides with its Taylor polynomial of order N. $\Diamond$

Definition 1.6. A *trigonometric*[5] *polynomial of degree* M is a function of the form

$$f(\theta) = \sum_{n=-M}^{M} a_n e^{in\theta},$$

where $a_n \in \mathbb{C}$ for $-M \le n \le M$, $n \in \mathbb{N}$. $\Diamond$

Exercise 1.7. Verify that if f is a trigonometric polynomial then its coefficients $\{a_n\}_{n\in\mathbb{Z}}$ are given by: $a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta$. $\Diamond$

In both the Fourier and Taylor series, the problem is how well and in what sense we can approximate a given function by using its Taylor polynomials (very specific polynomials) or by using its *partial Fourier sums* (very specific trigonometric polynomials).

1.3. Fourier series and Fourier coefficients

We begin to answer our questions. The Fourier coefficients $\widehat{f}(n) = a_n$ are calculated using the formula suggested by Exercise 1.7:

$$\widehat{f}(n) = a_n := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta. \tag{1.5}$$

Aside 1.8. The notation $:=$ indicates that we are defining the term on the left of the $:=$, in this case a_n. Occasionally we may need to use $=:$ instead, when we are defining a quantity on the right. $\Diamond$

[5]This particular trigonometric polynomial is 2π-periodic; see Section 1.3.1.

One way to explain the appearance of formula (1.5) is to assume that the function f is equal to a trigonometric series,

$$f(\theta) = \sum_{n=-\infty}^{\infty} a_n e^{in\theta},$$

and then to proceed formally[6], operating on both sides of the equation. Multiply both sides by an exponential function, and then integrate, taking the liberty of interchanging the sum and the integral:

$$\begin{aligned} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-ik\theta}\,d\theta &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{n=-\infty}^{\infty} a_n e^{in\theta}e^{-ik\theta}\,d\theta \\ &= \sum_{n=-\infty}^{\infty} a_n \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{in\theta}e^{-ik\theta}\,d\theta \;=\; a_k. \end{aligned}$$

The last equality holds because

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{in\theta}e^{-ik\theta}\,d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(n-k)\theta}\,d\theta = \delta_{n,k}, \tag{1.6}$$

where the *Kronecker*[7] *delta* $\delta_{n,k}$ is defined by

$$\delta_{n,k} = \begin{cases} 1, & \text{if } k = n; \\ 0, & \text{if } k \neq n. \end{cases} \tag{1.7}$$

We explore the geometric meaning of equation (1.6) in Chapter 5. In language we will meet there, the equation says that the exponential functions $\{e^{in\theta}\}_{n\in\mathbb{N}}$ form an *orthonormal set*.

Aside 1.9. The usual rules of calculus apply to complex-valued functions. For example, if u and v are the real and imaginary parts of the function $f : [a,b] \to \mathbb{C}$, that is $f = u + iv$ where u and v are real-valued, then $f' := u' + iv'$. Likewise, for integration, $\int f = \int u + i\int v$. Here we are assuming that the functions u and v are differentiable in the first case, and integrable in the second. $\Diamond$

---

[6]Here the term *formally* means that we work through a computation 'following our noses', without stopping to justify every step. In this example we don't worry about whether the integrals or series converge, or whether it is valid to exchange the order of the sum and the integral. A rigorous justification of our computation could start from the observation that it is valid to exchange the sum and the integral if the Fourier series converges uniformly to f (see Chapters 2 and 4 for some definitions). Formal computations are often extremely helpful in building intuition.

[7]Named after Leopold Kronecker, German mathematician and logician (1823–1891).

Notice that in formula (1.5), a necessary condition on f for the coefficients to exist is that the complex-valued function $f(\theta)e^{-in\theta}$ should be an integrable[8] function on $[-\pi, \pi)$. In fact, if $|f|$ is integrable, then so is $f(\theta)e^{-in\theta}$ and since $\left|\int g\right| \leq \int |g|$,

$$\begin{aligned} |a_n| &= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta\right| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta)e^{-in\theta}|\,d\theta, \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta)|\,d\theta < \infty, \qquad\qquad (\text{since } |e^{-in\theta}| = 1.) \end{aligned}$$

Aside 1.10. It can be verified that the absolute value of the integral of an integrable complex-valued function is less than or equal to the integral of the absolute value of the function, the so-called *triangle inequality* for integrals, not to be confused with the triangle inequality for complex numbers: $|a + b| \leq |a| + |b|$, or the triangle inequality for integrable functions: $\int |f + g| \leq \int |f| + \int |g|$. They are all animals in the same family. $\Diamond$

Next we explore in what sense the Fourier series (1.2) approximates the original function f. We begin with some examples.

Exercise 1.11. Find the Fourier series for the trigonometric polynomial $f(\theta) = 3e^{-2i\theta} - e^{-i\theta} + 1 + e^{i\theta} - \pi e^{4i\theta} + (1/2)e^{7i\theta}$. $\Diamond$

Example 1.12. (*Periodic Ramp Function*) Consider the Fourier coefficients and the Fourier series for the function

$$f(\theta) = \theta, \qquad \text{for } -\pi \leq \theta < \pi.$$

The n^{th} Fourier coefficient is given by

$$\widehat{f}(n) = a_n := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} \theta e^{-in\theta}\,d\theta,$$

using integration by parts in the last equality. Therefore,

$$\widehat{f}(n) = \begin{cases} (-1)^{n+1}/(in), & \text{if } n \neq 0; \\ 0, & \text{if } n = 0. \end{cases}$$

[8]A function $g(\theta)$ is said to be *integrable* on $[-\pi, \pi)$ if $\int_{-\pi}^{\pi} g(\theta)\,d\theta$ is well-defined; in particular the value of the integral is a finite number. In harmonic analysis the notion of integral used is the *Lebesgue integral.* We expect the reader to be familiar with Riemann-integrable functions meaning that the function g is bounded and the integral exists in the sense of Riemann. Riemann-integrable functions are Lebesgue integrable. In Chapter 2 we will briefly review the Riemann integral. See **[Tao06a**, Chapters 11], **[Tao06b**, Chapters 19].

Thus the Fourier series of $f(\theta) = \theta$ is given by

$$f(\theta) \sim \sum_{\{n \in \mathbb{Z}: n \neq 0\}} \frac{(-1)^{n+1}}{in} e^{in\theta} = 2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(n\theta).$$

Note that both f and its Fourier series are *odd functions*. It turns out that the function and the series coincide for all $\theta \in (-\pi, \pi)$. At the endpoints $\theta = \pm\pi$ the series converges to zero (the midpoint of the jump), since $\sin(n\pi) = 0$ for all $n \in \mathbb{N}$. Notice that in this example the function and the series do not coincide at the endpoints where the jump discontinuity occurs. $\Diamond$

Exercise 1.13. Fill in the details in the integration-by-parts argument in the calculation in Example 1.12. $\Diamond$

1.3.1. 2π-Periodic functions. We are considering functions

$$f : [-\pi, \pi) \to \mathbb{C}$$

that are complex-valued, and defined on a bounded half-open interval in the real line. We extend these functions periodically to the whole of $\mathbb{R}$. For example, Figure 1.3 shows part of the graph of the periodic extension to $\mathbb{R}$ of the real-valued function $f(\theta) = \theta$ defined on $[-\pi, \pi)$ in Example 1.12.

Exercise 1.14. Use MATLAB to reproduce the plot in Figure 1.3. Then modify your MATLAB code to create a plot over the interval $[-3\pi, 3\pi]$ for the periodic extension of the function defined by $f(\theta) = \theta^2$ for $\theta \in [-\pi, \pi)$. $\Diamond$

Definition 1.15. In mathematical language, a complex valued function f defined on $\mathbb{R}$ is 2π*-periodic* if

$$f(x + 2\pi) = f(x), \qquad \text{for all } x \in \mathbb{R}.$$

The building blocks we are using are 2π-periodic functions: $\sin(n\theta)$, $\cos(n\theta)$, $e^{-in\theta}$. Finite linear combinations of 2π-periodic functions are 2π-periodic.

Let $\mathbb{T}$ denote the *unit circle*:

$$\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\} = \{z = e^{i\theta} : -\pi \leq \theta < \pi\}. \tag{1.8}$$

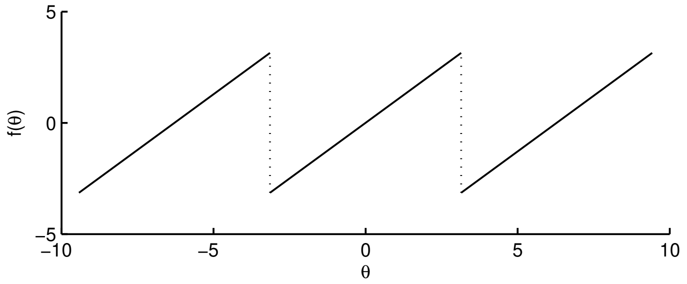

Figure 1.3. Graph of the *sawtooth* or *periodic ramp function* given by extending $f(\theta) = \theta$ periodically from the interval $[-\pi, \pi)$ to $[-3\pi, 3\pi)$.

We can identify a 2π-periodic function f on $\mathbb{R}$ with a function $g : \mathbb{T} \to \mathbb{C}$ defined on the *unit circle* $\mathbb{T}$ as follows. Given $z \in \mathbb{T}$ there is a unique $\theta \in [-\pi, \pi)$ such that $z = e^{i\theta}$. Define $g(z) := f(\theta)$. For all $n \in \mathbb{Z}$ it is also true that $z = e^{i(\theta+2n\pi)}$. The 2π-periodicity of f permits us to use all these angles in the definition of g without ambiguity: $g(z) = f(\theta + 2n\pi) = f(\theta)$.

Note that, if g is to be continuous on the unit circle $\mathbb{T}$, we will need $f(-\pi) = f(\pi)$, and this implies continuity of f on $\mathbb{R}$. Similarly for differentiability. When we say $f : \mathbb{T} \to \mathbb{C}$ and $f \in C^k(\mathbb{T})$, we mean that the function f is differentiable k times, and that f, f', $\dots$, $f^{(k)}$ are all continuous functions on $\mathbb{T}$. In words, *f is a k times continuously differentiable function from the unit circle into the complex plane.* In particular $f^{(\ell)}(-\pi) = f^{(\ell)}(\pi)$ for all $0 \leq \ell \leq k$.

If f is a 2π-periodic and integrable function, then the integral of f over each interval of length 2π takes the same value. For example,

$$\int_{-\pi}^{\pi} f(\theta)\, d\theta = \int_{0}^{2\pi} f(\theta)\, d\theta = \int_{\frac{\pi}{2}}^{\frac{5\pi}{2}} f(\theta)\, d\theta = \cdots .$$

Exercise 1.16. Verify that if f is 2π-periodic and integrable, then $\int_a^{a+2\pi} f(\theta)\, d\theta = \int_{-\pi}^{\pi} f(\theta)\, d\theta$, for all $a \in \mathbb{R}$. $\Diamond$

If f is a 2π*-periodic trigonometric polynomial of degree* M, in other words a function of the form $f(\theta) = \sum_{k=-M}^{M} a_k e^{ik\theta}$, then the Fourier series of f coincides with f itself (Exercises 1.7 and 1.18).

1.3.2. 2π**- and** L**-periodic Fourier series and coefficients.** Up to now, we have mostly considered functions defined on the interval $[-\pi, \pi)$ and their 2π-periodic extensions. Later we will also use functions defined on the unit interval $[0, 1)$, or a symmetric version of it $[-1/2, 1/2)$, and their 1-periodic extensions. More generally, we can consider functions defined on a general interval $[a, b)$ of length $L = b - a$, and extended periodically to $\mathbb{R}$ with period L.

An *L-periodic function* f defined on $\mathbb{R}$ is a function such that $f(x) = f(x + L)$ for all $x \in \mathbb{R}$. We can define the Fourier coefficients and the Fourier series for an L-periodic function f. The formulae come out slightly differently to account for the rescaling:

$$L\text{-Fourier coefficients} \quad \widehat{f}^{\,L}(n) = \frac{1}{L}\int_a^b f(\theta)e^{-2\pi in\theta/L}d\theta,$$

$$L\text{-Fourier series} \quad \sum_{n=-\infty}^{\infty} \widehat{f}^{\,L}(n)e^{2\pi in\theta/L}.$$

In the earlier case $[a, b) = [-\pi, \pi)$, we had $2\pi/L = 1$. Note that the building blocks are now the L-periodic exponential functions $e^{2\pi in\theta/L}$, while the L-periodic trigonometric polynomials are *finite* linear combinations of these building blocks.

Exercise 1.17. Verify that for each $n \in \mathbb{Z}$, the function $e^{2\pi in\theta/L}$ is L-periodic. $\Diamond$

In Chapters 7, 9, and 12, we will consider 1-periodic functions. Their Fourier coefficients and series will read

$$\widehat{f}(n) = \int_0^1 f(x)e^{-2\pi inx}\,dx, \qquad \sum_{n=-\infty}^{\infty} \widehat{f}(n)e^{2\pi inx}.$$

Why is it important to consider functions defined on a general interval $[a, b)$? A key distinction in Fourier analysis is between functions f that are defined on a bounded interval, or equivalently periodic functions on $\mathbb{R}$, and functions f that are defined on $\mathbb{R}$ but that are not periodic. For non-periodic functions on $\mathbb{R}$, it turns out that the natural "Fourier quantity" is not a Fourier series but another function, known as the *inverse Fourier transform* of f. One way to develop and understand the Fourier transform is to consider Fourier series on symmetric intervals $[-L/2, L/2)$ of length L, and then let $L \to \infty$; see Section 7.1.

Exercise 1.18. Let f be an L-periodic trigonometric polynomial of degree M, that is $f(\theta) = \sum_{n=-M}^{M} a_n e^{2\pi i n\theta/L}$. Verify that f coincides with its L-Fourier series. $\Diamond$

1.4. History, and motivation from the physical world

'It was this pliability which was embodied in Fourier's intuition, commonly but falsely called a theorem, according to which the trigonometric series *"can express any function whatever between definite values of the variable."* This familiar statement of Fourier's "theorem," taken from Thompson and Tait's "Natural Philosophy," is much too broad a one, but even with the limitations which must to-day be imposed upon the conclusion, its importance can still be most fittingly described as follows in their own words: The theorem *"is not only one of the most beautiful results of modern analysis, but may be said to furnish an indispensable instrument in the treatment of nearly recondite question [sic] in modern physics. To mention only sonorous vibrations, the propagation of electric signals along a telegraph wire, and the conduction of heat by the earth's crust, as subjects in their generality intractable without it, is to give but a feeble idea of its importance."* '

Edward B. Van Vleck
Address to the American Association
for the Advancement of Science, 1913.
Quoted in [**Bre**, p. 12].

For the last 200 years, Fourier Analysis has been of immense practical importance, both for theoretical mathematics (especially in the subfield now called harmonic analysis, but also in number theory and elsewhere), and in understanding, modeling, predicting, and controlling the behavior of physical systems. Older applications in mathematics, physics and engineering include the way that heat diffuses in a solid object, and the wave motion undergone by a string or a two-dimensional surface such as a drumhead. Recent applications include telecommunications (radio, wireless phones) and more generally data storage and transmission (the JPEG format for images on the Internet).

The mathematicians involved in the discovery, development and applications of trigonometric expansions include, in the 18th, 19th, and early 20th centuries: Brook Taylor (1685–1731), Daniel Bernoulli

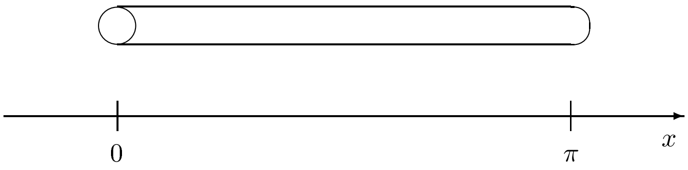

Figure 1.4. Sketch of a one-dimensional bar.

(1700–1782), Jean Le Rond d'Alembert (1717–1783), Leonhard Euler (1707–1783) (especially for the vibrating string), Joseph Louis Lagrange (1736–1813), Jean Baptiste Joseph Fourier (1768–1830), Johann Peter Gustave Lejeune Dirichlet (1805–1859), and Charles De La Vallée Poussin (1866–1962), among others. See the books by Ivor Grattan-Guinness [**Grat**, Chapter 1] and by Thomas Körner [**Kör**] for more on historical development.

We describe briefly some of the physical models that these mathematicians tried to understand, and whose solutions led them to believe that generic functions should be expandable in trigonometric series.

Temperature distribution in a one-dimensional bar. Our task is to find the temperature $u(x,t)$ in a bar of length π, where $x \in [0,\pi]$ represents the position of a point on the bar, and $t \in [0,\infty)$ represents time. See Figure 1.4. We assume that the initial temperature (when $t=0$) is given by a known function $f(x) = u(x,0)$, and that at both endpoints of the bar the temperature is held at zero, giving the boundary conditions $u(0,t) = u(\pi,t) = 0$. There are many other plausible boundary conditions.

The physical principle governing the diffusion of heat is expressed by the *Heat Equation*:

$$(\partial/\partial t)\,u = (\partial^2/\partial x^2)\,u, \quad \text{or} \quad u_t = u_{xx}. \tag{1.9}$$

We want $u(x,t)$ to be a solution of the linear partial differential equation (1.9), with initial and boundary conditions

$$u(x,0) = f(x), \qquad u(0,t) = u(\pi,t) = 0. \tag{1.10}$$

We refer to equation (1.9) together with the initial conditions and boundary conditions (1.10) as an *initial value problem*.

As Fourier did in his *Théorie Analytique de la Chaleur*, we assume (or, more accurately, guess) that there will be a *separable* solution, meaning a solution that is the product of a function of x and a function of t: $u(x,t) = \alpha(x)\beta(t)$. Then

$$(\partial/\partial t)\, u(x,t) = \alpha(x)\beta'(t), \qquad (\partial^2/\partial x^2)\, u(x,t) = \alpha''(x)\beta(t).$$

By equation (1.9), we conclude that $\alpha(x)\beta'(t) = \alpha''(x)\beta(t)$, and

$$\beta'(t)/\beta(t) = \alpha''(x)/\alpha(x). \tag{1.11}$$

Because the left side of equation (1.11) depends only on t and the right side depends only on x, both sides of the equation must be equal to a constant. Call the constant k. We can decouple equation (1.11) into a system of two linear ordinary differential equations, whose solutions we know:

$$\begin{cases} \beta'(t) = k\beta(t) & \longrightarrow \quad \beta(t) = C_1 e^{kt}, \\ \alpha''(x) = k\alpha(x) & \longrightarrow \quad \alpha(x) = C_2 \sin(\sqrt{-k}\, x) + C_3 \cos(\sqrt{-k}\, x). \end{cases}$$

The boundary conditions tell us which choices to make:

$$\begin{aligned} &\alpha(0) = 0 \quad \longrightarrow \quad \alpha(x) = C_2 \sin(\sqrt{-k}\, x), \\ &\alpha(\pi) = 0 \quad \longrightarrow \quad \sqrt{-k} = n \in \mathbb{N}, \qquad \text{and thus } k = -n^2, \text{ with } n \in \mathbb{N}. \end{aligned}$$

We have found the following separable solutions of the initial value problem. For each $n \in \mathbb{N}$, there is a solution of the form

$$u_n(x,t) = Ce^{-n^2 t} \sin(nx).$$

Aside 1.19. A note on notation: Here we are using the standard convention from analysis, where C denotes a constant that depends on constants from earlier in the argument. In this case C depends on C_1 and C_2. An extra complication in this example is that C_1, C_2 and therefore C may all depend on the natural number n; it would be more illuminating to write C_n in $u_n(x,t) = C_n e^{-n^2 t} \sin(nx)$. $\Diamond$

Finite linear combinations of these solutions are also solutions of equation (1.9), since $\partial/\partial t$ and $\partial^2/\partial x^2$ are linear. We would like to say that *infinite* linear combinations

$$u(x,t) = \sum_{n=1}^{\infty} b_n e^{-n^2 t} \sin(nx)$$

are also solutions. To satisfy the initial condition $u(x,0) = f(x)$ at time $t = 0$, we would have

$$f(x) = u(x,0) = \sum_{n=1}^{\infty} b_n \sin(nx).$$

So as long as we can write the initial temperature distribution f as a superposition of sine functions, we will have a solution to the equation. Here is where Fourier suggested that *all functions* have expansions into sine and cosine series. However, it was a long time before this statement was fully explored. See Section 3.3 for some highlights in the history of this quest.

1.5. Project: Other physical models

Note: For each project, we ask our students both to give a presentation in class and to write a report. For brevity, we mention these expectations only here in this first project.

In Table 1.1 we present several physical models schematically. In each case, one can do a heuristic analysis similar to that done in Section 1.4 for the heat equation in a one-dimensional bar. The initial conditions given for some models in the table specify the value of u at time $t = 0$, that is $u(x,0) = f(x)$. The boundary conditions specify the value of u at the ends or edges of the physical region in that model.

(a) Find solutions to one of the physical models from Table 1.1, following the approach used in Section 1.4.

(b) Search for some information about the physical model that you chose to study. When were the equations discovered? Who discovered them? What do the equation and its solutions mean in physical terms? Write a short essay to accompany the mathematics you are doing in part (a). Some references to get you started: [**Bre**], [**Grat**], [**SS03**, Chapter 1] and [**Kör**].

Table 1.1. Physical models: solutions of their partial differential equations

PDE	Boundary/Initial Conditions	Solution
Vibrating String		
$u_{tt} = u_{xx}$	$u(0,t) = u(\pi,t) = 0$	$u(x,t) = \sum_n b_n e^{int} \sin(nx)$
	$u(x,0) = \sum_n b_n \sin(nx)$	
Temperature in a bar with insulated ends		
$u_t = u_{xx}$	$u_x(0,t) = u_x(\pi,t) = 0$	$u(x,t) = \sum_n c_n e^{-n^2 t} \cos(nx)$
	$u(x,0) = \sum_n c_n \cos(nx)$	
Steady-state temperature in a disc		
$u_{rr} + r^{-1}u_r + r^{-2}u_{\theta\theta} = 0$	$u(0,\theta) = \sum_n a_n e^{in\theta}$	$u(r,\theta) = \sum_n a_n r^{\lvert n\rvert} e^{in\theta}$
Steady-state temperature in a rectangle		
$u_{xx} + u_{yy} = 0$	$u(x,0) = \sum_k A_k \sin(kx)$	$u(x,y) = \sum_k \left(\frac{\sinh[k(1-y)]}{\sinh k} A_k\right.$
	$u(x,1) = \sum_k B_k \sin(kx)$	$\left.+ \frac{\sinh(ky)}{\sinh k} B_k\right) \sin(kx)$

Chapter 2

Interlude: Analysis concepts

In this chapter we describe some useful classes of functions, sets of measure zero (Section 2.1), and various ways (known as *modes of convergence*) in which a sequence of functions can approximate another function (Section 2.2). We then describe situations in which we are allowed to interchange limiting operations (Section 2.3). Finally, we state several density results that appear in different guises throughout the book (Section 2.4). A leitmotif in analysis, in particular in harmonic analysis, is that often it is easier to prove results for a class of simpler functions that are dense in a larger class, and then to use limiting processes to pass from the approximating simpler functions to the more complicated limiting function.

This chapter should be read once to get a glimpse of the variety of function spaces and modes of convergence that can be considered. It should be revisited as necessary while the reader gets acquainted with the theory of Fourier series and integrals, and with the other types of time–frequency decompositions described in this book.

2.1. Nested classes of functions on bounded intervals

Here we introduce some classes of functions on a bounded interval I (the interval can be open, closed, or neither) and on the unit circle $\mathbb{T}$. In Chapters 7 and 8 we introduce analogous function classes on $\mathbb{R}$, and we go beyond functions and introduce *distributions* (generalized functions) in Section 8.2.

2.1.1. Riemann-integrable functions. We review the concept of the *Riemann integral*[1], following the exposition in the book by Terence Tao[2] [**Tao06a**, Chapter 11].

The development of the Riemann integral is in terms of approximation by step functions. We begin with some definitions.

Definition 2.1. Given a bounded interval I, a *partition* P of I is a finite collection of disjoint intervals $\{J_k\}_{k=1}^n$ such that $I = \bigcup_{k=1}^n J_i$. The intervals J_i are contained in I and are necessarily bounded. $\lozenge$

Definition 2.2. The *characteristic function* $\chi_J(x)$ of an interval J is defined to be the function that takes the value one if x lies in J, and zero otherwise:

$$\chi_J(x) = \begin{cases} 1, & \text{if } x \in J; \\ 0, & \text{if } x \notin J. \end{cases} \qquad \lozenge$$

The building blocks in the Riemann integration theory are finite linear combinations of characteristic functions of disjoint intervals, called *step functions*.

Definition 2.3. A function h defined on I is a *step function* if there is a partition P of the interval I, and there are real numbers $\{a_J\}_{J\in P}$, such that

$$h(x) = \sum_{J\in P} a_J \chi_J(x),$$

where $\chi_J(x)$ is the characteristic function of the interval J. $\lozenge$

[1]Named after the German mathematician Georg Friedrich Bernhard Riemann (1826–1866).

[2]Terence Tao (born 1975) is an Australian mathematician. He was awarded the Fields Medal in 2006.

For step functions there is a natural definition of the integral. Let us first consider an example.

Example 2.4. The function

$$h(x) = 2\chi_{[-5,-2)}(x) - 3\chi_{[-1,1)}(x) + 7\chi_{[6,7)}(x)$$

is a step function defined on the interval $[-5,7)$. The associated partition are the intervals $[-5,-2)$, $[-2,-1)$, $[-1,1)$, $[1,6)$ and $[6,7)$. See Figure 2.1. If we ask the reader to compute the integral of this function over the interval $[-5,7)$, she will add (with appropriate signs) the areas of the rectangles in the picture to get $(2\times 3)+(-3\times 2)+(7\times 1)=7$. The function h coincides with the following step function,

$$2\chi_{[-5,-3])}(x) + 2\chi_{(-3,-2)}(x) - 3\chi_{[-1,1)}(x) + 7\chi_{[6,7)}(x),$$

defined on the interval $[-5,7)$, with associated partition the intervals $[-5,-3]$, $(-3,-2)$, $[-2,-1)$, $[-1,1)$, $[1,6)$ and $[6,7)$. The areas under the rectangles defined by this partition also add up to 7. $\Diamond$

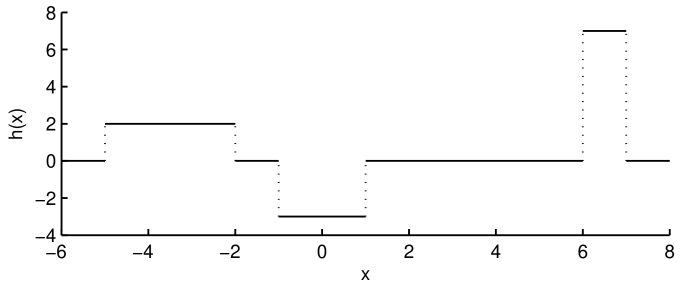

Figure 2.1. Graph of the step function in Example 2.4: $h(x) = 2\chi_{[-5,-2)}(x) - 3\chi_{[-1,1)}(x) + 7\chi_{[6,7)}(x)$.

Generalizing, we define the *integral of a step function h associated to the partition P over the interval I* by

$$\int_I h(x)\,dx := \sum_{J\in P} a_J|J|, \tag{2.1}$$

where $|J|$ denotes the length of the interval J. A given step function can be associated to more than one partition, but this definition of the integral is independent of the partition chosen.

The integral of the characteristic function of a subinterval J of I is its length $|J|$.

Exercise 2.5. Given interval $J \subset I$, show that $\int_I \chi_J(x)\,dx = |J|$. $\Diamond$

Definition 2.6. A function f is *bounded on* I if there is a constant $M > 0$ such that $|f(x)| \leq M$ for all $x \in I$. If so, we say that f is *bounded by* M. $\Diamond$

Exercise 2.7. Check that step functions are always bounded. $\Diamond$

Definition 2.8. A function $f : I \to \mathbb{R}$ is *Riemann integrable* if f is bounded on I, and the following condition holds,

$$\sup_{h_1 \leq f} \int_I h_1(x)\,dx = \inf_{h_2 \geq f} \int_I h_2(x)\,dx < \infty,$$

where the supremum and infimum are taken over step functions h_1 and h_2 such that $h_1(x) \leq f(x) \leq h_2(x)$ for all $x \in I$. We denote the common value by

$$\int_I f(x)\,dx, \quad \int_a^b f(x)\,dx, \quad \text{or simply} \quad \int_I f,$$

where a and b ($a \leq b$) are the endpoints of the interval I. We denote the class of *Riemann-integrable functions on* I by $\mathcal{R}(I)$. $\Diamond$

Riemann integrability is a property preserved under linear transformations, taking maximums and minimums, absolute values, and products. It also preserves order, meaning that if f and g are Riemann-integrable functions over I and if $f < g$ then $\int_I f < \int_I g$. Lastly, the *triangle inequality*

$$\Big| \int_I f \Big| \leq \int_I |f|$$

holds for Riemann-integrable functions. See [**Tao06a**, Section 11.4].

Exercise 2.9. (*The Riemann Integral is Linear*) Given two step functions h_1 and h_2 defined on I, verify that for c, $d \in \mathbb{R}$, the linear combination $ch_1 + dh_2$ is again a step function. Also check that $\int_I \big(ch_1(x) + dh_2(x)\big)\,dx = c\int_I h_1(x)\,dx + d\int_I h_2(x)\,dx$. Use Definition 2.8 to show that this linearity property holds for all Riemann-integrable functions (not just the step functions). **Hint:** The partitions P_1 and P_2 associated to the step functions h_1 and h_2 might

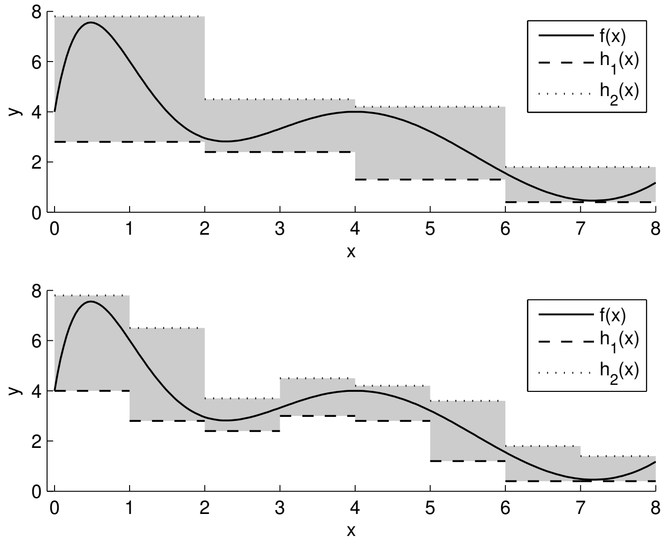

Figure 2.2. This figure illustrates the definition of the Riemann integral. We see the approximation of a function f from below by a step function h_1 and from above by a step function h_2, at a coarse scale (upper plot) and at a finer scale (lower plot). The upper and lower plots use different partitions of the interval $[0, 8]$. The shaded areas represent the integral of $h_2 - h_1$. We have used the function $f(x) = 4 + x(x - 1.5)(x - 4)^2(x - 9)\exp(-x/2.5)/12$.

be different. Try to find a partition that works for both (a *common refinement*). $\lozenge$

Exercise 2.10. (*Riemann Integrability is Preserved by Products*) Show that if f, $g : I \to \mathbb{R}$ are Riemann integrable then their product fg is Riemann integrable. $\lozenge$

One can deduce the following useful proposition from Definition 2.8.

Proposition 2.11. *A bounded function f is Riemann integrable over an interval I if and only if for each $\varepsilon > 0$ there are step functions h_1 and h_2 defined on I such that $h_1(x) \leq f(x) \leq h_2(x)$ for all $x \in I$,* *and such that the integral of their difference is bounded by ε. More precisely,*

$$0 \leq \int_I \big(h_2(x) - h_1(x)\big)\, dx < \varepsilon.$$

In particular, if f is Riemann integrable then there exists a step function h such that

$$\int_I |f(x) - h(x)|\, dx < \varepsilon.$$

See Figure 2.2, in which the shaded area represents the integral of the difference of the two step functions. Notice how taking a finer partition allows us to use step functions that approximate f more closely, thereby reducing the shaded area.

Aside 2.12. Readers familiar with the definition of the Riemann integral via Riemann sums may like to observe that the integral of a step function that is below (respectively above) a given function f is always below a corresponding *lower* (respectively above a corresponding *upper*) *Riemann sum*[3] for f. $\Diamond$

Our first observation is that if f is Riemann integrable on $[-\pi, \pi)$, then the zeroth Fourier coefficient of f is well defined:

$$a_0 = \widehat{f}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)\, d\theta.$$

We say that a *complex-valued function* $f : I \to \mathbb{C}$ is *Riemann integrable* if its real and imaginary parts are Riemann integrable. If f is Riemann integrable, then the complex-valued functions $f(\theta)e^{-in\theta}$ are Riemann integrable for each $n \in \mathbb{Z}$, and all the Fourier coefficients $a_n = \widehat{f}(n)$, $n \neq 0$, are also well defined.

Exercise 2.13. Verify that if $f : [-\pi, \pi) \to \mathbb{C}$ is Riemann integrable then the function $g : [-\pi, \pi) \to \mathbb{C}$ defined by $g(\theta) = f(\theta)e^{-in\theta}$ is Riemann integrable. $\Diamond$

[3]A *Riemann sum* associated to a bounded function $f : I \to \mathbb{R}$ and to a given partition P of the interval I is a real number $R(f, P)$ defined by $R(f, P) = \sum_{J \in P} f(x_J)\,|J|$, where x_J is a point in J for each $J \in P$. The *upper (resp. lower) Riemman sum*, $U(f, P)$ (resp. $L(f, P)$ associated to f and P is obtained similarly replacing $f(x_J)$ by $\sup_{x \in J} f(x)$ (resp. $\inf_{x \in J} f(x)$). Note that since f is bounded both the infimum and supremum over J exist.

Here are some examples of familiar functions that are Riemann integrable and for which we can define Fourier coefficients. See Theorem 2.32 toward the end of the Chapter, for a complete characterization of Riemann-integrable functions.

Example 2.14. A uniformly continuous functions on an interval is Riemann integrable. In particular a continuous function on a closed interval is Riemann integrable; see [**Tao06a**, Section 11.5]. A monotone[4] bounded function on an interval is also Riemann integrable; see [**Tao06a**, Section 11.6]. $\Diamond$

Aside 2.15. We are assuming the reader is familiar with the notions of continuity and uniform continuity. $\Diamond$

Here is an example of a function that is not Riemann integrable.

Example 2.16. (*Dirichlet's Example*[5]) The function

$$f(x) = \begin{cases} 1, & \text{if } x \in [a,b] \cap \mathbb{Q}; \\ 0, & \text{if } x \in [a,b] \setminus \mathbb{Q} \end{cases}$$

is bounded (by one), but it is not Riemann integrable on $[0,1]$. It is also discontinuous everywhere. $\Diamond$

Exercise 2.17. Show that the Dirichlet function is not Riemann integrable. $\Diamond$

2.1.2. Lebesgue integrable functions and L^p spaces. There are other more general notions of integral; see for example the book by Robert G. Bartle [**Bar01**]. One such notion is that of the *Lebesgue*[6] *integral*. The Lebesgue integral is studied in depth in courses on *measure theory*. In this section we simply aim to give some ideas about what the Lebesgue integral is and what the L^p spaces are, and we skim over the many subtleties of measure theory[7].

[4]A function $f : I \to \mathbb{R}$ is *monotone* if it is either increasing on the whole of I, or decreasing on the whole of I; that is, either for all $x, y \in I$, $x \leq y$, we have $f(x) \leq f(y)$, or for all $x, y \in I$, $x \leq y$, we have $f(x) \geq f(y)$.

[5]Named after the German mathematician Johann Peter Gustav Lejeune Dirichlet (1805–1859).

[6]Named after the French mathematician Henri Léon Lebesgue (1875–1941).

[7]For a brief introduction see [**Tao06b**, Chapters 18, 19]. For more in-depth presentations at a level intermediate between advanced undergraduate and graduate students see [**SS05**] or [**Bar66**]. Graduate textbooks include the books by Gerald B. Folland [**Fol**] and by Halsey L. Royden [**Roy**].

Every Riemann-integrable function on a bounded interval I is also Lebesgue-integrable on I, and the values of its Riemann and Lebesgue integrals are the same. The Dirichlet function in Example 2.16 is Lebesgue integrable but not Riemann integrable, with Lebesgue integral equal to zero; see Remark 2.37. In practice when you see an integral, it will usually suffice to think of Riemann integrals even if sometimes we really mean Lebesgue integrals.

In parallel to the definition of the Riemann integral just presented, we sketch the definition of the Lebesgue integral over the bounded interval I. We use as scaffolding the so-called simple functions instead of step functions.

Simple functions are finite linear combinations of characteristic functions of disjoint *measurable sets*. The *characteristic*[8] *function* of a set $A \subset \mathbb{R}$ is defined to be

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A; \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

In particular, intervals are measurable sets, and we can assign to them a *measure*, namely their length. However there are sets that are non-measurable, and we need to exclude such sets. Once we assign a measure to each measurable set, we can define the Lebesgue integral of the characteristic function of a measurable set to be the measure of the set. The Lebesgue integral of a simple function is then defined to be the corresponding finite linear combination of the measures of the underlying sets, exactly as in (2.1), except that now the sets J are disjoint measurable subsets of the measurable set I, and the length $|J|$ is replaced by the measure $m(J)$ of the set J. One can then define Lebesgue integrability of a function $f : I \to \mathbb{R}$ in analogy to Definition 2.8, replacing the boundedness condition by requiring the function to be measurable on I, and replacing step functions by simple functions. A function $f : I \to \mathbb{R}$ is *measurable* if the pre-image under f of each measurable subset of $\mathbb{R}$ is a measurable subset of I. It all boils down to understanding measurable sets and their measures.

[8]In statistics and probability the characteristic function of a set A is called the *indicator function* of the set, and denoted $\mathbb{I}_A$. The term *characteristic function* refers to the Fourier transform of a probability density (so-called because it *characterizes* the probability distribution).

In this book we do not define measurable sets, or their measure, except for sets of measure zero in $\mathbb{R}$ (which we discuss in Section 2.1.4), and countable unions of disjoint intervals (whose measure is given, quite naturally, by the sum of the lengths of the intervals).

A measurable function is said to belong to $L^1(I)$ if

$$\int_I |f(x)|\,dx < \infty,$$

where the integral is in the sense of Lebesgue. In fact, there is a continuum of *Lebesgue spaces*, the L^p *spaces*. The space $L^p(I)$ is defined for each real number p such that $1 \le p < \infty$ as the collection of measurable functions f on I, such that $|f|^p \in L^1(I)$. For $p = \infty$ we define $L^\infty(I)$ as the collection of measurable functions f on I, such that f is "essentially bounded".

All the functions involved are assumed to be measurable. Let us denote by $\mathcal{M}(I)$ the collection of measurable functions defined on I. In Definition 2.33 we state precisely what *essentially bounded* means. For now, just think of L^∞ functions as *bounded* functions on I. Let us denote by $B(I)$ the space of bounded measurable functions on I. There are essentially bounded functions that are not bounded, so $B(I) \subsetneq L^\infty(I)$.

These spaces are *normed spaces*[9] (after defining properly what it means for a function to be zero in L^p, see Remark 2.37) with the L^p norm defined by

$$\|f\|_{L^p(I)} := \Big(\int_I |f(x)|^p\,dx\Big)^{1/p}, \quad 1 \le p < \infty, \tag{2.3}$$

$$\|f\|_{L^\infty(I)} := \operatorname*{ess\,sup}_{x\in I} |f(x)|, \quad \|f\|_{B(I)} := \sup_{x\in I} |f(x)|. \tag{2.4}$$

In Definition 2.34 we state precisely what *essential supremum of* f means.

The L^p spaces are *complete normed spaces*. Such spaces are so common and important that they have a special name, *Banach*[10] spaces. Furthermore the Lebesgue spaces $L^p(I)$ are the *completion*

[9] See Appendix A for the definition.

[10] Named after the Polish mathematician Stefan Banach (1892–1945). See Appendix A for the definition of a complete normed space. One learns about Banach spaces in a course on functional analysis. Graduate textbooks include **[Fol]** and **[Sch]**.

of $\mathcal{R}(I)$ in the L^p norm. This fact is so important that we state it (without proof) as a theorem, for future reference.

Theorem 2.18. *For each p with $1 \leq p \leq \infty$, the space $L^p(I)$ is the completion of the Riemann-integrable functions $\mathcal{R}(I)$ in the L^p norm. In particular, every* Cauchy sequence[11] *in $L^p(I)$ converges to a function in $L^p(I)$, $\mathcal{R}(I)$ is* dense[12] *in $L^p(I)$.*

Aside 2.19. The reader may be familiar with the construction of the real numbers $\mathbb{R}$ as the completion of the rational numbers $\mathbb{Q}$ in the Euclidean metric $d(x, y) := |x - y|$. One can think of $\mathcal{R}(I)$ as playing the role of $\mathbb{Q}$ and $L^p(I)$ as playing the role of $\mathbb{R}$. There is a procedure that allows the completion of any metric space, mirroring the procedure used in the construction of the real numbers. See for example [**Tao06b**, Exercise 12.4.8]. $\Diamond$

Among the L^p spaces, both L^1 and L^∞ play special roles. The space of *square-integrable* functions

$$L^2(I) := \left\{ f : I \to \mathbb{C} : \int_I |f(x)|^2 \, dx < \infty \right\}$$

is also very important. The term *square-integrable* emphasizes that the integral of the square of (the absolute value of) $f \in L^2(I)$ is finite. Functions in $L^2(I)$ are also said to have *finite energy*.

The L^2 norm is induced by the *inner product*

$$\langle f, g \rangle_{L^2(I)} := \int_I f(x)\overline{g(x)} \, dx, \tag{2.5}$$

meaning that $\|f\|_{L^2(I)} = \sqrt{\langle f, f \rangle_{L^2(I)}}$.

The space $L^2(I)$ of square-integrable functions over I is a *complete inner-product vector space* (a *Hilbert*[13] *space*), and so it has geometric

---

[11] Named after the French mathematician Augustin-Louis Cauchy (1789–1857). The idea of a Cauchy sequence is that the terms must get arbitrarily close to each other as $n \to \infty$. In our setting, the sequence $\{f_n\} \subset L^p(I)$ is a *Cauchy sequence in* $L^p(I)$ if for each $\varepsilon > 0$ there is an $N > 0$ such that for all n, $m > N$, $\|f_n - f_m\|_{L^p(I)} < \varepsilon$. Every convergent sequence is Cauchy, but the converse does not hold unless the space is complete.

[12] Meaning that we can approximate functions in $L^p(I)$ with Riemann integrable functions in the L^p norm. More precisely, given $f \in L^p(I)$, $\epsilon > 0$, there is $g \in \mathcal{R}(I)$ such that $\|f - g\|_{L^p(I)} \leq \epsilon$. See Section 2.4.

[13] Named after the German mathematician David Hilbert (1862–1943).

properties very similar to those of n-dimensional Euclidean space $\mathbb{R}^n$. In particular the notion of *orthogonality* is very important[14].

In Chapter 5 we will be especially interested in the Hilbert space $L^2(\mathbb{T})$ of square-integrable functions on $\mathbb{T}$, with the L^2 norm

$$\|f\|_{L^2(\mathbb{T})} := \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta)|^2\, d\theta\right)^{1/2}$$

induced by the inner product

$$\langle f, g\rangle := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\overline{g(\theta)}\, d\theta.$$

The factor $1/(2\pi)$ is just a normalization constant that forces the trigonometric functions $e_n(\theta) := e^{in\theta}$ to have norm one in $L^2(\mathbb{T})$.

Exercise 2.20. Verify that $\langle e_n, e_m\rangle = \delta_{n,m}$. In particular, the norm of e_n is one: $\|e_n\|_{L^2(\mathbb{T})} = 1$. In the language of Chapter 5, the trigonometric system is an *orthonormal* system. ◊

2.1.3. Ladder of function spaces on $\mathbb{T}$. We have been concerned with function spaces defined by integrability properties. Earlier we encountered function spaces defined by differentiability properties. It turns out that for functions defined on a given closed bounded interval, these function spaces are nicely nested. The inclusions are summarized in Figure 2.3 in page 33.

Continuous functions on a closed bounded interval $I = [a, b]$ are Riemann integrable, and hence Lebesgue integrable. The collection of continuous functions on $I = [a, b]$ is denoted by $C([a, b])$. Next, $C^k([a, b])$ denotes the collection of k times continuously differentiable functions on $[a, b]$, in other words functions whose k^{th} derivatives exist and are continuous. For convenience we sometimes use the expression 'f is C^k' to mean '$f \in C^k$'. Finally, $C^\infty([a, b])$ denotes the collection of functions on $[a, b]$ that are differentiable infinitely many times. (At the endpoints a and b, we mean continuous or differentiable from the right and from the left respectively.) Notice that there exist functions that are differentiable but not continuously differentiable.

[14]See Appendix A for a more detailed description of inner-product vector spaces and Hilbert spaces. Another reference is [**SS05**, Chapters 4, 5].

Exercise 2.21. Verify that the function defined by $f(x) = x^2 \sin(1/x)$ for $x \neq 0$, $f(0) = 0$, is differentiable everywhere but its derivative is not continuous at $x = 0$. $\lozenge$

Recall that in Section 1.3.2 we discussed periodic functions on the unit circle $\mathbb{T} = [-\pi, \pi)$, as well as functions that are continuously differentiable on $\mathbb{T}$. In that context we meant that $f : \mathbb{T} \to \mathbb{C}$ is C^k if f is k times continuously differentiable on $\mathbb{T}$ and if, when f is extended periodically, $f^{(j)}(-\pi) = f^{(j)}(\pi)$ for all $0 \leq j \leq k$. This boundary condition implies that when we restrict f to the closed interval $[-\pi, \pi]$ it is C^k, in particular it is bounded and Riemann integrable on $[-\pi, \pi)$. A 2π-periodic function is integrable on $[-\pi, \pi)$ if and only if it is integrable on $[-\pi, \pi]$. Define $L^p(\mathbb{T})$ to be $L^p([-\pi, \pi])$.

For complex-valued functions defined on $\mathbb{T}$, or on any closed bounded interval $I = [a, b]$, all these spaces are nicely nested. There is a ladder of function spaces in $\mathbb{T}$, running from the small space $C^\infty(\mathbb{T})$ of smooth functions up to the large space $L^1(\mathbb{T})$ of Lebesgue-integrable functions and beyond to the even larger space $\mathcal{M}$ of measurable functions. This is the top 3/4 of the ladder in Figure 2.3.

Earlier we encountered a different family of function spaces on $\mathbb{T}$, namely the collections of 2π-periodic trigonometric polynomials of degree at most N, for $N \in \mathbb{N}$. We denote these spaces by $\mathcal{P}_N(\mathbb{T})$. More precisely, $\mathcal{P}_N(\mathbb{T}) := \Big\{ \sum_{|n| \leq N} a_n e^{in\theta} : a_n \in \mathbb{C},\ |n| \leq N \Big\}$.

These spaces are clearly nested and increasing as N increases, and they are certainly subsets of $C^\infty(\mathbb{T})$. Thus we can extend the ladder to one starting at $\mathcal{P}_0(\mathbb{T})$ (the constant functions) and climbing all the way to $C^\infty(\mathbb{T})$, by including the nested spaces of trigonometric polynomials. This is the bottom 1/4 of the ladder in Figure 2.3.

Remark 2.22. In Figure 2.3, p_1 and p_2 are real numbers such that $1 < p_1 < 2 < p_2 < \infty$, and N_1 and N_2 are natural numbers such that $N_1 \leq N_2$. The blank box appears because there is no commonly used symbol for the class of differentiable functions. Notice that as we move down the ladder into the smaller function classes, the functions become better behaved. For instance, the L^p spaces on $\mathbb{T}$ are nested and decreasing as the real number p increases from 1 to ∞ (Exercise 2.26). $\lozenge$

$\mathcal{M}(\mathbb{T})$	Measurable functions on $\mathbb{T}$
$\cup$	
$L^1(\mathbb{T})$	Lebesgue integrable: $\int_{\mathbb{T}} \lvert f(x)\rvert\, dx < \infty$
$\cup$	
$L^{p_1}(\mathbb{T})$	$\int_{\mathbb{T}} \lvert f(x)\rvert^{p_1}\, dx < \infty\,, \quad 1 < p_1 < 2$
$\cup$	
$L^2(\mathbb{T})$	square integrable: $\int_{\mathbb{T}} \lvert f(x)\rvert^2\, dx < \infty$
$\cup$	
$L^{p_2}(\mathbb{T})$	$\int_{\mathbb{T}} \lvert f(x)\rvert^{p_2}\, dx < \infty\,, \quad 2 < p_1 < \infty$
$\cup$	
$L^\infty(\mathbb{T})$	essentially bounded: $\operatorname{ess\,sup}_{x\in\mathbb{T}} \lvert f(x)\rvert < \infty$
$\cup$	
$B(\mathbb{T})$	bounded/measurable: $\lvert f(x)\rvert \le M$ for all $x \in \mathbb{T}$
$\cup$	
$\mathcal{R}(\mathbb{T})$	Riemann integrable
$\cup$	
$\mathcal{C}(\mathbb{T})$	continuous
$\cup$	
$\boxed{\phantom{C^1(\mathbb{T})}}$	differentiable
$\cup$	
$C^1(\mathbb{T})$	continuously differentiable: f' exists and is continuous
$\cup$	
$C^2(\mathbb{T})$	twice continuously differentiable: f, f' exist and are continuous
$\cup$	
$\vdots$	$\vdots$
$\cup$	
$C^k(\mathbb{T})$	k times continuously differentiable: f, f', $\ldots$, $f^{(k)}$ exist and are continuous
$\cup$	
$\vdots$	$\vdots$
$\cup$	
$C^\infty(\mathbb{T})$	infinitely often differentiable: f, f', f'', $\ldots$, exist and are continuous
$\cup$	
$\vdots$	$\vdots$
$\cup$	
$\mathcal{P}_{N_2}(\mathbb{T})$	trigonometric polynomials of degree $N_2 > N_1$
$\cup$	
$\mathcal{P}_{N_1}(\mathbb{T})$	trigonometric polynomials of degree $N_1 > 0$
$\cup$	
$\vdots$	$\vdots$
$\cup$	
$\mathcal{P}_0(\mathbb{T})$	constant functions

Figure 2.3. Ladder of nested classes of functions $f : \mathbb{T} \to \mathbb{C}$. See Remark 2.22.

Two warnings are in order. First, when the underlying space is changed from the circle $\mathbb{T}$ to another space X, the L^p spaces need not be nested: for real numbers p, q with $1 \le p \le q \le \infty$, $L^p(X)$ need not contain $L^q(X)$. For example, the L^p spaces on the real line $\mathbb{R}$ are not nested. (By contrast, the C^k classes are always nested: $C^k(X) \supset C^{k+1}(X)$ for all spaces X in which differentiability is defined, and for all $k \in \mathbb{N}$.) Second, there are important function classes that do not fall into the nested chain of classes shown here.

The step functions defined in Section 2.1.1 are Riemann integrable, although most of them have discontinuities. Our ladder has a branch:

$$\{\text{ step functions on } I\} \subset \mathcal{R}(I).$$

The simple functions mentioned in Section 2.1.2 are bounded functions that are not necessarily Riemann integrable. However, they are Lebesgue integrable, and in fact simple functions are in $L^p(I)$ for each p such that $1 \le p \le \infty$. Our ladder has a second branch:

$$\{\text{ simple functions on } I\} \subset B(I).$$

Exercise 2.23. Compute the L^p norm of a step function defined on an interval I. Compute the L^p norm of a simple function defined on an interval I. Use the notation $m(A)$ for the measure of a measurable set $A \subset I$. $\Diamond$

We will encounter other classes of functions on $\mathbb{T}$, $[a, b]$, or $\mathbb{R}$, such as functions that are *piecewise continuous* or *piecewise smooth*, *monotone functions*, functions of *bounded variation*, *Lipschitz*[15] *functions*, and *Hölder-continuous*[16] *functions*. We define what we mean as they appear in the text. As you encounter these spaces, you are encouraged to try to fit them into the ladder of spaces in Figure 2.3, which will start to look more like a tree with branches and leaves.

The following example and exercises justify some of the inclusions in the ladder of function spaces. Operate with the Lebesgue integrals as you would with Riemann integrals. Try to specify which properties of the integral you are using.

[15]Named after the German mathematician Rudolf Otto Sigismund Lipschitz (1832–1903).

[16]Named after the German mathematician Otto Ludwig Hölder (1859–1937).

Example 2.24. Let I be a bounded interval. If $f : I \to \mathbb{C}$ is bounded by $M > 0$ and $f \in L^p(I)$, then for all q such that $p \leq q < \infty$,

$$\int_I |f(x)|^q \, dx \leq M^{q-p} \int_I |f(x)|^p \, dx < \infty,$$

and so in this situation the L^q-norm is controlled by the L^p norm, and $f \in L^q(I)$. In the language of norms, the preceding inequality can be written as $\|f\|^q_{L^q(I)} \leq \|f\|^{q-p}_{L^\infty(I)} \|f\|^p_{L^p(I)}$, since $\|f\|_{L^\infty(I)}$ is the infimum of the numbers M that bound f. $\lozenge$

The following estimate, called *Hölder's inequality*, is an indispensable tool in the study of L^p spaces.

Lemma 2.25 (Hölder's Inequality)**.** *Suppose I is a bounded interval, $1 \leq s \leq \infty$, $1/s + 1/t = 1$, $f \in L^s(I)$, and $g \in L^t(I)$. Then the product fg is integrable, and furthermore*

$$\left| \int_I f(x) g(x) \, dx \right| \leq \|f\|_{L^s(I)} \|g\|_{L^t(I)}; \tag{2.6}$$

Hölder's inequality reappears in later chapters, and it is discussed and proved, together with other very important inequalities, at the end of the book. The special case when $s = t = 2$ is the *Cauchy–Schwarz*[17] *inequality.* Inequalities are a fundamental part of analysis, and we will see them over and over again throughout the book.

Exercise 2.26. Let I be a bounded interval. Show that if $1 \leq p \leq q < \infty$ and $h \in L^q(I)$ then $h \in L^p(I)$. Moreover, show that if $h \in L^q(I)$ then $\|h\|_{L^p(I)} \leq |I|^{\frac{1}{p} - \frac{1}{q}} \|h\|_{L^q(I)}$. Thus for those exponents, $L^q(I) \subset L^p(I)$. **Hint:** Apply Hölder's inequality on I, with $s = q/p > 1$, $f = |h|^p$, $g = 1$, and $h \in L^q(I)$. $\lozenge$

2.1.4. Sets of measure zero. For many purposes we often need only that a function has a given property except on a particular set of points, provided such exceptional set is small. We introduce the appropriate notion of smallness, given by the concept of a *set of measure zero*.

[17]This inequality is named after the French mathematician Augustin-Louis Cauchy (1789–1857), and the German mathematician Hermann Amandus Schwarz (1843–1921). The Cauchy–Schwarz inequality was discovered 25 years earlier by the Ukranian mathematician Viktor Yakovlevich Bunyakovskii (1804–1889), and it is also known as the Cauchy–Bunyakovskii–Schwarz inequality.

Definition 2.27. A set $E \subset \mathbb{R}$ has *measure zero* if, for each $\varepsilon > 0$, there are open intervals I_j such that E is covered by the collection $\{I_j\}_{j\in\mathbb{N}}$ of intervals, and the total length of the intervals is less than ε:

$$\text{(i)}\quad E \subset \bigcup_{j=1}^{\infty} I_j, \quad \text{and} \quad \text{(ii)}\quad \sum_{j=1}^{\infty} |I_j| < \varepsilon.$$

We say that a given property holds *almost everywhere* if the set of points where the property does not hold has measure zero. We use the abbreviation *a.e.* for *almost everywhere*[18]. ◊

In particular, every countable set has measure zero. There are uncountable sets that have measure zero. The most famous example is the *Cantor*[19] *set*, discussed briefly in Example 2.29.

Exercise 2.28. Show that finite subsets of $\mathbb{R}$ have measure zero. Show that countable subsets of $\mathbb{R}$ have measure zero. In particular the set $\mathbb{Q}$ of rational numbers has measure zero. ◊

Example 2.29. (*The Cantor set*) The Cantor set C is obtained from the closed interval $[0,1]$ by a sequence of successive deletions of open intervals called the *middle thirds*, as follows. Remove the middle-third interval $(1/3, 2/3)$ of $[0,1]$. We are left with two closed intervals: $[0,1/3]$ and $[2/3,1]$. Remove the two middle-third intervals of the remaining closed intervals. We are left with four closed intervals. This process can be continued indefinitely, so that the set A of all points removed from the closed interval $[0,1]$ is the union of a collection of disjoint open intervals. Thus $A \subset [0,1]$ is an open set. The Cantor set C is defined to be the set of points that remains after A is deleted: $C := [0,1] \setminus A$. The Cantor set C is an uncountable set.However, C has measure zero, because its complement A has *measure one*[20]. ◊

[18]In probability theory the abbreviation *a.s.* for *almost surely* is used instead; in French *p.p.* for *presque partout*; in Spanish *c.s.* for *casi siempre*, etc. etc.

[19]Named after the German mathematician Georg Ferdinand Ludwig Philipp Cantor (1845–1918).

[20]Recall that a natural way of defining the length or measure of a subset A of $\mathbb{R}$ which is a countable union of disjoint intervals $\{I_j\}$ is to declare the measure of A to be the sum of the lengths of the intervals I_j. Also the following property holds, if A is a measurable set in X then its complement $X\setminus A$ is also measurable, furthermore, $m(X\setminus A) = m(X) - m(A)$.

Exercise 2.30. Show that the Cantor set C is closed. Use Definition 2.27 to show that C has measure zero. Alternatively, show that the measure of its complement is 1. Show that C is uncountable. $\Diamond$

Aside 2.31. We assume the reader is familiar with basic notions of point set topology on $\mathbb{R}$, such as the concepts of *open* and *closed* sets in $\mathbb{R}$. To review these ideas, see [**Tao06b**, Chapter 12]. $\Diamond$

Here is an important characterization of Riemann-integrable functions; it appeared in Lebesgue's doctoral dissertation. A proof can be found in Stephen Abbott's book [**Abb**, Section 7.6].

Theorem 2.32 (Lebesgue's Theorem, 1901)**.** *A bounded function $f : I \to \mathbb{C}$ is Riemann integrable if and only if f is continuous almost everywhere.*

We can now define what the terms *essentially bounded* and *essential supremum* mean.

Definition 2.33. A function $f : \mathbb{T} \to \mathbb{C}$ is *essentially bounded* if it is bounded except on a set of measure zero. More precisely, there exists a constant $M > 0$ such that $|f(x)| \leq M$ a.e. in I. $\Diamond$

Definition 2.34. Let $f : I \to \mathbb{C}$ be an essentially bounded function. The *essential supremum* $\operatorname{ess\,sup}_{x\in I} |f(x)|$ of f is the smallest constant M such that $|f(x)| \leq M$ a.e. on the interval I. In other words, $\operatorname{ess\,sup}_{x\in I} |f(x)| := \inf\{M \geq 0 : |f(x)| \leq M \text{ a.e. in } I\}$. $\Diamond$

Exercise 2.35. Verify that if f is bounded on I then f is essentially bounded. In other words, the space $\mathcal{B}(I)$ of bounded functions on I is a subset of $L^\infty(I)$. Show that it is a proper subset. Furthermore, verify that if f is bounded then $\operatorname{ess\,sup}_{x\in I} |f(x)| = \sup_{x\in I} |f(x)|$. $\Diamond$

Exercise 2.36. Verify that Exercise 2.26 holds true when $q = \infty$. $\Diamond$

Remark 2.37. For the purposes of Lebesgue integration theory, functions that are equal almost everywhere are regarded as being the same. To be really precise, one should talk about equivalence classes of L^p integrable functions: f is equivalent to g if and only if $f = g$ a.e. In that case $\int_I |f - g|^p = 0$, and if $f \in L^p(I)$ then so is g. We say $f = g$ in $L^p(I)$ if $f = g$ a.e. In that case, $\int_I |f|^p = \int_I |g|^p$. In particular the Dirichlet function g in Example 2.16 is equal a.e. to the zero function, which is in $L^p([0,1])$. Hence the Dirichlet function g is in $L^p([0,1])$, its integral in the Lebesgue sense vanishes, and its L^p norm is zero for all $1 \leq p \leq \infty$. $\Diamond$

2.2. Modes of convergence

Given a bounded interval I and functions g_n, $g : I \to \mathbb{C}$, what does it mean to say that $g_n \to g$? In this section we discuss six different ways in which the functions g_n could approximate g: pointwise; almost everywhere; uniformly; uniformly outside a set of measure zero; in the sense of convergence in the mean (convergence in L^1); in the sense of mean-square convergence (convergence in L^2). The first four of these modes of convergence are defined in terms of the behavior of the functions g_n and g at individual points x. Therefore in these modes, the interval I can be replaced by any subset X of the real numbers. The last three of these modes are defined in terms of integrals of powers of the difference $|g_n - g|$. In these cases I can be replaced by any set X of real numbers for which the integral over the set is defined. Some of these modes of convergence are stronger than others, uniform convergence being the strongest. Figure 2.4 in page 41 illustrates the interrelations between five of these modes of convergence.

In Chapter 3, we will be concerned with pointwise and uniform convergence of Fourier series, and in Chapter 5 we will be concerned with mean-square convergence of Fourier series. At the end of the book, we come full circle and show that convergence in $L^p(\mathbb{T})$ of Fourier series is a consequence of the boundedness in $L^p(\mathbb{T})$ of the *periodic Hilbert transform.*

See [**Bar66**, pp.68–72] for yet more modes of convergence, such as convergence in measure and almost uniform convergence (not to be confused with uniform convergence outside a set of measure zero).

We begin with the four modes of convergence that are defined in terms of individual points x. In what follows X is a subset of the real numbers.

Definition 2.38. A sequence of functions $g_n : X \to \mathbb{C}$ *converges pointwise* to $g : X \to \mathbb{C}$ if for each $x \in X$, $\lim_{n\to\infty} g_n(x) = g(x)$. Equivalently, g_n converges pointwise to g if for each $\varepsilon > 0$ and for each $x \in X$ there is an $N = N(\varepsilon, x) > 0$ such that for all $n > N$,

$$|g_n(x) - g(x)| < \varepsilon. \qquad \Diamond$$

Almost-everywhere convergence is a little weaker than pointwise convergence.

Definition 2.39. A sequence of functions $g_n : X \to \mathbb{C}$ *converges almost everywhere*, or *pointwise a.e.*, to $g : X \to \mathbb{C}$ if it converges pointwise to g except on a set of measure zero: $\lim_{n\to\infty} g_n(x) = g(x)$ a.e. in X. Equivalently, g_n converges a.e. to g if there exists a set E of measure zero, $E \subset X$, such that for each $x \in X \setminus E$ and for each $\varepsilon > 0$ there is an $N = N(\varepsilon, x) > 0$ such that for all $n > N$,

$$|g_n(x) - g(x)| < \varepsilon. \qquad \Diamond$$

Uniform convergence is stronger than pointwise convergence. Informally, given a fixed ε, for uniform convergence the same N must work for all x, while for pointwise convergence we may use different numbers N for different points x.

Definition 2.40. A sequence of bounded functions $g_n : X \to \mathbb{C}$ *converges uniformly* to $g : X \to \mathbb{C}$ if given $\varepsilon > 0$ there is an $N = N(\varepsilon) > 0$ such that for all $n > N$,

$$|g_n(x) - g(x)| < \varepsilon \quad \text{for all } x \in X. \qquad \Diamond$$

Here is a slightly weaker version of uniform convergence.

Definition 2.41. A sequence of L^∞-functions $g_n : X \to \mathbb{C}$ *converges uniformly outside a set of measure zero*, or *in* L^∞, to $g : X \to \mathbb{C}$ if

$$\lim_{n\to\infty} \|g_n - g\|_{L^\infty(X)} = 0.$$

In this case, we write $g_n \to g$ in $L^\infty(X)$. $\Diamond$

Remark 2.42. If the functions g_n and g are continuous and bounded on $X \subset \mathbb{R}$, then $\|g_n - g\|_{L^\infty(X)} = \sup_{x\in X} |g_n(x) - g(x)|$, and in this case convergence in $L^\infty(X)$ and uniform convergence coincide. This is why the L^∞-norm for continuous functions is often called the *uniform norm*. The metric space of continuous functions over a closed bounded interval with the metric induced by the uniform norm is a complete metric space (a Banach space); see [**Tao06b**, Section 4.4]. In particular, if the functions g_n in Definition 2.40 are continuous and converge uniformly, then the limiting function is itself continuous; see Theorem 2.58. $\Diamond$

We move to those modes of convergence that are defined via integrals. The two most important are convergence in L^1 and in L^2. The integral used in these definitions is the Lebesgue integral, and by integrable we mean Lebesgue integrable. Here we restrict to integrals on bounded intervals I.

Definition 2.43. A sequence of integrable functions $g_n : I \to \mathbb{C}$ *converges in the mean*, or *in* $L^1(I)$, to $g : I \to \mathbb{C}$ if for each $\varepsilon > 0$ there is a number $N > 0$ such that for all $n > N$

$$\int_I |g_n(x) - g(x)|\, dx < \varepsilon.$$

Equivalently, g_n converges in L^1 to g if $\lim_{n\to\infty} \|g_n - g\|_{L^1(I)} = 0$. We write $g_n \to g$ in $L^1(I)$. $\Diamond$

Definition 2.44. A sequence of square-integrable functions $g_n : I \to \mathbb{C}$ *converges in mean-square*, or *in* $L^2(I)$, to $g : I \to \mathbb{C}$ if for each $\varepsilon > 0$ there is an $N > 0$ such that for all $n > N$,

$$\int_I |g_n(x) - g(x)|^2\, dx < \varepsilon.$$

Equivalently, g_n converges in $L^2(I)$ to g if $\lim_{n\to\infty} \|g_n - g\|_{L^2(I)} = 0$. We write $g_n \to g$ in $L^2(I)$. $\Diamond$

Convergence in $L^1(I)$ and in $L^2(I)$ are particular cases of *convergence in* $L^p(I)$, defined replacing $L^2(I)$ by $L^p(I)$ in Definition 2.44.

We have already noted that the L^p spaces are complete normed spaces, also known as Banach spaces. Therefore in Definitions 2.43 and 2.44 the limit functions must be in $L^1(I)$ and $L^2(I)$ respectively.

Exercise 2.45. Verify that uniform convergence on a bounded interval I implies the other modes of convergence on I. $\Diamond$

Uniform convergence:
$|g_n(x) - g(x)| \to 0$
uniformly in x.
For each $\varepsilon > 0$ there is an N that works for all $x \in \mathbb{T}$.

$\swarrow \qquad \searrow$

Pointwise convergence: $\lvert g_n(x) - g(x)\rvert \to 0$ for all $x \in \mathbb{T}$. For each $\varepsilon > 0$ and each $x \in \mathbb{T}$ there is an $N = N(x, \varepsilon)$ that works.	L^2 *(mean-square) convergence*: $\int_{\mathbb{T}} \lvert g_n(x) - g(x)\rvert^2 \, dx \to 0$.
$\Downarrow$	$\Downarrow$
Pointwise a.e. convergence: $\lvert g_n(x) - g(x)\rvert \to 0$ for almost all $x \in \mathbb{T}$. For each $\varepsilon > 0$ and a.e. $x \in \mathbb{T}$ there is an $N = N(x, \varepsilon)$ that works.	L^1 *(mean) convergence*: $\int_{\mathbb{T}} \lvert g_n(x) - g(x)\rvert \, dx \to 0$.

Figure 2.4. Relations between five of the six modes of convergence discussed in Section 2.2, for functions g_n, $g : \mathbb{T} \to \mathbb{C}$, showing which modes imply which other modes. All potential implications between these five modes other than those shown here are false.

Exercise 2.46. Show that for $1 \leq q < p \leq \infty$ convergence in $L^p(I)$ implies convergence in $L^q(I)$. **Hint:** Use Hölder's Inequality (inequality (2.6)). $\Diamond$

The following examples show that in Figure 2.4 the implications go only in the directions indicated. In these examples we work on the unit interval $[0, 1]$.

Example 2.47. (*Pointwise but not uniform*) The sequence of functions defined by $f_n(x) = x^n$ for $x \in [0, 1]$ converges pointwise to the function f defined by $f(x) = 0$ for $0 \leq x < 1$ and $f(1) = 1$, but it does not converge uniformly to f or to any other function. See Figure 2.5. $\diamondsuit$

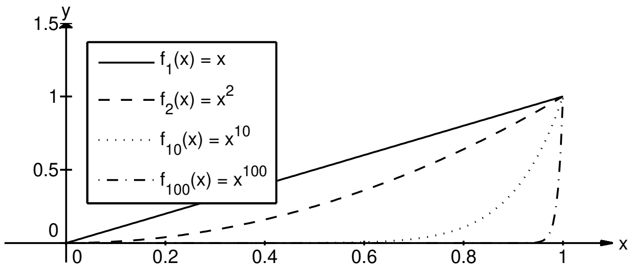

Figure 2.5. The functions $f_n(x) = x^n$ converge pointwise, but not uniformly, on $[0, 1]$. Their limit is the function $f(x)$ that takes the value 0 for $x \in [0, 1)$ and 1 at $x = 1$. Graphs of f_n are shown for $n = 1$, 2, 10, and 100.

Example 2.48. (*Pointwise but not in the mean*) The sequence of functions $g_n : [0, 1] \to \mathbb{R}$ defined by $g_n(x) = n$ if $0 < x < 1/n$ and $g_n(x) = 0$ otherwise, converges pointwise everywhere to $g(x) = 0$, but it does not converge in the mean or in L^2, nor does it converge uniformly. See Figure 2.6. $\diamondsuit$

Example 2.49. (*In the mean but not pointwise*) Given a positive integer n, let m, k be the unique non-negative integers such that $n = 2^k + m$ and $0 \leq m < 2^k$. The sequence of functions $h_n(x) : [0, 1) \to \mathbb{R}$ defined by $h_n(x) = 1$ if $m/2^k \leq x < (m + 1)/2^k$ and $h_n(x) = 0$ otherwise, converges in the mean to the zero function, but it does not converge pointwise at any point $x \in [0, 1)$. This sequence of functions is sometimes called the *walking functions*. See Figure 2.7. $\diamondsuit$

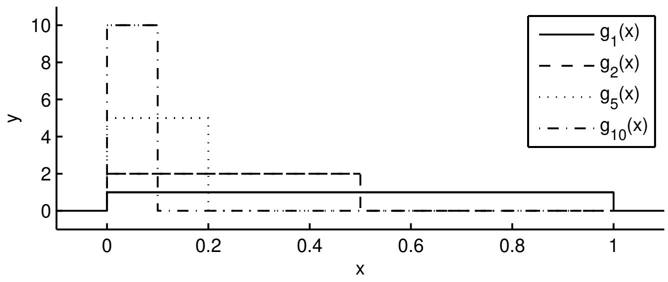

Figure 2.6. The functions g_n defined in Example 2.48 converge pointwise on $(0, 1]$, but not in the mean, in L^2, or uniformly. Their limit is the function $g(x)$ that takes the value 0 for $x \in (0, 1]$; g is undefined at $x = 0$. Graphs of g_n are shown for $n = 1$, 2, 5, and 10.

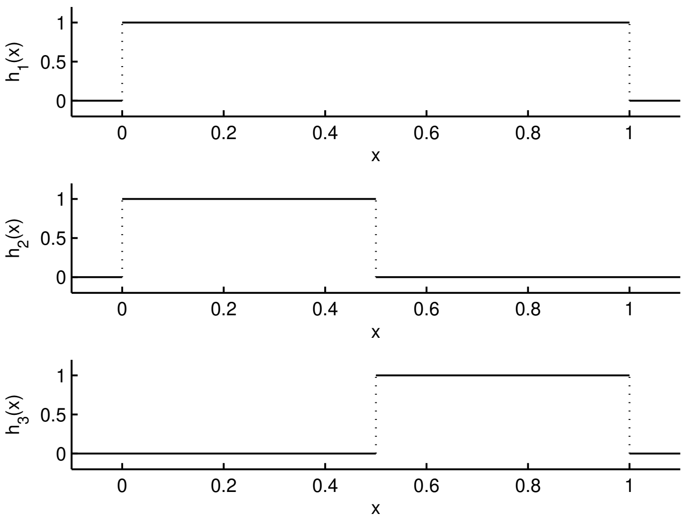

Figure 2.7. The first three walking functions of Example 2.49. The functions $h_n(x)$ converge in mean, but do not converge pointwise at any x in $[0, 1)$.

Despite these examples, there is a connection between pointwise almost everywhere convergence and convergence in L^p. For p such that $1 \leq p < \infty$, if $f_n, f \in L^p(I)$, and if $f_n \to f$ in L^p, then there is a subsequence $\{f_{n_k}\}_{k \in \mathbb{N}}$ that converges to f pointwise a.e. on I. See [**SS05**, Corollary 2.2].

Exercise 2.50. Verify that the sequences described in Examples 2.47, 2.48, and 2.49 have the convergence properties claimed there. Adapt the sequences in the examples to have the same convergence properties on the interval $[a, b)$. $\Diamond$

Exercise 2.51. Where would uniform convergence outside a set of measure zero, and L^p convergence for a given p with $1 \leq p \leq \infty$, fit in Figure 2.4? Prove the implications that involve these two modes. Devise examples to show that other potential implications involving these two modes do not hold. $\Diamond$

2.3. Interchanging limit operations

When can you interchange the order of two integral signs? Uniform convergence on closed bounded intervals is a very strong form of convergence, which allows interchanges with other limiting operations such as integration and differentiation. We list here without proof several results from advanced calculus that will be used throughout the book. Some involve uniform convergence as one of the limit operations; others involve other limit operations such as interchanging two integrals.

Theorem 2.52 (Interchange of Limits and Integrals)**.** *If $\{f_n\}_{n \in \mathbb{N}}$ is a sequence of integrable functions that converges uniformly to f in $[a, b]$ then f is integrable on $[a, b]$ and*

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b \lim_{n\to\infty} f_n = \int_a^b f.$$

That is, the limit of the integrals of a uniformly convergent sequence of functions on $[a, b]$ is the integral of the limit function on $[a, b]$.

Notice that the functions in Example 2.48 do not converge uniformly, and they provide an example where this interchange fails. In fact, $\int_0^1 g_n = 1$ for all n and $\int_0^1 g = 0$, so

$$\lim_{n\to\infty} \int_0^1 g_n = 1 \neq 0 = \int_0^1 \lim_{n\to\infty} g_n = \int_0^1 g.$$

Likewise, we can interchange infinite summation and integration provided the convergence of the partial sums is uniform.

Exercise 2.53. Given a closed bounded interval I and integrable functions $g_n : I \to \mathbb{C}$, $n \in \mathbb{N}$, suppose the partial sums $\sum_{n=0}^{N} g_n(x)$ converge uniformly on I as $N \to \infty$. Show that

$$\int_I \sum_{n=0}^{\infty} g_n(x)\,dx = \sum_{n=0}^{\infty} \int_I g_n(x)\,dx. \qquad \Diamond$$

It is therefore important to know when partial sums converge uniformly to their series. The *Weierstrass M-Test*[21] is a useful technique for establishing uniform convergence of partial sums of functions.

Theorem 2.54 (Weierstrass M-Test)**.** *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of bounded real-valued functions on a subset X of the real numbers. Assume there is a sequence of positive real numbers such that $|f_n(x)| \leq a_n$ for all $x \in X$, and $\sum_{n=1}^{\infty} a_n < \infty$. Then the series $\sum_{n=1}^{\infty} f_n$ converges uniformly to a real-valued function on X.*

Under certain conditions, involving uniform convergence of a sequence of functions and their derivatives, one can exchange differentiation and the limit.

Theorem 2.55 (Interchange of Limits and Derivatives)**.** *Let $f_n : [a,b] \to \mathbb{C}$ be a sequence of C^1 functions converging uniformly on $[a,b]$ to the function f. Assume that the derivatives $(f_n)'$ also converge uniformly on $[a,b]$ to some function g. Then f is C^1, and $f' = g$, that is*

$$\lim_{n\to\infty} (f_n)' = \Big(\lim_{n\to\infty} f_n\Big)'.$$

In other words, we can interchange limits and differentiation provided the convergence of the functions and the derivatives is uniform.

[21]Named after the German mathematician Karl Theodor Wilhelm Weierstrass (1815–1897).

Exercise 2.56. Show by exhibiting counterexamples that each hypothesis in Theorem 2.55 is necessary. ♢

Exercise 2.57. What hypotheses on a series $\sum_{n=0}^{\infty} g_n(x)$ are necessary so that it is legal to interchange sum and differentiation? In other words, when is $\big(\sum_{n=0}^{\infty} g_n(x)\big)' = \sum_{n=0}^{\infty} (g_n)'(x)$? ♢

Here is another useful result in the same vein, already mentioned in Remark 2.42.

Theorem 2.58. *The uniform limit of continuous functions on a subset X of the real numbers is continuous on X.*

Theorem 2.58 says that if $g_n \to g$ uniformly, and the functions g_n are continuous, then for all $x_0 \in X$ we can interchange the limit as x approaches x_0 with the limit as n tends to infinity:

$$\lim_{x\to x_0} \lim_{n\to\infty} g_n(x_0) = \lim_{n\to\infty} \lim_{x\to x_0} g_n(x). \tag{2.7}$$

Showing more steps will make the above statement clearer. First, we see that $\lim_{x\to x_0} \lim_{n\to\infty} g_n(x) = \lim_{x\to x_0} g(x) = g(x_0)$, because g_n converges uniformly to g, and g is continuous at x_0. Second, we see that $g(x_0) = \lim_{n\to\infty} g_n(x_0) = \lim_{n\to\infty} \lim_{x\to x_0} g_n(x)$, because g_n converges uniformly to g, and for all $n \in \mathbb{N}$, g_n is continuous at x_0.

Example 2.47 shows a sequence of continuous functions on $[0, 1]$ that converge pointwise to a function that is discontinuous (at $x = 1$). Thus the convergence cannot be uniform.

There is a partial converse, called *Dini's Theorem*[22], to Theorem 2.58. It says that the pointwise limit of a decreasing sequence of functions is also their uniform limit.

Theorem 2.59 (Dini's Theorem)**.** *Suppose that f, $f_n : I \to \mathbb{R}$ are continuous functions. Assume that (i) f_n converges pointwise to f as $n \to \infty$; and (ii) f_n is a decreasing sequence, namely $f_n(x) \geq f_{n+1}(x)$ for all $x \in I$. Then f_n converges uniformly to f.*

Exercise 2.60. Prove Dini's Theorem, and show that hypothesis (ii) is essential. A continuous version of Example 2.48, where the *steps* are replaced by *tents*, might help. ♢

[22]Named after the Italian mathematician Ulisse Dini (1845–1918).

In Weierstrass's M-Test, Theorem 2.54, if the functions $\{f_n\}_{n=1}^{\infty}$ being summed up are continuous then the series $\sum_{n=1}^{\infty} f_n(x)$ is a continuous function.

Interchanging integrals is another instance of these interchange-of-limit operations. The theorems that allow for such interchange go by the name *Fubini's Theorems*[23]. Here is one such theorem, valid for continuous functions.

Theorem 2.61 (Fubini's Theorem)**.** *Let I and J be closed intervals, $F : I \times J \to \mathbb{C}$ be a continuous function. Assume*

$$\iint_{I\times J} |F(x,y)|\, dA < \infty,$$

where dA denotes the differential of area. Then

$$\int_I \int_J F(x,y)\, dx\, dy = \int_J \int_I F(x,y)\, dy\, dx = \iint_{I\times J} F(x,y)\, dA.$$

Usually Fubini's Theorem refers to an L^1-version of Theorem 2.61 that is stated and proved in every measure theory book. You can find a two-dimensional version in [**Tao06b**, Section 19.5].

Interchanging limit operations is a delicate maneuver that cannot always be accomplished. We illustrate throughout the book settings where this interchange is allowed and settings where it is illegal, such as the ones just described involving uniform convergence.

Unlike the Riemann integral, the Lebesgue theory allows for the interchange of a pointwise limit and an integral. There are several landmark theorems that you learn in a course on measure theory. We state one such result.

Theorem 2.62 (Lebesgue Dominated Convergence Theorem)**.** *Consider a sequence of measurable functions f_n defined on the interval I, converging pointwise a.e. to a function f. Suppose there exists a dominating function $g \in L^1(I)$, meaning that $|f_n(x)| \leq g(x)$ a.e. for all $n > 0$. Then*

$$\lim_{n\to\infty} \int_I f_n(x)\, dx = \int_I f(x)\, dx.$$

[23] Named after the Italian mathematician Guido Fubini (1879–1943).

Exercise 2.63. Verify that the functions in Example 2.48 cannot be dominated by an integrable function g. $\Diamond$

Remark 2.64. The Lebesgue Dominated Convergence Theorem holds on $\mathbb{R}$ (see Appendix A). $\Diamond$

2.4. Density

Having different ways to measure convergence of functions to other functions provides ways of deciding when we can approximate functions in a given class by functions in another class. In particular we can decide when a subset is dense in a larger set of functions.

Definition 2.65. Given a normed space X of functions defined on a bounded interval I, with norm denoted by $\|\cdot\|_X$, we say that a subset $A \subset X$ is *dense in* X if given any $f \in X$ there exists a sequence of functions $f_n \in A$ such that $\lim_{n\to\infty} \|f_n - f\|_X = 0$. Equivalently, A is dense in X if given any $f \in X$ and any $\varepsilon > 0$ there exists a function $h \in A$ such that

$$\|h - f\|_X < \varepsilon. \qquad \Diamond$$

One can deduce from the definition of the Riemann-integrable functions on a bounded interval I that the step functions (see Definition 2.3) are dense in $\mathcal{R}(I)$ with respect to the L^1-norm; see Proposition 2.11. We state this result as a theorem, for future reference.

Theorem 2.66. *Given a Riemann-integrable function on a bounded interval I, $f \in \mathcal{R}(I)$, and given $\varepsilon > 0$, there exists a step function h such that*

$$\|h - f\|_{L^1(I)} < \varepsilon.$$

Remark 2.67. We can choose the step functions that approximate a given Riemann-integrable function to lie entirely below the function, or entirely above it. $\Diamond$

In Chapter 4 we will learn that trigonometric polynomials can approximate uniformly periodic continuous functions on $\mathbb{T}$. This important result is Weierstrass' Theorem (Theorem 3.4). In Chapter 9 we will learn that on closed bounded intervals, continuous functions

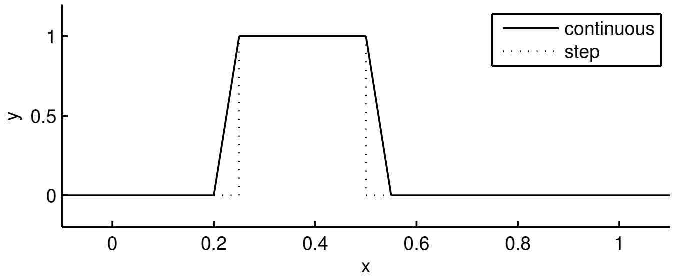

Figure 2.8. Graph of a continuous function that approximates a step function, both pointwise a.e. and in L^p norm.

can be approximated uniformly by step functions and that they can also be approximated in the L^2 norm (or L^p norm) by step functions.

It is perhaps more surprising that continuous functions can approximate step functions pointwise almost everywhere and in the L^2 norm (or L^p norm). We prove this result in Chapter 4, when we introduce the concepts of convolution and approximations of the identity. The result seems more plausible if you draw a picture of a step function and then remove the discontinuities by putting in suitable steep line segments; see Figure 2.8.

Assuming these results, proved in later chapters, we can establish the density of the continuous functions in the class of Riemann-integrable functions and as well as in the L^p spaces.

Theorem 2.68. *The continuous functions on a closed bounded interval I are dense in $\mathcal{R}(I)$ with respect to the L^p norm.*

Proof. Start with $f \in \mathcal{R}(I)$. There is a step function h that is ε-close in the L^p norm to f, that is, $\|f - h\|_{L^p(I)} < \varepsilon$. There is also a continuous function $g \in C(I)$ that is ε-close in the L^p norm to the step function h, that is, $\|h - g\|_{L^p(I)} < \varepsilon$. By the triangle inequality, the continuous function g is 2ε-close in the L^p norm to our initial function f: $\|f - g\|_{L^p(I)} \leq \|f - h\|_{L^p(I)} + \|h - g\|_{L^p(I)} \leq 2\varepsilon$. Hence continuous functions can get arbitrarily close in the L^p norm to any Riemann-integrable function. $\square$

Remark 2.69. We will use a precise version of this result for $p = 1$ in Chapter 4 (see Lemma 4.10), where we require the approximating continuous functions to be uniformly bounded by the same bound that controls the Riemann-integrable function. Here is the idea. Given real-valued $f \in \mathcal{R}(I)$ and $\varepsilon > 0$, there exists a real-valued continuous function g on I (by Theorem 2.68) such that $\|g - f\|_{L^1(I)} < \varepsilon$. Let f be bounded by $M > 0$, that is, $|f(x)| \leq M$ for all $x \in I$. If g is bounded by M there is nothing to be done. If g is not bounded by M, consider the new function defined on I by $g_0(x) = g(x)$ if $|g(x)| \leq M$, $g(x) = M$ if $g(x) > M$, and $g(x) = -M$ if $g(x) < -M$. One can show that g_0 is continuous, $g_0(x) \leq M$ for all $x \in I$, and g_0 is closer to f than g is. In particular the L^1-norm of the difference is at most ε, and possibly even smaller. $\Diamond$

Exercise 2.70. Show that the function g_0 defined in Remark 2.69 satisfies the properties claimed there. $\Diamond$

Exercise 2.71. Let X be a normed space of functions over a bounded interval I. Suppose A is a dense subset of X, and $A \subset B \subset X$. Then B is a dense subset of X (all with respect to the norm of the ambient space X). $\Diamond$

Exercise 2.72. (*Density is a Transitive Property*) Let X be a normed space of functions over a bounded interval I. Suppose A is a dense subset of B, and B is a dense subset of X. Then A is a dense subset of X (all with respect to the norm of the ambient space X). $\Diamond$

Notice that the notion of density is strongly dependent on the norm or metric of the ambient space. For example, when considering nested L^p spaces on a closed bounded interval, $C(I) \subset L^2(I) \subset L^1(I)$. Suppose we knew that the continuous functions are dense in $L^2(I)$ (with respect to the L^2 norm), and the square-integrable functions are dense in $L^1(I)$ (with respect to the L^1 norm). Even so, we could not use transitivity to conclude that the continuous functions are dense in $L^1(I)$. We need to know more about the relation between the L^1 and the L^2 norms. In fact, to complete the argument, it suffices to know that $\|f\|_{L^1(I)} \leq C\|f\|_{L^2(I)}$ (see Exercise 2.26), because that would imply that if the continuous functions are dense in $L^2(I)$ (with respect to the L^2 norm), then they are also dense with respect to the

L^1 norm, and now we can use the transitivity. Such a norm inequality is stronger than the statement that $L^2(I) \subset L^1(I)$. It says that the *identity map* or *embedding* $E : L^2(I) \to L^1(I)$, defined by $Ef = f$ for $f \in L^2(I)$, is a *continuous map* or a *continuous embedding*.

Exercise 2.73. Use Exercise 2.26 to show that the embedding $E : L^q(I) \to L^p(I)$, $Ef = f$, is a continuous embedding. That is, show that given $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|Ef - Eg\|_{L^p(I)} < \varepsilon$ if $\|f - g\|_{L^q(I)} < \delta$. ♢

We do not prove the following statements about various subsets of functions being dense in $L^p(I)$, since they rely on Lebesgue integration theory. However, we use these results in subsequent chapters, for example to show that the trigonometric system is a basis for $L^2(\mathbb{T})$.

We know that $L^p(I)$ is the completion of $\mathcal{R}(I)$ with respect to the L^p norm. In particular, the Riemann-integrable functions are dense in $L^p(I)$ with respect to the L^p norm. With this fact in mind, we see that any set of functions that is dense in $\mathcal{R}(I)$ with respect to the L^p norm is also dense in $L^p(I)$ with respect to the L^p norm, by transitivity. For future reference we state here the specific density of the continuous functions in $L^p(I)$.

Theorem 2.74. *The continuous functions on a closed bounded interval I are dense in $L^p(I)$, for all p with $1 \leq p \leq \infty$.*

Exercise 2.75. Given $f \in L^p(I)$, $1 \leq p < \infty$, $\varepsilon > 0$, and assuming all the theorems and statements in this section, show that there exists a step function h on I such that $\|h(x) - f(x)\|_{L^p(I)} < \varepsilon$. ♢

2.5. Project: Monsters, Take I

In this project we explore some unusual functions. Here we quote T. Körner [**Kör**, Chapter 12], who in turn cites Hermite and Poincaré:

> Weierstrass's nowhere differentiable continuous function was the first in a series of examples of functions exhibiting hitherto unexpected behavior. [...] Many mathematicians held up their hands in (more or less) genuine horror at *'this dreadful plague of continuous nowhere differentiable functions'* (Hermite), with which nothing mathematical could be done. [...]
>
> Poincaré wrote that in the preceding half century,

'we have seen a rabble of functions arise whose only job, it seems, is to look as little as possible like decent and useful functions. No more continuity, or perhaps continuity but no derivatives. Moreover, from the point of view of logic, it is these strange functions which are the most general; whilst those one meets unsearched for and which follow simple laws are seen just as very special cases to be given their own tiny corner. Yesterday, if a new function was invented it was to serve some practical end, today they are specially invented only to show up the arguments of our fathers, and they will never have any other use.'

(*Collected Works*, Vol. 11, p.130)

(a) Read about, and write an explanation of, the construction of a function that is continuous everywhere, but nowhere differentiable. Weierstrass showed that the function $\sum_{n=0}^{\infty} b^n \cos(a^n \pi x)$ where a is an odd integer, b lies strictly between 0 and 1, and ab is strictly larger than $1+3\pi/2$, has these properties. Bressoud works through the proof for Weierstrass's example in [**Bre**, Chapter 6]. Körner shows in [**Kör**, Chapter 11] that the same properties hold for $\sum_{n=0}^{\infty} \frac{1}{n!} \sin\big((n!)^2 x\big)$, another version of Weierstrass's example. See [**SS03**, Section 4.3].

(b) The following quote from [**JMR**, p.3] indicates that Fourier and wavelet analysis can be used to analyze such functions: *'Holschneider and Tchamitchian have shown that wavelet analysis is more sensitive and efficient than Fourier analysis for studying the differentiability of the Riemann function at a given point.'* What did they mean? What is the Riemann function? Here are some possible references to start you off: [**HT**], [**Jaf95**], [**Jaf97**], [**JMR**, Chapter 10], [**KL**, Section 4, p.382]. You will be better equipped for this search after Chapters 3–10.

(c) As a final note, Hermite and Poincaré were wrong, in the sense that there are examples of nowhere differentiable functions that turn up in real life, including the world of mathematical art: fractals, Brownian motion[24] and the k^{th} derivative of a k times differentiable Daubechies wavelet (see Chapter 10). Write an essay supporting this statement, choose one of these examples and investigate its definition, history, properties.

[24] Named after the Scottish botanist and paleobotanist Robert Brown (1773-1858).

Chapter 3

Pointwise convergence of Fourier series

In this chapter we study pointwise and uniform convergence of Fourier series for smooth functions (Section 3.1). We also discuss the connection between the decay of the Fourier coefficients and smoothness (Section 3.2). We end with a historical account of convergence theorems (Section 3.3).

We assume that the reader is familiar with the notions of pointwise and uniform convergence. In Chapter 2 we review these concepts, as well as several other modes of convergence that arise naturally in harmonic analysis.

3.1. Pointwise convergence: why do we care?

For the periodic ramp function in Example 1.12, the Fourier expansion converges *pointwise* to $f(\theta) = \theta$ everywhere except at the odd multiples of π, where it converges to zero (the halfway height in the vertical jump). For the toy model of a voice signal in Example 1.1, we were able to reconstruct the signal perfectly using only four numbers (two frequencies and their corresponding amplitudes), because the signal was very simple. In practice, complicated signals such as real voice signals transmitted through phone lines require many more amplitude coefficients and frequencies for accurate reconstruction, sometimes thousands or millions. Should we use all of these coefficients in reconstructing our function? Or perhaps we could get a reasonable approximation to our original signal using only a subset of the coefficients? If so, which ones should we keep? One simple approach is to truncate at specific frequencies $-N$ and N, and to use only the coefficients corresponding to frequencies such that $|n| \leq N$. Instead of the full Fourier series, we obtain the N^{th} *partial Fourier sum of f*,

$$S_N f(\theta) = \sum_{|n|\leq N} a_n e^{in\theta} = \sum_{|n|\leq N} \widehat{f}(n) e^{in\theta}, \tag{3.1}$$

which is a trigonometric polynomial of degree N.

Is $S_N f$ like f? In general, we need to know whether $S_N f$ looks like f as $N \to \infty$, then judge which N will be appropriate for truncation so that the reconstruction is accurate enough for the purposes of the application at hand.

For the periodic ramp function f in Example 1.12, the partial Fourier sums are

$$S_N f(x) = \sum_{|n|\leq N} \frac{(-1)^{n+1}}{in} e^{in\theta} = 2\sum_{n=1}^{N} \frac{(-1)^{n+1}}{n} \sin(n\theta).$$

In Figure 3.1 we plot f together with $S_N f$, for $N = 2$, 5, and 20.

Exercise 3.1. Create a MATLAB script to reproduce Figure 3.1. Experiment with other values of N. Describe what you see. Near the jump discontinuity you will see an overshoot that remains large as N increases. This is the *Gibbs phenomenon*[1]. Repeat the exercise for the *periodic step function*

$$g(\theta) = \begin{cases} 1, & \text{if } \theta \in [0, \pi); \\ -1, & \text{if } \theta \in [-\pi, 0). \end{cases}$$

[1]Named after the American mathematician and perhaps primarily physical chemist Josiah Willard Gibbs (1839–1903), who reported it in a letter to *Nature* in 1899. The Gibbs phenomenon has to do with how poorly a Fourier series converges near a jump discontinuity of the function f. More precisely, for each N there is a neighborhood of the discontinuity, decreasing in size as N increases, on which the partial sums $S_N f$ *always overshoot the left- and right-hand limits of f by about 9%*.

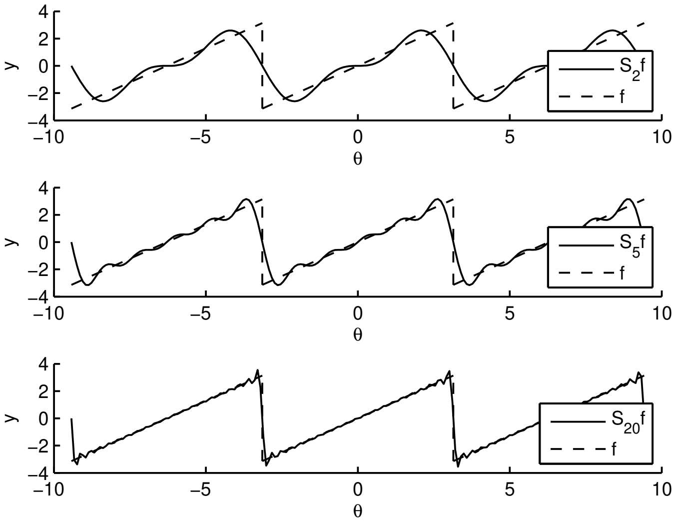

Figure 3.1. Graphs of the periodic ramp function f and its partial Fourier sums $S_N f$, for $N = 2$, 5, and 20, extended periodically from (π, π) to $\mathbb{R}$. Compare with Figure 3.2.

Project 3.4 outlines a more detailed investigation of the Gibbs phenomenon. ◊

Exercise 3.2. Compute the partial Fourier sums for the trigonometric polynomial in Exercise 1.11. ◊

Here is an encouraging sign that pointwise approximation by partial Fourier sums may work well for many functions. As the example in Exercise 3.2 suggests, if f is a 2π-periodic trigonometric polynomial of degree M, in other words a function of the form

$$f(\theta) = \sum_{|k| \leq M} a_k e^{ik\theta},$$

then for $N \geq M$, its partial Fourier sums $S_N f(x)$ coincide with $f(x)$. Therefore $S_N f$ converges pointwise to f, for all θ and for all trigonometric polynomials f. *A fortiori*, the convergence is uniform on $\mathbb{T}$, since $S_M f = S_{M+1} f = \cdots = f$.

> *The partial Fourier sums $S_N f$ converge uniformly to f for all trigonometric polynomials f.*

Exercise 3.3. Show that the Taylor polynomials $P_N(f,0)$ converge uniformly for all polynomials f. See formula (1.4). $\Diamond$

Approximating with partial Fourier sums presents some problems. For example, there are continuous functions for which one does not get pointwise convergence of the partial Fourier sums everywhere, as du Bois-Reymond[2] showed. One of the deepest theorems in twentieth-century analysis, due to Lennart Carleson[3], says that one does, however, get *almost everywhere* convergence for *square-integrable functions on* $\mathbb{T}$, and in particular for continuous functions; see Section 3.3 for more details, and Chapter 2 for definitions.

One can obtain better results in approximating a function f if one combines the Fourier coefficients in ways different from the partial sums $S_N f$. For example, averaging over the partial Fourier sums provides a smoother truncation method in terms of the *Cesáro*[4] *means*,

$$\sigma_N f(\theta) = \big(S_0 f(\theta) + S_1 f(\theta) + \cdots + S_{N-1} f(\theta)\big)/N, \tag{3.2}$$

that we will study more carefully in Section 4.4. In particular, we will prove Fejér's[5] Theorem (Theorem 4.32):

> *The Cesáro means $\sigma_N f$ converge uniformly to f for continuous functions f on $\mathbb{T}$.*

Therefore the following *uniqueness principle* holds:

> *If $f \in C(\mathbb{T})$ and $\widehat{f}(n) = 0$ for all n, then $f = 0$.*

Just observe that if $\widehat{f}(n) = 0$ for all n then $\sigma_N f = 0$ for all $N > 0$, but the Cesáro means $\sigma_N f$ converge uniformly to f. Therefore f must be identically equal to zero.

The Cesáro means are themselves trigonometric polynomials, so we can deduce from this result a theorem named after the German mathematician Karl Theodor Wilhelm Weierstrass (1815–1897).

---

[2]The German mathematician Paul David Gustav du Bois-Reymond (1831–1889).

[3]Carleson was born in Stockholm, Sweden in 1928. He was awarded the Abel Prize in 2006 "for his profound and seminal contributions to harmonic analysis and the theory of smooth dynamical systems".

[4]Named after the Italian mathematician Ernesto Cesáro (1859–1906).

[5]Named after the Hungarian mathematician Lipót Fejér (1880–1959).

Theorem 3.4 (Weierstrass's Approximation Theorem for Trigonometric Polynomials). *Every continuous function on $\mathbb{T}$ can be approximated uniformly by trigonometric polynomials. Equivalently, the trigonometric functions are dense in the continuous functions on $\mathbb{T}$ with respect to the uniform norm.*

Aside 3.5. There is a version of Weierstrass's Theorem for continuous functions on closed and bounded intervals and plain polynomials. Namely, continuous functions on $[a, b]$ can be approximated uniformly by polynomials. Equivalently, the polynomials are dense in $C([a, b])$ with respect to the uniform norm. See [**Tao06b**, Section 14.8, Theorem 14.8.3]. Both versions are special cases of a more general result, the Stone[6]–Weierstrass Theorem, which can be found in more advanced textbooks such as [**Fol**, Chapter 4, Section 7]. $\Diamond$

To sum up, continuous periodic functions on the circle can be uniformly approximated by trigonometric polynomials, but if we insist on approximating a continuous periodic function f by the particular trigonometric polynomials given by its partial Fourier sums $S_N f$, then even pointwise convergence can fail. However, it turns out that if we assume that f is smoother, we will get both pointwise and uniform convergence.

3.2. Smoothness vs convergence

In this section we present a first convergence result for functions that have at least two continuous derivatives. In the proof we obtain some decay of the Fourier coefficients which can be generalized to the case when the functions are k times differentiable. We then explore whether decay of the Fourier coefficients implies smoothness, and we study the rate of convergence of the partial Fourier sums for smooth functions.

3.2.1. A first convergence result.

Let us consider a function $f \in C^2(\mathbb{T})$, so f is 2π-periodic and twice continuously differentiable on the unit circle $\mathbb{T}$. We show that the numbers $S_N f(\theta)$ do converge, and that they converge to $f(\theta)$, for each $\theta \in \mathbb{T}$. In the following theorem

[6]Named after the American mathematician Marshall Harvey Stone (1903–1989).

we show that the partial sums converge uniformly to the Fourier series Sf; afterwards we argue that the limit must be f.

Theorem 3.6. *Let $f \in C^2(\mathbb{T})$. Then for each $\theta \in \mathbb{T}$ the limit of the partial sums*

$$\lim_{N\to\infty} S_N f(\theta) =: \sum_{n=-\infty}^{\infty} \widehat{f}(n)e^{in\theta} =: Sf(\theta)$$

exists. Moreover, the convergence to the limit is uniform on $\mathbb{T}$.

Proof. The partial Fourier sums of f are given by

$$S_N f(\theta) = \sum_{|n|\leq N} \widehat{f}(n)e^{in\theta}.$$

Fix θ, and notice that the Fourier series of f evaluated at θ is a power series in the complex variable $z = e^{i\theta}$:

$$\sum_{n=-\infty}^{\infty} \widehat{f}(n)e^{in\theta} = \sum_{n=-\infty}^{\infty} \widehat{f}(n)\left(e^{i\theta}\right)^n.$$

It suffices to show that this numerical series is absolutely convergent, or in other words that the series of its absolute values is convergent, that is

$$\sum_{n=-\infty}^{\infty} |\widehat{f}(n)e^{in\theta}| = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)| < \infty,$$

because then the Fourier series itself must converge, and the convergence is uniform, by the Weierstrass M-Test (Theorem 2.54).

We can find new expressions for the Fourier coefficients $\widehat{f}(n)$, for $n \neq 0$, by integrating by parts with $u = f(\theta)$, $v = -e^{-in\theta}/(in)$:

$$\begin{aligned}
\widehat{f}(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta \\
&= \frac{1}{2\pi}\left[-\frac{1}{in}f(\theta)e^{-in\theta}\right]_{-\pi}^{\pi} - \frac{1}{2\pi}\int_{-\pi}^{\pi} -\frac{1}{in}f'(\theta)e^{-in\theta}\,d\theta \\
&= \frac{1}{in}\frac{1}{2\pi}\int_{-\pi}^{\pi} f'(\theta)e^{-in\theta}\,d\theta \;=\; \frac{1}{in}\widehat{f'}(n)
\end{aligned}$$

Using the same argument for f' we conclude that $\widehat{f}(n) = -\widehat{f''}(n)/n^2$.

The integration by parts is permissible since f is assumed to have two continuous derivatives. The boundary terms vanish because of the continuity and 2π-periodicity of f, f', and the trigonometric functions (hence of their product). We conclude that for $n \neq 0$,

$$\widehat{f'}(n) = in\widehat{f}(n), \qquad \widehat{f''}(n) = -n^2\widehat{f}(n). \tag{3.3}$$

We can now make a direct estimate of the coefficients, for $n \neq 0$,

$$\begin{aligned} |\widehat{f}(n)| &= \frac{1}{n^2}|\widehat{f''}(n)|^2 = \frac{1}{n^2}\left|\frac{1}{2\pi}\int_{-\pi}^{\pi} f''(\theta)e^{-in\theta}\,d\theta\right| \\ &\leq \frac{1}{n^2}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f''(\theta)|\,d\theta\right), \end{aligned}$$

since $|e^{-in\theta}| = 1$. But f'' is continuous on the closed interval $[-\pi, \pi]$. So it is bounded by a finite constant $C > 0$, and therefore its average over any bounded interval is also bounded by the same constant: $\frac{1}{2\pi}\int_{-\pi}^{\pi} |f''(\theta)|\,d\theta \leq C$. We conclude that there is $C > 0$ such that

$$|\widehat{f}(n)| \leq C/n^2 \qquad \text{for all } n \neq 0.$$

The comparison test then implies that $\sum_{n\in\mathbb{Z}} |\widehat{f}(n)|$ converges, and the *Weierstrass M-Test* (Theorem 2.54) implies that the Fourier series $\sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{inx}$ converges uniformly and hence pointwise if f is twice continuously differentiable. □

In fact, we have shown more.

Theorem 3.7. *Suppose f is an integrable function on $\mathbb{T}$ such that $\sum_{n=-\infty}^{\infty} |\widehat{f}(n)| < \infty$. Then the partial Fourier sums $S_N f$ converge uniformly to the Fourier series Sf. Moreover, Sf will be a periodic and continuous function on $\mathbb{T}$.*

Aside 3.8. The equations (3.3) are instances of a *time–frequency dictionary* that we summarize in Table 4.1. Specifically, *derivatives of f are transformed into multiplication of $\widehat{f}$ by polynomials.* As a consequence, linear differential equations are transformed into algebraic equations. ◊

We can restate concisely Theorem 3.6 as follows.

If $f \in C^2(\mathbb{T})$, then its Fourier series converges uniformly on $\mathbb{T}$.

Remark 3.9. We have not yet established whether the Fourier series evaluated at θ converges to $f(\theta)$. ◊

If $f \in C^2(\mathbb{T})$, it is true that f is the pointwise limit of its Fourier series, though it takes more work to prove this result directly. It is a consequence of the following elementary lemma about complex numbers.

Lemma 3.10. *Suppose that the sequence of complex numbers $\{a_n\}_{n\geq 1}$ converges. Then the sequence of its averages also converges, and their limits coincide; that is*

$$\lim_{n\to\infty} \big(a_1 + a_2 + \cdots + a_n\big)/n = \lim_{n\to\infty} a_n.$$

Exercise 3.11. Prove Lemma 3.10. Give an example to show that if the sequence of averages converges, that does not imply that the original sequence converges. $\Diamond$

In Section 4.4 we show that for continuous functions f on $\mathbb{T}$, the Cesáro means $\sigma_N f(\theta)$ (see (3.2)), which are averages of the partial Fourier sums, do converge uniformly to $f(\theta)$ as $N \to \infty$, this is Fejer's Theorem (Theorem 4.32). Lemma 3.10 shows that since for $f \in C^2(\mathbb{T})$ the partial Fourier sums converge (Theorem 3.6), then the Cesáro sums also converge, and to the same limit. It follows that if $f \in C^2(\mathbb{T})$, then the partial sums $S_N f(\theta)$ of the Fourier series for f do converge pointwise to $f(\theta)$, for each $\theta \in \mathbb{T}$. Summing up, we have established the following theorem (modulo Fejér's Theorem).

Theorem 3.12. *If $f \in C^2(\mathbb{T})$, then the sequence of partial Fourier sums, $S_N f$, converges uniformly to f on $\mathbb{T}$.*

On the other hand, while the partial Fourier sums converge uniformly for C^1 functions, the argument we just gave does not suffice to guarantee pointwise, let alone uniform, convergence, since we can no longer argue by comparison, as we just did, to obtain absolute convergence. In this case, the estimate via integration by parts on the absolute value of the coefficients is insufficient; all we get is $|a_n| \leq C/n$, and the corresponding series does not converge. A more delicate argument is required just to check convergence of the sequence $\{S_N f(\theta)\}$ for $f \in C^1(\mathbb{T})$. But once the convergence of the series is guaranteed, then the same result about convergence of Cesáro sums for continuous functions applies, and the limit has no other choice than to coincide with $f(\theta)$. The following result is proved in [**Stri00a**, Chapter 12, Theorem 12.2.2].

Theorem 3.13. *If $f \in C^1(\mathbb{T})$, then its Fourier series converges uniformly on $\mathbb{T}$ to f.*

3.2.2. Smoothness vs rate of decay of Fourier coefficients.

In Section 3.2.1, we integrated twice by parts and obtained formulae (3.3) for the Fourier coefficients of the first and second derivatives of a C^2-function. If the function is C^k, then we can iterate the procedure k times, obtaining a formula for the Fourier coefficients of the k^{th} derivative of f in terms of the Fourier coefficients of f. Namely, for $n \neq 0$,

$$\widehat{f^{(k)}}(n) = (in)^k \widehat{f}(n). \tag{3.4}$$

Furthermore, we obtain an estimate about the rate of decay of the Fourier coefficients of f if the function is k times continuously differentiable on $\mathbb{T}$, for $n \neq 0$,

$$|\widehat{f}(n)| \leq C/n^k.$$

Here we have used once more the fact that $f^{(k)}$ is bounded on the closed interval $[-\pi, \pi]$, hence its Fourier coefficients are bounded: $|\widehat{f^{(k)}}(n)| \leq C$, where $C = C(k)$ is a constant independent of n but dependent on k. What we just have shown is the following.

Theorem 3.14. *If $f \in C^k(\mathbb{T})$, then its Fourier coefficients $\widehat{f}(n)$ decay at least like n^{-k}.*

If the Fourier coefficients of an integrable function decay like n^{-k} is it true that f is C^k? No. It is not even true that f is C^{k-1}. Consider the following example when $k = 1$.

Example 3.15. Recall that the *characteristic function* of the interval $[0, 1]$ on $\mathbb{T}$ is defined to be 1 for all $\theta \in [0, 1]$ and zero on $\mathbb{T} \setminus [0, 1]$, and it is denoted $\chi_{[0,1]}$. Its Fourier coefficients are given by

$$\widehat{\chi_{[0,1]}}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \chi_{[0,1]}(\theta) e^{-in\theta} \, d\theta = \frac{1}{2\pi} \int_0^1 e^{-in\theta} \, d\theta = \frac{1 - e^{-in}}{2\pi in}.$$

Clearly $|\widehat{\chi_{[0,1]}}(n)| \leq (\pi n)^{-1}$, so the Fourier coefficients decay like n^{-1}. However, the characteristic function is not continuous. ♢

One can construct similar examples where the Fourier coefficients decay like n^{-k}, $k \geq 2$, and the function is not in $C^{k-1}(\mathbb{T})$. However, it follows from Corollary 3.17 that such function must be in $C^{k-2}(\mathbb{T})$ for $k \geq 2$.

Theorem 3.16. *Let $f : \mathbb{T} \to \mathbb{C}$ be 2π-periodic and integrable, $\ell \geq 0$. If $\sum_{n\in\mathbb{Z}} |\widehat{f}(n)|\,|n|^{\ell} < \infty$, then f is C^{ℓ}.*

Furthermore the partial Fourier sums converge uniformly to f, and the derivatives up to order ℓ of the partial Fourier sums converge uniformly to the corresponding derivatives of f.

Corollary 3.17. *If $|\widehat{f}(n)| \leq C|n|^{-k}$ for $k \geq 2$, and $n \neq 0$, then f is C^{ℓ}, where $\ell = k-2$ if k is an integer, and $\ell = [k]-1$ otherwise. Here $[k]$ denotes the* integer part *of k, in other words the largest integer less than or equal to k.*

Exercise 3.18. Deduce Corollary 3.17 from Theorem 3.16. $\diamondsuit$

We sketch the proof of Theorem 3.16.

Proof of Theorem 3.16. For $\ell = 0$, this is Theorem 3.7 since the hypothesis simplifies to $\sum_{n=-\infty}^{\infty} |\widehat{f}(n)| < \infty$, which implies the uniform convergence of the partial Fourier sums $S_N f$ to some periodic and continuous function Sf. Féjer's Theorem (Theorem 4.32) shows that the averages of the partial Fourier sums, namely the Cesàro means $\sigma_N f$, see (3.2), converge uniformly to $f \in C(\mathbb{T})$. Finally Exercise 3.11 implies that $Sf = f$, and hence the partial Fourier sums converge uniformly to f. For the case $\ell = 1$, see Exercise 3.19. The general case follows by induction on ℓ. $\square$

Exercise 3.19. Suppose that $f : \mathbb{T} \to \mathbb{C}$ is a continuous function such that $\sum_{n\in\mathbb{Z}} |\widehat{f}(n)|\,|n|$ converges. Show that $S_N f \to f$ uniformly, and that $(S_N f)' \to h$ uniformly for some function h. Show that f is C^1 and that $h = f'$. **Hint:** Theorem 2.55 may be useful.

Now prove Theorem 3.16 by induction. $\diamondsuit$

In this section we learned that the Fourier coefficients of smooth functions go to zero as $|n| \to \infty$. We also found the following.

> *The smoother the function, the faster the rate of decay of its Fourier coefficients.*

What if f is merely continuous? Is it still true that the Fourier coefficients decay to zero? The answer is yes.

Lemma 3.20 (Riemann–Lebesgue Lemma for continuous functions). *Let* $f \in C(\mathbb{T})$. *Then* $\lim_{|n|\to\infty} \widehat{f}(n) = 0$.

Proof. Observe that because $e^{\pi i} = -1$, we can write the n^{th} Fourier coefficient of $f \in C(\mathbb{T})$ in a slightly different form, namely

$$\begin{aligned}\widehat{f}(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta \;=\; -\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}e^{\pi i}\,d\theta \\ &= -\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in(\theta-\pi/n)}\,d\theta \\ &= -\frac{1}{2\pi}\int_{-\pi-\frac{\pi}{n}}^{\pi-\frac{\pi}{n}} f\left(\alpha+\pi/n\right)e^{-in\alpha}\,d\alpha,\end{aligned}$$

where we have made the change of variable $\alpha = \theta - \pi/n$. But since the integrand is 2π-periodic, and we are integrating over an interval of length 2π, we can shift the integral back to $\mathbb{T}$ without altering its value, and conclude that

$$\widehat{f}(n) = -\frac{1}{2\pi}\int_{-\pi}^{\pi} f\left(\theta+\pi/n\right)e^{-in\theta}\,d\theta.$$

Averaging the original integral representation of $\widehat{f}(n)$ with this new representation gives yet another integral representation, which involves a difference of values of f, and for which we can use the hypothesis that f is continuous. More precisely,

$$\widehat{f}(n) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[f(\theta) - f\left(\theta+\pi/n\right)\right]e^{-in\theta}\,d\theta.$$

We get an estimate for $|\widehat{f}(n)|$ by the triangle inequality for integrals,

$$|\widehat{f}(n)| \le \frac{1}{4\pi}\int_{-\pi}^{\pi}\left|f(\theta) - f\left(\theta+\pi/n\right)\right|\,d\theta.$$

We can now take the limit as $|n| \to \infty$, and since f is uniformly continuous on $[-\pi,\pi]$, the sequence $g_n(\theta) := |f(\theta) - f\left(\theta+\pi/n\right)|$ converges uniformly to zero. Therefore we can interchange the limit and the integral (see Theorem 2.52), obtaining

$$0 \le \lim_{|n|\to\infty}|\widehat{f}(n)| \le \frac{1}{4\pi}\int_{-\pi}^{\pi}\lim_{|n|\to\infty}\left|f(\theta) - f\left(\theta+\pi/n\right)\right|\,d\theta = 0,$$

as required. □

Exercise 3.21. Show that if f is continuous and 2π-periodic, then it is uniformly continuous on $[-\pi, \pi]$. Moreover, the functions $g_n(\theta) := |f(\theta) - f(\theta + \pi/n)|$ converge uniformly to zero. ♢

We will see other versions of the Riemann–Lebesgue Lemma in subsequent chapters.

3.2.3. Differentiability vs rate of convergence. Now that we know something about the rate of decay of the Fourier coefficients of smooth functions, let us say something about the rate of convergence of their Fourier sums.

> *The smoother f is, the faster is the rate of convergence of the partial Fourier sums to f.*

Theorem 3.22. *If $f \in C^k(\mathbb{T})$ for $k \geq 2$, then there exists a constant $C > 0$ independent of θ such that*

$$|S_N f(\theta) - f(\theta)| \leq C/N^{k-1}. \tag{3.5}$$

If $f \in C^1(\mathbb{T})$, then there exists a constant $C > 0$ independent of θ such that

$$|S_N f(\theta) - f(\theta)| \leq C/\sqrt{N}. \tag{3.6}$$

These results automatically provide uniform convergence of the partial Fourier sums for C^k-functions, $k \geq 1$. See [**Stri00a**, Thm 12.2.2] for the proofs of this theorem.

If a function satisfies inequality (3.5) for $k \geq 2$, it does not necessarily mean that f is C^{k-1}. Therefore f is not necessarily C^k, as is shown in Example 3.23 for $k = 2$. It is not difficult to adapt that example to the case $k > 2$.

Example 3.23. (*Plucked String*) The graph of the function $f : \mathbb{T} \to \mathbb{R}$ given by $f(\theta) = \pi/2 - |\theta|$, is shown in Figure 3.2. This function is continuous but not differentiable at $\theta \in \{n\pi : n \in \mathbb{Z}\}$, and therefore it is not C^2 (or even C^1). Its Fourier coefficients are given by

$$\widehat{f}(n) = \begin{cases} 0, & \text{if } n \text{ is even;} \\ 2/(\pi n^2), & \text{if } n \text{ is odd.} \end{cases}$$

In particular $\sum_{n\in\mathbb{Z}} |\widehat{f}(n)| < \infty$, and so by Theorem 3.16, $S_N f \to f$ uniformly on $\mathbb{T}$. Furthermore, the following estimates hold:

$$(3.7) \qquad |S_N f(\theta) - f(\theta)| \leq \frac{2\pi^{-1}}{N-1}, \qquad |S_N f(0) - f(0)| \geq \frac{2\pi^{-1}}{N+2}.$$

The second estimate implies that the rate of convergence cannot be improved uniformly in θ. ◊

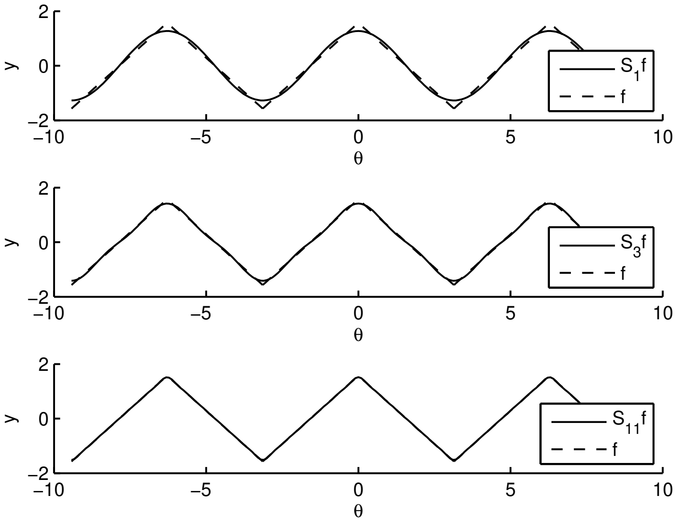

Figure 3.2. Graphs of the plucked string function $f(\theta) = \pi/2 - |\theta|$ from Example 3.23 and its partial Fourier sums $S_N f$ for $N = 1$, 3, and 9, extended periodically from $(-\pi, \pi)$ to $\mathbb{R}$. Compare the convergence of $S_N f$ with that shown for the ramp function in Figure 3.1.

Exercise 3.24. Compute the Fourier coefficients of the plucked string function in Example 3.23. Rewrite the Fourier series as a purely cosine series expansion. Also show that the inequalities (3.7) hold. **Hint:** $1/n - 1/(n+1) = 1/\big(n(n+1)\big) \leq 1/n^2 \leq 1/\big(n(n-1)\big) = 1/(n-1) - 1/n$. The left- and right-hand sides in the string of inequalities, when added up, form telescoping series. This observation allows you to obtain the following bounds: $1/N \leq \sum_{n\geq N} 1/n^2 \leq 1/(N-1)$. If this hint is not enough, see [**Kör**, Example 10.1, p.35]. $\Diamond$

Exercise 3.25. Show that if f is an integrable even function, then $\widehat{f}(n) = \widehat{f}(-n)$ for all $n \in \mathbb{Z}$. This means that the corresponding Fourier series is a purely cosine series expansion.

3.3. A suite of convergence theorems

Here we state, without proofs and in chronological order, the main results concerning pointwise convergence of the Fourier series for continuous functions and for Lebesgue-integrable functions. The Fourier series converges to $f(\theta)$ for functions f that are slightly more than just continuous at θ, but not necessarily differentiable there. In this suite of theorems we illustrate how delicate the process of mathematical thinking can be, with a number of intermediate results and puzzling examples, which slowly help bring to light the right concepts and the most complete and satisfactory results.

This is a good time to revisit Chapter 2. In this section we use ideas introduced there, such as sets of measure zero, almost everywhere convergence, Lebesgue-integrable functions and convergence in L^p.

Theorem 3.26 (Dirichlet, 1829). *Suppose that $f : \mathbb{T} \to \mathbb{C}$ is continuous and its derivative f' is continuous and bounded, except possibly at finitely many points. Then for all $\theta \in \mathbb{T}$,*

$$\lim_{N\to\infty} S_N f(\theta) = \big[f(\theta^+) + f(\theta^-)\big]/2, \tag{3.8}$$

where $f(\theta^+)$ is the limit when α approaches θ from the right, and $f(\theta^-)$ is the limit when α approaches θ from the left. In particular, $S_N f(\theta) \to f(\theta)$ as $N \to \infty$ at all points $\theta \in \mathbb{T}$ where f is continuous.

Thinking wishfully, we might dream that the partial Fourier sums for a continuous function would converge everywhere. That dream was shattered by the following result.

Theorem 3.27 (Du Bois-Reymond, 1873). *There is a continuous function $f : \mathbb{T} \to \mathbb{C}$ such that $\limsup_{N\to\infty} |S_N f(0)| = \infty$.*

Here the partial sums of the Fourier series of a continuous function fail to converge at the point $x = 0$. The modern proof of this result uses the Uniform Boundedness Principle, which we will encounter in Section 9.4.5. For a constructive proof see Project 3.5.

A reproduction of Dirichlet's paper can be found in [**KL**, Chapter 4], as well as a comparison with the following result due to Jordan[7].

Theorem 3.28 (Jordan, 1881)**.** *If f is a function of* bounded variation[8] *on $\mathbb{T}$, then the Fourier series of f converges to $[f(\theta^+)+f(\theta^-)]/2$ at every point $\theta \in \mathbb{T}$, and in particular to $f(\theta)$ at every point of continuity. The convergence is uniform on closed intervals of continuity.*

The class of functions that have bounded variation in I is denoted by $BV(I)$. Bounded variation implies Riemann integrability, that is $BV(I) \subset \mathcal{R}(I)$. The conditions in Theorem 3.26 imply bounded variation, and so Dirichlet's Theorem follows from Jordan's Theorem. A proof of Dirichlet's Theorem can be found in [**Kör**, Chapter 16, p.59]. A similar proof works for Jordan's Theorem. The following result, Dini's Criterion, can be used to deduce Dirichlet's Theorem, at least at points of continuity.

Theorem 3.29 (Dini, 1878)**.** *Let f be a 2π-periodic function. Suppose there is $\delta > 0$ such that $\int_{|t|<\delta} |f(\theta+t) - f(\theta)|/|t|\,dt < \infty$ for some θ. Then $\lim_{N\to\infty} S_N f(\theta) = f(\theta)$. (In particular f must be continuous at θ.)*

A 2π-periodic function f satisfies a *uniform Lipschitz or Hölder condition of order* α, for $0 < \alpha \leq 1$, if there exists $C > 0$ such that $|f(\theta+t) - f(\theta)| \leq C|t|^\alpha$ for all t and θ. The class of 2π-periodic functions that satisfy a uniform Lipschitz condition of order α is denoted by $C^\alpha(\mathbb{T})$. For all $0 < \alpha \leq \beta \leq 1$ we have $C(\mathbb{T}) \subset C^\alpha(\mathbb{T}) \subset C^\beta(\mathbb{T})$. If $f \in C^\alpha(\mathbb{T})$ for $0 < \alpha \leq 1$, then it satisfies Dini's condition uniformly at all $\theta \in \mathbb{T}$, and hence the partial Fourier sums converge pointwise

[7] Camille Jordan, French mathematician (1838–1922).

[8] That is, the *total variation* on $\mathbb{T}$ of f is bounded. The total variation of f is defined to be $V(f) = \sup_P \sum_{n=1}^N |f(\theta_n) - f(\theta_{n-1})|$, where the supremum is taken over all partitions $P : \theta_0 = -\pi < \theta_1 < \cdots < \theta_{N-1} < \theta_N = \pi$ of $\mathbb{T}$. Step functions are functions of bounded variation, as are monotone functions, and you can compute their variation exactly.

to f. If $\alpha > 1/2$ and $f \in C^{\alpha}(\mathbb{T})$ the convergence of the partial Fourier sums is absolute [**SS03**, Chapter 3, Exercises 15–16]. A proof of Dini's result can be found in [**Graf08**, Theorem 3.3.6, p.189].

These criteria depend only on the values of f in an arbitrarily small neighborhood of θ. This is a general and surprising fact known as the *Riemann Localization Principle*, which was part of a Memoir presented by Riemann in 1854, but only published after his death in 1867. See [**KL**, Chapter 5, Sections 1 and 4].

Theorem 3.30 (Riemann Localization Principle). *Let f and g be integrable functions whose Fourier series converge. If f and g agree on an open interval, then the Fourier series of their difference $f - g$ converges uniformly to zero on compact subsets of the interval. In other words, for an integrable function f, the convergence of the Fourier series to $f(\theta)$ only depends on the values of f in a neighborhood of θ.*

At the beginning of the twentieth century, pointwise convergence everywhere of the Fourier series was ruled out for periodic and continuous functions. However, positive results could be obtained by slightly improving the continuity. In this chapter we proved such a result for smooth (C^2) functions, and we have just stated a much refined result where the improvement over continuity is encoded in Dini's condition. Can one have a theorem guaranteeing pointwise convergence everywhere for a larger class than the class of Dini-continuous functions? Where can we draw the line?

In 1913, the Russian mathematician Nikolai Nikolaevich Luzin (1883–1950) conjectured that: *Every square-integrable function, and thus in particular every continuous function, equals the sum of its Fourier series almost everywhere.*

Ten years later Andrey Nikolaevich Kolmogorov (1903–1987), another famous Russian mathematician, found a Lebesgue-integrable function whose partial Fourier sums *diverge almost everywhere*. Three years later he constructed an even more startling example [**Kol**].

Theorem 3.31 (Kolmogorov, 1926). *There is a Lebesgue-integrable function $f : \mathbb{T} \to \mathbb{C}$ such that $\limsup_{N\to\infty} |S_N f(\theta)| = \infty$ for all $\theta \in \mathbb{T}$.*

Recall that the minimal requirement on f so that we can compute its Fourier coefficients is that f is Lebesgue-integrable $\big(f \in L^1(\mathbb{T})\big)$. Kolmogorov's result tells us that we can cook up a Lebesgue-integrable function whose partial Fourier sums $S_N f$ diverge at *every* point θ! His function is Lebesgue integrable, but it is not continuous; in fact it is essentially unbounded on every interval.

After Kolmogorov's example, experts believed that it was only a matter of time before the same sort of example could be constructed for a continuous function. It came as a big surprise when, half a century later, Carleson proved Luzin's conjecture, which implies that the partial Fourier sums of every continuous function converge at almost every point θ.

Theorem 3.32 (Carleson, 1966)**.** *Suppose $f : \mathbb{T} \to \mathbb{C}$ is square-integrable. Then $S_N f(\theta) \to f(\theta)$ as $N \to \infty$, except possibly on a set of measure zero.*

In particular, Carleson's conclusion holds if f is continuous on $\mathbb{T}$, or more generally if f is Riemann-integrable on $\mathbb{T}$. Carleson's Theorem implies that Kolmogorov's example cannot be a square-integrable function. In particular it cannot be Riemann integrable. Du Bois-Reymond's example does not contradict Carleson's Theorem. Although the function is square integrable (since it is continuous), the partial Fourier series are allowed to diverge on a set of measure zero (which can be large, as we mentioned in Section 2.1.4).

The next question is whether, given a set of measure zero in $\mathbb{T}$, one can construct a continuous function whose Fourier series diverges exactly on the given set. The next theorem answers this question.

Theorem 3.33 (Kahane[9], Katznelson[10], 1966)**.** *If $E \subset \mathbb{T}$ is a set of measure zero, then there is a continuous function $f : \mathbb{T} \to \mathbb{C}$ such that* $\limsup_{N\to\infty} |S_N f(\theta)| = \infty$ *for all $\theta \in E$.*

The proof of Theorem 3.33 can be found in Katznelson's book *An introduction to harmonic analysis* [**Kat**, Chapter II, Sec 3]. The same chapter includes a construction of a Kolmogorov-type example.

[9]Jean-Pierre Kahane, French mathematician (born 1926).
[10]Yitzhak Katznelson, Israeli mathematician (born 1934)

A year later the following non-trivial extension of Carleson's Theorem was proved by the American mathematician Richard Hunt.

Theorem 3.34 (Hunt[11]). *Carleson's Theorem holds for all $f \in L^p(\mathbb{T})$ and for all $p > 1$.*

It took almost 40 years for the Carleson–Hunt Theorem to make its way into graduate textbooks in full detail. There are now several more accessible accounts of this famous theorem, for instance in [**Graf08**] and in [**Ari**].

Now we see where to draw the line between those functions for which the partial Fourier series converge pointwise to $f(x)$ for all x, and the ones for which they do not. For example, $f \in C^k(\mathbb{T})$ is enough, but f being continuous is not. However, for continuous functions we have convergence almost everywhere, and for each set of measure zero there is a continuous function whose partial Fourier series diverges on that set. As Körner writes [**Kör**, p.75],

> *'The problem of pointwise convergence is thus settled. There are few questions which have managed to occupy even a small part of humanity for 150 years. And of those questions, very few indeed have been answered with as complete and satisfactory an answer as Carleson has given to this one.'*

3.4. Project: The Gibbs phenomenon

We have seen that if a function f is just a little better than continuous, for instance (Theorem 3.13) if f is in the space $C^1(\mathbb{T})$, then the Fourier series of f converges uniformly to f on $\mathbb{T}$. What if f is not continuous but instead has a jump discontinuity at some point x_0? We saw in Exercise 3.1 for the *sawtooth* (or *periodic ramp*) function, and for the step function, that a blip, or overshoot, is visible in the graph of the partial sum $S_N(f)(\theta)$ on either side of the discontinuity, and that while the width of these blips decreases as $n \to \infty$, the height does not. The existence of these blips is known as the *Gibbs phenomenon*. In this project we explore the Gibbs phenomenon for Fourier series.

[11] Richard Allen Hunt, American mathematician (1937–2009).

(a) Use MATLAB to calculate the values of, and to plot, several partial Fourier sums $S_N f$ for a function f with a jump discontinuity. For instance, try the *square wave* function

$$g(x) := \begin{cases} 1, & \text{if } 0 \leq x < \pi; \\ -1, & \text{if } -\pi \leq x < 0. \end{cases}$$

(Extend this function periodically and the name will explain itself.) Concentrate on the blips near zero (you should see other blips near $\pm\pi$ as well). Animate your plots to create a movie showing what happens as N increases. Estimate the height of the blip. Is this height related to the size or location of the jump? If so, how? What happens to the x-location and the width of the blips as N increases?

(b) Prove analytically that the height of the blip for the square wave function in part (a) does not tend to zero as $N \to \infty$. One approach is as follows. Calculate the Fourier coefficients $\widehat{g}\,(n)$ and the partial Fourier sums of g. Get a closed formula (no sums involved) for the derivative of the partial Fourier sums derivative $[S_{2N-1}(g)]'$. **Hint:** the following trigonometric identity may help: $\sum_{n=1}^{N} \cos\big((2n-1)x\big) = \sin(2Nx)(2\sin x)$. Use calculus to find the critical points of $S_{2N-1}(g)$ (points where the derivative is equal to zero), and check that the point $x_N = \pi/(2N)$ is actually a local maximum. Verify that

$$S_{2N-1}(g)(x_N) \sim (2/\pi) \int_0^{\pi} \big(\sin u/u\big)\, du =: (2/\pi)\,\mathrm{Si}(\pi).$$

Look up, use Taylor series, or use MATLAB to calculate, the numerical value of the integral $\mathrm{Si}(\pi)$. Interpret your results to explain the blips.

(c) Extend the results of part (b) to the case of a more general function with a jump discontinuity.

(d) Explore the literature related to the Gibbs phenomenon, and write a report, with references, summarizing the results of your literature search. Here are some references to get you started: Körner [**Kör**, Chapters 8 and 17], and Prestini [**Pre**] give overviews of the history, while more detail is in Gibbs' two letters to *Nature*, [**Gib1898**], [**Gib1899**], and Michelson's letter to the same journal [**Mic**], and in the papers by Wilbraham [**Wil**], Bôcher [**Bôc**], and Michelson and Stratton [**MS**].

(e) Can the graphs of the partial Fourier sums of g converge in any reasonable sense to the graph of g? Can we say something about the arclengths of the graphs Γ_N of $S_N g$? Here the graph of $S_n f$ is given by $\Gamma_N = \{(x, S_N f(x)) : x \in [-\pi, \pi)\}$. See the short article [**Stri00b**].

(f) Do wavelets show Gibb's phenomenon? This question will make more sense once you get acquainted with Chapters 9-11.

3.5. Project: Monsters, Take II

This project is a follow up of Project 2.5.

(a) Read an account or two of how to construct a function that is continuous everywhere but whose Fourier series diverges at some point. (See Theorem 3.27.) Write an explanation of the construction. Your intended audience is another student at the same stage of their studies as you are now. The original construction by du Bois-Reymond is explained in [**Kör**, Chapter 18, Theorem 18.1]; see also [**SS03**, Chapter 3, Section 2.2]. You can also search for the original paper (in German).

(b) As in part (a) but for Kolmogorov's example of an integrable function whose Fourier series diverges almost everywhere. You can go further, with Kolmogorov's example of an integrable function whose Fourier series diverges everywhere. Start your search with the original paper [**Kol**], or with [**Zyg59**, Chapter VIII].

Appendix A

Useful tools

In this appendix we record the definitions and the main properties of vector spaces, normed spaces, and inner-product vector spaces (Section A.1). We also record definitions, examples and main properties of Banach spaces and Hilbert spaces (Section A.2). We specialize to the Lebesgue spaces on the line $L^p(\mathbb{R})$. We collect basic definitions, density results, and interchange of limit and integral results valid in $\mathbb{R}$ that we use in the book (Section A.3).

A.1. Vector spaces, norms, inner products

We recall here definitions and properties of vector spaces, normed and inner product spaces. We list some common examples that we use in the book. We pay particular attention to the geometry of inner-product vector spaces, and state the basic theorems related to orthogonality: the Cauchy–Schwarz inequality, the Pythagorean theorem and the triangle inequality (Section A.1.1). We discuss the notions of a (Schauder) basis and unconditional bases in Banach spaces (Section A.2.2). We also discuss orthonormal bases (Section A.1.2) and orthogonal projections onto closed subspaces (Section A.2.4).

Definition A.1. A *vector space over* $\mathbb{C}$ is a set V, together with operations of addition and scalar multiplication, such that if x, y, $z \in V$ and λ, λ_1, $\lambda_2 \in \mathbb{C}$ then

(i) *V is closed under addition and scalar multiplication*: $x+y \in V$, and $\lambda x \in V$.

(ii) *Addition is commutative and associative*: $x+y=y+x$, and $x+(y+z)=(x+y)+z$.

(iii) There exists a *zero vector* $0 \in V$ such that $x+0=0+x=x$ for all $x \in V$.

(iv) Each $x \in V$ has an *additive inverse* $-x \in V$ such that $x+(-x)=(-x)+x=0$.

(v) *Addition is distributive w.r.t. scalar multiplication*: $\lambda(x+y)=\lambda x+\lambda y$.

(vi) *Scalar multiplication is associative*: $\lambda_1(\lambda_2 x)=(\lambda_1\lambda_2)x$. $\diamondsuit$

We have already encountered some vector spaces over $\mathbb{C}$.

Example A.2. The set $\mathbb{C}^n = \{(z_1, z_2, \dots, z_n) : z_j \in \mathbb{C}\}$ of n-tuples of complex numbers, with the usual component-by-component addition and scalar multiplication. $\diamondsuit$

Example A.3. The collections of functions on $\mathbb{T}$ we have been working with (Section 2.1) with the usual pointwise addition and scalar multiplication: $C(\mathbb{T})$, $C^k(\mathbb{T})$, $C^\infty(\mathbb{T})$; the collection $\mathcal{R}(\mathbb{T})$ of Riemann-integrable functions on $\mathbb{T}$; the Lebesgue spaces $L^p(\mathbb{T})$, in particular $L^1(\mathbb{T})$, $L^2(\mathbb{T})$, and $L^\infty(\mathbb{T})$; and the collection $\mathcal{P}_N(\mathbb{T})$ of 2π-periodic trigonometric polynomials of degree less than or equal to N. See the ladder of functional spaces in $\mathbb{T}$ in Figure 2.3. $\diamondsuit$

Example A.4. The corresponding collections of functions on $\mathbb{R}$ discussed in the book: $C(\mathbb{R})$, $C^k(\mathbb{R})$, $C^\infty(\mathbb{R})$; the collection $C_c(\mathbb{R})$ of compactly supported continuous functions; similarly $C_c^k(\mathbb{R})$ and $C_c^\infty(\mathbb{R})$; the collection $\mathcal{S}(\mathbb{R})$ of Schwartz functions; infinitely often differentiable and rapidly decaying functions all of whose derivatives are also rapidly decaying (see Section 7.2); the Lebesgue spaces $L^p(\mathbb{R})$ (see Appendix A). $\diamondsuit$

Example A.5. The space $\ell^2(\mathbb{Z})$ of sequences defined by equation (5.1) in Section 5.2, with componentwise addition and scalar multiplication. See also Example A.13. $\diamondsuit$

It is useful to be able to measure closeness between vectors in a vector space. One way to do so is via a norm.

Definition A.6. A *norm* on a vector space V over $\mathbb{C}$ is a function $\|\cdot\| : V \to \mathbb{R}$ such that for all x, $y \in V$, $\alpha \in \mathbb{C}$:

(i) $\|x\| \geq 0$ (*positive*),

(ii) $\|x\| = 0$ if and only if $x = 0$ (*positive definite*),

(iii) $\|\alpha x\| = |\alpha|\,\|x\|$ (*homogeneous*),

(iv) It satisfies the *triangle inequality* $\|x + y\| \leq \|x\| + \|y\|$.

A vector space that has a norm is called a *normed vector space.* ♢

Example A.7. The set $\mathbb{R}$ with the Euclidean norm. The set $\mathbb{C}$ with norm given by the absolute value function. ♢

Example A.8. The set $\mathbb{R}^n$ with norm the ℓ^p-norm, defined for each $x = (x_1, x_2, \ldots, x_n)$, $1 \leq p < \infty$, by $\|x\|_p = \Big(\sum_{k=1}^n |x_k|^p\Big)^{1/p}$. ♢

Example A.9. The bounded continuous functions f over an interval I (or $\mathbb{R}$) for which the uniform norm defined to be $\|f\|_{L^\infty(I)} := \sup_{x\in I} |f(x)|$ is finite. ♢

Example A.10. The continuous functions over a closed interval I (or $\mathbb{R}$) for which the L^p-norm defined by $\|f\|_{L^p(I)} = (\int_I |f(x)|^p\, dx)^{1/p}$ is finite. ♢

Example A.11. The Lebesgue spaces $L^p(I)$ over intervals I (or $\mathbb{R}$) with the L^p-norm. See Chapter 2 and Appendix A for more on these examples. ♢

Some vector spaces V come equipped with an extra piece of structure, called an inner product.

Definition A.12. A (*strictly positive-definite*) *inner product* on a vector space V over $\mathbb{C}$ is a function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$ such that, for all x, y, $z \in V$, α, $\beta \in \mathbb{C}$,

(i) $\langle x, y\rangle = \overline{\langle y, x\rangle}$. (*skew-symmetric* or *Hermitian.*)

(ii) $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$. (*Linear in the first variable.*)

(iii) $\langle x, x\rangle \geq 0$. (*Positive-definite.*)

(iv) $\langle x, x\rangle = 0 \iff x = 0$. $\big($(iii) *& (iv) together: strictly positive-definite.*$\big)$

A vector space with such an inner product is called an *inner-product vector space.* $\Diamond$

Items (i) and (ii) imply that the inner product is *conjugate-linear* in the second variable, in other words $\langle z, \alpha x + \beta y\rangle = \overline{\alpha}\langle z, x\rangle + \overline{\beta}\langle y, x\rangle$.

Every inner-product vector space V can be made into a normed space, since the quantity $\|x\| := \langle x, x\rangle^{1/2}$ is always a norm on V, as we ask you to verify in Exercise A.25. This norm is known as the *induced norm* on V.

Here are some examples of inner-product vector spaces over $\mathbb{C}$.

Example A.13. $V = \mathbb{C}^n$. The elements of $\mathbb{C}^n$ are of the form $z = (z_1, z_2, \ldots, z_n)$, with $z_j \in \mathbb{C}$. The inner product and its induced norm are defined for $z, w \in \mathbb{C}^n$ by $\langle z, w\rangle := z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n}$ and

$$\|z\|_{\mathbb{C}^n} := \langle z, z\rangle^{1/2} = (|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2)^{1/2}. \qquad \Diamond$$

Example A.14. (*Little ℓ^2 of the Integers, over $\mathbb{C}$*) $V = \ell^2(\mathbb{Z})$. The elements of $\ell^2(\mathbb{Z})$ are doubly infinite sequences of complex numbers, written $\{a_n\}_{n\in\mathbb{Z}}$, with $a_n \in \mathbb{C}$ and $\sum_{n\in\mathbb{Z}} |a_n|^2 < \infty$. The inner product and its induced norm are defined for $x = \{a_n\}_{n\in\mathbb{Z}}$ and $y = \{b_n\}_{n\in\mathbb{Z}}$ in $\ell^2(\mathbb{Z})$ by $\langle x, y\rangle := \sum_{n=-\infty}^{\infty} a_n\overline{b_n}$, and

$$\|x\|_{\ell^2(\mathbb{Z})} := \langle x, x\rangle^{1/2} = \Big(\sum_{n=-\infty}^{\infty} |a_n|^2\Big)^{1/2}. \qquad \Diamond$$

Example A.15. (*Square-Integrable Functions on the Circle*) $V = L^2(\mathbb{T})$. The inner product and its induced norm are defined for $f, g \in L^2(\mathbb{T})$ and $\mathbb{T} = [-\pi, \pi)$ by $\langle f, g\rangle := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}\,dx$ and

$$\|f\|_{L^2(\mathbb{T})} := \langle f, f\rangle^{1/2} = \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2\,dx\Big)^{1/2}. \qquad \Diamond$$

Example A.16. (*Square-Integrable Functions on the Line*) $V = L^2(\mathbb{R})$. The inner product and its induced norm are defined for $f, g \in L^2(\mathbb{R})$ by $\langle f, g\rangle := \int_{\mathbb{R}} f(x)\overline{g(x)}\,dx$ and

$$\|f\|_{L^2(\mathbb{R})} := \langle f, f\rangle^{1/2} = \Big(\int_R |f(x)|^2\,dx\Big)^{1/2}. \qquad \Diamond$$

The fact that the infinite sums in Example A.14, and the integrals in Examples A.15 and A.16, yield finite quantities is not automatic. It is a direct consequence of the indispensable *Cauchy–Schwarz Inequality*, which holds for all inner-product vector spaces and which we state, together with two other important inequalities, in Section A.1.1.

A.1.1. Orthogonality.

Orthogonality is one of the most important ideas in the context of inner-product vector spaces.

In the two-dimensional case of the vector space $V = \mathbb{R}^2$ over $\mathbb{R}$, the usual inner product has a natural geometric interpretation. Given two vectors $x = (x_1, x_2)$, $y = (y_1, y_2)$ in $\mathbb{R}^2$,

$$\langle x, y\rangle := x_1\overline{y_1} + x_2\overline{y_2} = \|x\|\,\|y\|\cos\theta,$$

where θ is the angle between the two vectors. In this setting, two vectors are said to be perpendicular or *orthogonal* when $\theta = \pi/2$, which happens if and only if $\langle x, y\rangle = 0$.

Let V be a vector space over $\mathbb{C}$, with inner product $\langle\cdot,\cdot\rangle$, and induced norm $\|\cdot\|$.

Definition A.17. Vectors $x, y \in V$ are *orthogonal* if $\langle x, y\rangle = 0$. We use the notation $x \perp y$. A set of vectors $\{x_\lambda\}_{\lambda\in\Lambda} \subset V$ is *orthogonal* if $x_{\lambda_1} \perp x_{\lambda_2}$ for all $\lambda_1 \neq \lambda_2$, $\lambda_i \in \Lambda$. Two subsets X, Y of V are *orthogonal*, denoted by $X \perp Y$, if $x \perp y$ for all $x \in X$ and $y \in Y$. ♢

Definition A.18. A collection of vectors $\{x_n\}_{n\in\mathbb{Z}}$ in an inner-product vector space is *orthonormal* if $\langle x_n, x_m\rangle = 1$ if $n = m$ and $\langle x_n, x_m\rangle = 0$ if $n \neq m$. ♢

Example A.19. Define vectors $e_k \in \mathbb{C}^N$ by $e_k(n) = (1/\sqrt{N})\, e^{2\pi ikn/N}$, for $k, n \in \{0, 1, 2, \ldots, N-1\}$. The vectors $\{e_k\}_{k=0}^{N-1}$ form an orthonormal family in $\mathbb{C}^N$ (Chapter 6). ♢

Example A.20. The trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are *orthonormal* in $L^2(\mathbb{T})$. ♢

Example A.21. The Haar functions are orthonormal in $L^2(\mathbb{R})$ (Chapter 9). ♢

The Pythagorean Theorem, the Cauchy–Schwarz Inequality, and the Triangle Inequality are well known in the two-dimensional case described above. They hold in every inner-product vector space V.

Theorem A.22 (Pythagorean Theorem)**.** *If x, $y \in V$, $x \perp y$, then*

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2.$$

Theorem A.23 (Cauchy–Schwarz Inequality)**.** *For all x, y in V,*

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

Theorem A.24 (Triangle Inequality)**.** *For all x, $y \in V$,*

$$\|x+y\| \leq \|x\| + \|y\|.$$

Exercise A.25. Prove Theorems A.22, A.23, and A.24, first for the model case $V = \mathbb{R}^2$, and then for a general inner-product vector space V over $\mathbb{C}$. Show that $\|x\| := \sqrt{\langle x, x\rangle}$ is a norm in V. ◊

Exercise A.26. Deduce from the Cauchy–Schwarz Inequality that if f and g are in $L^2(\mathbb{T})$, then $|\langle f, g\rangle| < \infty$. Similarly, if $x = \{a_n\}_{n\in\mathbb{Z}}$, $y = \{b_n\}_{n\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$, then $|\langle x, y\rangle| < \infty$. ◊

Exercise A.27. Suppose that V is an inner-product vector space, Λ is an arbitrary index set, and $\{x_\lambda\}_{\lambda\in\Lambda}$ is an orthogonal family of vectors. Show that these vectors are *linearly independent*; in other words, for any finite subset of indices $\{\lambda_1, \dots, \lambda_n\} \subset \Lambda$ such that n is no larger than the dimension of V (which may be infinite), we have $a_1 x_{\lambda_1} + \cdots + a_n x_{\lambda_n} = 0$ if and only if $a_1 = \cdots = a_n = 0$. ◊

A.1.2. Orthonormal bases. Orthogonality implies *linear independence* (Exercise A.27), and if our geometric intuition is right, orthogonality is in some sense the most linearly independent a set could be. In Example A.19 we exhibited a collection of N vectors in the N-dimensional vector space $\mathbb{C}^N$. Since these N vectors are linearly independent, they constitute a *basis* of $\mathbb{C}^N$. Moreover the set $\{e_k\}_{k=0}^{N-1}$ is an *orthonormal basis* of $\mathbb{C}^N$. We call it the *N-dimensional Fourier basis* (see Chapter 6).

The trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are an orthonormal basis of $L^2(\mathbb{T})$ (see Chapter 5), and the Haar functions are an orthonormal basis of $L^2(\mathbb{R})$ (see Chapter 9). It takes some time to verify these facts. Having an infinite orthonormal family (hence a linearly independent family) tells us that the space is infinite-dimensional. There is no guarantee that there are no other functions in the space orthogonal to the given orthonormal family, or in other words that *the orthonormal system is complete.*

Definition A.28. Consider an orthonormal family $\{x_n\}_{n\in\mathbb{N}}$ in an inner-product vector space V over $\mathbb{C}$. We say that the family is *complete*, or that $\{x_n\}_{n\in\mathbb{N}}$ is a *complete system*, or that the vectors in the family form an *orthonormal basis*, if given any vector $x \in V$, x can be expanded into a series of the basis elements which is convergent in the norm induced by the inner product. That is, there exists a sequence of complex numbers $\{a_n\}_{n\in\mathbb{N}}$ (the *coefficients*) such that $\lim_{N\to\infty} \Big\| x - \sum_{n=1}^{N} a_n\, x_n \Big\| = 0$. Equivalently, $x = \sum_{n=1}^{\infty} a_n x_n$ where equality holds in the norm of V. The coefficients a_n of x are uniquely determined. They can be calculated by pairing the vector with the basis elements: $a_n = \langle x, x_n\rangle$. $\Diamond$

How can we tell whether a given orthonormal family X in a vector space V is actually an orthonormal basis for V? One criterion for completeness is that the only vector orthogonal to all the vectors in X is the zero vector. Another is that Plancherel's Identity must hold. These ideas are summarized in Theorem A.29, which holds for all inner-product vector spaces V over $\mathbb{C}$. In Chapter 5 we prove Theorem A.29 for the particular case of $V = L^2(\mathbb{T})$ with the trigonometric basis. Showing that the same proofs extend to the general case is left as an exercise.

Theorem A.29. *Let V be an inner-product vector space over $\mathbb{C}$ and $X = \{x_n\}_{n\in\mathbb{N}}$ an orthonormal family in V. Then the following are equivalent:*

(i) *X is a complete system, hence an orthonormal basis of V.*

(ii) *If $x \perp X$ then $x = 0$.*

(iii) (Plancherel's Identity) *For all $x \in V$, $\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, x_n\rangle|^2$.*

A.2. Banach spaces and Hilbert spaces

Some, but not all, normed spaces satisfy another property, to do with convergence of Cauchy sequences of vectors. Namely, they are *complete normed vector spaces*, also known as *Banach spaces*. *Complete* *inner-product vector spaces* are known as *Hilbert spaces*. In this context we discuss unconditional bases in Banach spaces. We discuss orthogonal complements and orthogonal projections onto closed subspaces of a Hilbert space.

A.2.1. Banach spaces. In the presence of a norm we can talk about sequences of vectors converging to another vector.

Definition A.30. Let V be a normed space. The sequence $\{x_n\}_{n\in\mathbb{N}}$ in V *converges to* $x \in V$ if $\lim_{n\to\infty} \|x_n - x\| = 0$. $\Diamond$

It is well known that if a sequence is converging to a point, then as $n \to \infty$ the vectors in the sequence are getting closer to each other, in the sense that they form a *Cauchy sequence*.

Definition A.31. Let V be a normed space. The sequence $\{x_n\} \in V$ is *Cauchy* if for every $\varepsilon > 0$ there exists $N > 0$ such that $\|x_n - x_m\| < \varepsilon$ for all n, $m > N$. $\Diamond$

The converse is not always true, except in so-called *complete normed spaces* or *Banach spaces*. For instance, the Cauchy sequence $1, 1.4, 1.41, 1.414, 1.4142, \ldots$ in the space $\mathbb{Q}$ of rational numbers does not converge to any point in $\mathbb{Q}$, although it does converge to a point in the larger vector space $\mathbb{R}$. Furthermore, *every* Cauchy sequence of real numbers converges to some real number. $\mathbb{Q}$ is not complete; $\mathbb{R}$ is complete, and this is a fundamental fact about the real numbers.

Definition A.32. A normed space $\mathcal{B}$ is *complete* if every sequence in $\mathcal{B}$ that is Cauchy with respect to the norm of $\mathcal{B}$ converges to a limit in $\mathcal{B}$. Complete normed spaces are called *Banach spaces*. $\Diamond$

Example A.33. The continuous functions on a closed and bounded interval, with the uniform norm, form a Banach space. This fact and its proof are part of an advanced calculus course; for example see [**Tao06b**, Theorem 14.4.5]. $\Diamond$

Example A.34. The spaces $L^p(I)$ (Chapter 2) are Banach spaces. So are the spaces $L^p(\mathbb{R})$ (Chapter 7). The proof of these facts belongs to a more advanced course. However both results are crucial for us, and we assume them without proof. $\Diamond$

A.2.2. Bases, Unconditional bases. Given a Banach space X, a family of vectors $\{x_n\}_{n\geq 1}$ is a *Schauder basis* or simply a *basis* of X if every vector can be written in a unique fashion as an infinite linear combination of the elements of the basis, where the convergence is naturally in the norm of the Banach space.

Definition A.35. (*Schauder Basis*) Let X be a (real or complex) Banach space, then the sequence $\{x_n\}_{n\geq 1}$ is a *Schauder basis* (or simply a *basis*) of X if the closed linear span of $\{x_n\}_{n\geq 1}$ is X and $\sum_{n=1}^{\infty} a_n x_n$ is zero only if each a_n is zero. $\Diamond$

The second condition in the definition of a Schauder basis, asserting a certain kind of independence, clearly depends very much on the order of the x_n, and it is certainly possible for a permutation of a basis to fail to be a basis. Many naturally occurring bases, such as the standard bases in ℓ^p when $1 \leq p < \infty$, are bases under any permutation. Bases with the property that they remain a basis under permutation are called *unconditional bases.* We state as a theorem what amounts to several equivalent definitions of this concept. The spaces $C([0,1])$ and $L^1(\mathbb{R})$ do not have unconditional bases. See [**Woj91**, Chapter II.D].

Theorem A.36. *(*Unconditional basis equivalent conditions*) Let X be a (real or complex) Banach space and let $\{x_n\}_{n\geq 1}$ be a basis of X. Then the following are equivalent.*

(i) $\{x_{\pi(n)}\}_{n\geq 1}$ *is a basis of X for every permutation $\pi : \mathbb{N} \to \mathbb{N}$.*

(ii) *Sums of the form $\sum_{n=1}^{\infty} a_n x_n$ converge unconditionally whenever they converge.*

(iii) *There exists a constant $C > 0$ such that, for every sequence of scalars $\{a_n\}_{n\geq 1}$ and every sequence of scalars $\{\sigma_n\}_{n\geq 1}$ of modulus at most 1, we have the inequality*

$$\Big\| \sum_{n=1}^{\infty} \sigma_n a_n x_n \Big\|_X \leq C \Big\| \sum_{n=1}^{\infty} a_n x_n \Big\|_X .$$

A.2.3. Hilbert spaces. Some Banach spaces have the additional geometric structure of an inner product, as is the case for $L^2(\mathbb{T})$.

Definition A.37. A *Hilbert space* $\mathcal{H}$ is an inner-product vector space such that $\mathcal{H}$ is complete. In other words, every sequence in $\mathcal{H}$ that is Cauchy with respect to the norm induced by the inner product of $\mathcal{H}$ converges to a limit in $\mathcal{H}$. $\Diamond$

Here we are using the word *complete* with a different meaning than that in Definition A.28, where we consider *complete systems of orthonormal functions*. The context indicates whether we are talking about a *complete space* or a *complete system* $\{f_n\}$ *of functions*.

Here are some canonical examples of Hilbert spaces.

Example A.38. $\mathbb{C}^n$, a finite-dimensional Hilbert space. $\Diamond$

Example A.39. $\ell^2(\mathbb{Z})$, an infinite-dimensional Hilbert space. $\Diamond$

Example A.40. $L^2(\mathbb{T})$, the Lebesgue square-integrable functions on $\mathbb{T}$, consisting of the collection $\mathcal{R}(\mathbb{T})$ of all Riemann square-integrable functions on $\mathbb{T}$, together with all the functions that arise as limits of Cauchy sequences in $\mathcal{R}(\mathbb{T})$ (similarly to the construction of the real numbers from the rational numbers). In other words, $L^2(\mathbb{T})$ is the *completion* of $\mathcal{R}(\mathbb{T})$ with respect to the L^2-metric. $\Diamond$

Example A.41. $L^2(\mathbb{R})$, the Lebesgue square-integrable functions on$\mathbb{R}$. See Section A.3. $\Diamond$.

A.2.4. Orthogonal complements and projections. Given a closed subspace V of a Hilbert space $\mathcal{H}$, the collection of vectors orthogonal to the given subspace forms a new subspace denoted by $V^\perp$. We can then decompose the Hilbert space as a direct sum of the closed subspace and its orthogonal complement.

Definition A.42. Given a Hilbert space $\mathcal{H}$ and a closed subspace V of $\mathcal{H}$, the collection $V^\perp$ of vectors orthogonal to V is the *orthogonal complement of* V *in* $\mathcal{H}$. $\Diamond$

Lemma A.43. *Given a Hilbert space $\mathcal{H}$ and a closed subspace V then its orthogonal complement in $\mathcal{H}$, $V^\perp$, is a closed subspace, and $V \cap V^\perp = \{0\}$. Furthermore, $(V^\perp)^\perp = V$.*

One can write $\mathcal{H}$ as the orthogonal direct sum of a closed subspace V and its orthogonal complement.

Definition A.44. Let V, W be two subspaces of a vector space. Then their direct sum is denoted by $V \oplus W$ and defined to be $V \oplus W = \{x + y : x \in V, y \in W\}$. $\Diamond$

Lemma A.45. *Let V, W be closed subspaces of the Hilbert space $\mathcal{H}$. Then (i) $V \oplus W$ is a closed subspace of $\mathcal{H}$, and (ii) $V \oplus V^{\perp} = \mathcal{H}$.*

The proof of Lemma A.45(ii) follows from a best-approximation or orthogonal-projection theorem, presented next.

Theorem A.46 (Orthogonal Projection)**.** *Given any closed subspace V of a Hilbert space $\mathcal{H}$, and given $x \in \mathcal{H}$, there exists a unique vector $P_V x \in V$ that minimizes the distance in $\mathcal{H}$ to V. That is*

$$\|x - y\|_{\mathcal{H}} \geq \|x - P_V x\|_{\mathcal{H}} \qquad \textit{for all } y \in V.$$

Furthermore, the vector $x - P_V x$ is orthogonal to $P_V x$. If $\mathcal{B} = \{x_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of V, then $P_V x = \sum_{n\in\mathbb{N}} \langle x, x_n \rangle_{\mathcal{H}} x_n$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ denotes the inner product in $\mathcal{H}$.

One can then use pictures such us Figure 5.1 safely. Note that $x - P_V x \in V^{\perp}$ so it is now clear that we can decompose any vector $x \in \mathcal{H}$ as the (unique) sum of a vector in V and a vector in $V^{\perp}$, namely: $x = P_V x + (x - P_V x)$. Therefore $\mathcal{H} = V \oplus V^{\perp}$.

Definition A.47. Given any closed subspace V of a Hilbert space $\mathcal{H}$, and given $x \in \mathcal{H}$, the *orthogonal projection of x onto V* is the unique vector $P_V x \in V$ whose existence (and uniqueness) is guaranteed by Theorem A.46.

Example A.48. Let $\mathcal{H} = L^2(\mathbb{T})$, and $V = \mathcal{P}_n(\mathbb{T})$, the closed subspace of all trigonometric polynomials of degree less than or equal to N. Given $f \in L^2(\mathbb{T})$, the N^{th} partial Fourier sum of $f \in L^2(\mathbb{T})$, denoted by $S_N f$, is the one that best approximates f in the L^2-norm. In other words, $S_N f$ is the orthogonal projection of f onto $\mathcal{P}_n(\mathbb{T})$. $\Diamond$

Some useful properties of the orthogonal projections are listed.

Lemma A.49. *Given a closed subspace V of a Hilbert space $\mathcal{H}$, let $P_V : \mathcal{H} \to \mathcal{H}$ be the orthogonal projection mapping. Then*

(i) *for all $x \in \mathcal{H}$, $P_V x \in V$,* (ii) *if $x \in V$ then $P_V x = x$,*

(iii) $P_V^2 = P_V$, (iv) $(I - P_V)$ *is the orthogonal projection onto* $V^\perp$.

Given two nested closed subspaces $V_1 \subset V_2$ of $\mathcal{H}$, then the larger subspace V_2 can be viewed as a Hilbert space on its own. We can find the orthogonal complement of V_1 in V_2, call this subspace W_1, then

$$V_2 = V_1 \oplus W_1.$$

Lemma A.50. *Given two nested closed subspaces $V_1 \subset V_2$ of $\mathcal{H}$, denote by P_1 and P_2 the orthogonal projections onto V_1 and V_2 respectively. Let W_1 be the orthogonal complement of V_1 in V_2. Then the difference $Q_1x := P_2x - P_1x$ is the orthogonal projection of $x \in \mathcal{H}$ onto W_1.*

Proof. First observe that both P_2x and P_1x belong to the subspace V_2, and so does their difference, $Q_1x \in V_2$. Assume that Q_1x is orthogonal to P_1x, and that $Q_1x \in W_1$ (the reader should justify these assertions). To show that Q_1x is the orthogonal projection of x onto W_1 we need only to check that $x - Q_1x$ is orthogonal to Q_1x. But $\langle x - Q_1x, Q_1x\rangle = \langle x - P_2x + P_1x, Q_1x\rangle = \langle x - P_2x, Q_1x\rangle$, where the last equality holds because $P_1x \perp Q_1x$. Finally P_2x is the orthogonal projection of x onto V_2, so $x - P_2x$ is orthogonal to V_2, in particular to $Q_1x \in V_2$. We conclude that $\langle x - Q_1x, Q_1x\rangle = 0$. This defines Q_1 as the orthogonal projection onto the closed subspace W_1. □

A.3. $L^p(\mathbb{R})$, density, interchanging limits on $\mathbb{R}$

In this section we collect some useful facts from analysis on the line that we have used throughout the book. Here we consider functions defined on the real line $\mathbb{R}$. The parallel discussion in Chapter 2 treats functions defined on bounded intervals. We state most results without proof, but indicate where to find complete proofs in the literature.

We present first the Lebesgue spaces $L^p(\mathbb{R})$ which have been a common thread in Chapters 7– 12 (Section A.3.1); next a list of density results (Section A.3.2); and finally two theorems that give sufficient conditions under which we can interchange limit operations and integrals: Fubini's Theorem about interchanging integrals, and the Lebesgue Dominated Convergence Theorem about interchanging limits and integrals (Section A.3.3).

A.3.1. Lebesgue spaces $L^p(\mathbb{R})$ on the line. For each real number p with $1 \leq p < \infty$, the Lebesgue space $L^p(\mathbb{R})$ consists of those functions such that $\int_{\mathbb{R}} |f(x)|^p \, dx < \infty$ (Section 7.7). Here the integral is in the sense of Lebesgue. The analogous spaces $L^p(\mathbb{T})$ of functions on the unit circle are treated in Section 2.1. We now use the ideas of normed and inner-product vector spaces, Banach spaces and Hilbert spaces, recorded in Section A.2

All the $L^p(\mathbb{R})$ spaces are *normed*, with the L^p norm given by

$$\|f\|_p = \Big(\int_{\mathbb{R}} |f(x)|^p \, dx\Big)^{1/p}.$$

Each $L^p(\mathbb{R})$ spaces is also *complete*, meaning that every Cauchy sequence in the space converges and its limit is an element of the space. Therefore the $L^p(\mathbb{R})$ spaces are *Banach spaces*.

Let f be a continuous function that lies in $L^p(\mathbb{R})$. Then its L^p integral coincides with an improper Riemann integral. Further, the tails of the integral go to zero:

$$\lim_{N\to\infty} \int_{|x|\geq N} |f(x)|^p \, dx = 0. \tag{A.1}$$

In fact, this decay of the tail holds for all functions in $L^p(\mathbb{R})$, continuous or not.

The space $L^2(\mathbb{R})$ of square-integrable functions on the line is a *Hilbert space*, with *inner product* given by

$$\langle f, g\rangle = \int_{\mathbb{R}} f(x)\overline{g(x)} \, dx.$$

For $p \neq 2$, the spaces $L^p(\mathbb{R})$ are not Hilbert spaces.

We say that an identity $f = g$, or a limit $\lim_{t\to t_0} f_t = f$, holds *in the L^p-sense* if $\|f - g\|_p = 0$ or $\lim_{t\to t_0} \|f_t - f\|_p = 0$. Equality of f and g in the L^p-sense occurs if and only if $f = g$ except possibly on a set of measure zero, that is if $f = g$ a.e.

A set $E \subset \mathbb{R}$ has measure zero if given $\epsilon > 0$ there are intervals $\{I_j\}_{j\in\mathbb{N}}$ in $\mathbb{R}$ such that $E \subset \cup_{j\in\mathbb{N}} I_j$ and $\sum_{j\in\mathbb{N}} |I_j| < \epsilon$.

A function f is said to be *essentially bounded*, written $f \in L^\infty(\mathbb{R})$, if f is bounded by some fixed constant except possibly on a set of measure zero. The space $L^\infty(\mathbb{R})$ is a Banach space with norm given by $\|f\|_\infty := \operatorname{ess\,sup}_{x\in\mathbb{R}} |f(x)|$. Here the *essential supremum* of $|f(x)|$, denoted by $\operatorname{ess\,sup}_{x\in\mathbb{R}} |f(x)|$, is defined to be the smallest number M such that $|f(x)| \leq M$ for all $x \in \mathbb{R}$ except possibly on a set of measure zero. When the function f is actually bounded, then $\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$.

Functions in $L^p(\mathbb{R})$ are always locally integrable. In symbols we write $L^p(\mathbb{R}) \subset L^1_{\mathrm{loc}}(\mathbb{R})$. The Lebesgue spaces on the line are not nested, unlike the Lebesgue spaces on bounded intervals.

Here is a useful result connecting Lebesgue spaces on the line and on intervals. We used the case $p = 2$ when introducing the windowed Fourier transform in Section 9.2.

Lemma A.51. *A function $f : \mathbb{R} \to \mathbb{C}$ belongs to $L^p(\mathbb{R})$ if and only if the functions $f_n := f\chi_{[n,n+1)} \in L^p(\mathbb{R})$ and $\sum_{n\in\mathbb{Z}} \|f_n\|^p_{L^p(\mathbb{R})} < \infty$. Furthermore, $\|f\|^p_{L^p(\mathbb{R})} = \sum_{n\in\mathbb{Z}} \|f_n\|^p_{L^p(\mathbb{R})}$.*

Notice that $\|f_n\|^p_{L^p(\mathbb{R})} = \int_n^{n+1} |f(x)|^p\,dx$. Then Lemma A.51 is plausible given the additive property of the Lebesgue integral: if I, J are disjoint intervals then $\int_{I\cup J} = \int_I + \int_J$.

The intervals $\{[n, n+1)\}_{n\in\mathbb{Z}}$ can be replaced by any partition of $\mathbb{R}$ into countable measurable sets $\{A_n\}_{n\in\mathbb{N}}$. More precisely if $\mathbb{R} = \cup_{n\in\mathbb{N}} A_n$, where the sets A_n are mutually disjoint and measurable, then $\|f\|^p_{L^p(\mathbb{R})} = \sum_{n=1}^\infty \int_{A_n} |f(x)|^p\,dx$.

A.3.2. Density. The Schwartz class, introduced in Chapter 7, is *dense* in each $L^p(\mathbb{R})$. In other words, we can approximate any function $f \in L^p(\mathbb{R})$ by functions $\phi_n \in \mathcal{S}(\mathbb{R})$ such that the L^p norm of the difference $f - \phi_n$ tends to zero as n tends to infinity: $\lim_{n\to\infty} \|f - \phi_n\|_p = 0$. Equivalently, given $\varepsilon > 0$ there exists a $\phi \in \mathcal{S}(\mathbb{R})$ such that $\|f - \phi\|_p < \varepsilon$.

Theorem A.52. *The Schwartz functions on $\mathbb{R}$ are dense in $L^p(\mathbb{R})$ with respect to the L^p norm. Furthermore, $L^p(\mathbb{R})$ is the completion of $\mathcal{S}(\mathbb{R})$ in the L^p norm.*

In fact more is true: the smaller class of compactly supported infinitely differentiable functions (*bump functions*) is also dense in $L^p(\mathbb{R})$.

Here is a suite of density theorems in $\mathbb{R}$ leading to other density results that we have met in the book.

Theorem A.53. *Each continuous functions on $\mathbb{R}$ that is in $L^p(\mathbb{R})$ can be approximated with respect to the L^p norm by C^∞ functions on $\mathbb{R}$ that are also in $L^p(\mathbb{R})$.*

Proof. Consider a positive kernel in $\mathcal{S}(\mathbb{R})$, such as a normalized Gaussian. Use it to create a family of good kernels $K_n(x) = nK(nx)$ as $n \to \infty$. Take $f \in C(\mathbb{R}) \cap L^p(\mathbb{R})$. By the Hausdorff–Young inequality, $f_n := f * K_n \in L^p$, and by the approximation-of-the-identity result, $\|f_n - f\|_p \to 0$ as $n \to \infty$. By the smoothing property of convolution, f_n inherits the smoothness of K_n. Therefore $f_n \in C^\infty(\mathbb{R}) \cap L^p(\mathbb{R})$ (Chapter 7). □

Theorem A.54. *The continuous functions on $\mathbb{R}$ that are also in $L^p(\mathbb{R})$ are dense in $L^p(\mathbb{R})$ with respect to the L^p norm.*

Proof. Follows from Theorem A.53 and Theorem A.52, using the fact that if A is dense in $L^p(\mathbb{R})$ and $A \subset B$ then B is dense in $L^p(\mathbb{R})$. In this case $A = C^\infty(\mathbb{R}) \cap L^p(\mathbb{R})$ and $B = C(\mathbb{R}) \cap L^p(\mathbb{R}) \subset L^p(\mathbb{R})$. □

Theorem A.55. *The continuous functions on $\mathbb{R}$ with compact support are dense in $C(\mathbb{R}) \cap L^p(\mathbb{R})$, with respect to the L^p norm. Hence they are dense in $L^p(\mathbb{R})$.*

Proof. All we need to know here about the Lebesgue integral is that if $f \in C(\mathbb{R}) \cap L^p(\mathbb{R})$ then its L^p integral coincides with an improper Riemann integral, and the tails of the integral go to zero as in equation (A.1). Then one proceeds as in the Proof of Lemma 9.43 on page 239. In this proof one in fact reduces the problem to the Weierstrass approximation theorem on the compact intervals $[-N, N]$. □

The results in Chapter 9.4 imply the following facts.

Theorem A.56. *The step functions with compact support can uniformly approximate continuous functions with compact support. Hence they can approximate them in the L^p norm as well.*

Combining Theorems A.55 and A.56, we conclude the following.

Theorem A.57. *The step functions with compact support are dense in* $L^p(\mathbb{R})$.

A.3.3. Interchanging limits on the line. In this section we state two theorems we have used in the book, about interchanging limit operations and integrals. This section runs parallel to Section 2.3. However, some results that hold for bounded intervals cannot be carried over to the line. Notably, uniform convergence on the line is not sufficient for interchanging a limit and an integral over $\mathbb{R}$, as the discussion after Project 9.7 shows. Interchanging two integrals is a special case of these interchange-of-limit operations. The result that allows for such interchange is called Fubini's Theorem.

Theorem A.58 (Fubini's Theorem). *Let* $F : \mathbb{R}^2 \to \mathbb{C}$ *be a function. Suppose* $F \in L^1(\mathbb{R}^2)$, *that is* $\iint_{\mathbb{R}^2} |F(x,y)|\,dA < \infty$, *where* dA *denotes the differential of area or Lebesgue measure in* $\mathbb{R}^2$. *Then*

$$\int_{\mathbb{R}}\int_{\mathbb{R}} F(x,y)\,dx\,dy = \int_{\mathbb{R}}\int_{\mathbb{R}} F(x,y)\,dy\,dx = \iint_{\mathbb{R}^2} F(x,y)\,dA.$$

See [**SS05**, Chapter 2, Section 3], or any book in measure theory, for a proof of Fubini's Theorem on $\mathbb{R}^n$.

Interchanging limit operations is a delicate maneuver that is not always justified. We have illustrated throughout the book settings where this interchange is allowed and settings where it is illegal, such as the one just described involving uniform convergence.

Unlike the Riemann integral, the Lebesgue theory allows for the interchange of a pointwise limit and an integral. There are several landmark theorems that you learn in a course on measure theory. We state one such result.

Theorem A.59 (Lebesgue's Dominated Convergence Theorem). *Consider a sequence of measurable functions* f_n *defined on* $\mathbb{R}$, *converging pointwise a.e. to a function* f. *Suppose there exists a dominating function* $g \in L^1(\mathbb{R})$, *meaning that* $|f_n(x)| \leq g(x)$ *a.e. for all* $n > 0$. *Then*

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\,dx = \int_{\mathbb{R}} f(x)\,dx.$$

In other words, the limit of the integrals is the integral of the limit.

Bibliography

[Abb] S. Abbott, *Understanding analysis*, Undergraduate Texts in Mathematics, Springer, 2001.

[Ahl] L. V. Ahlfors, *Complex Analysis*, third ed., McGraw–Hill, 1979.

[Ari] J. Arias de Reyna, *Pointwise convergence of Fourier series*, Lecture Notes in Math. **1785**, Springer–Verlag, Berlin, 2002.

[Bae] A. Baernstein II, *Some sharp inequalities for conjugate functions*, Indiana Univ. Math. J. **27** (1978), 833–852.

[Bar66] R. G. Bartle, *The elements of integration and Lebesgue measure*, John Wiley & Sons, Inc., 1966.

[Bar01] R. G. Bartle, *A modern theory of integration*, Graduate Studies in Mathematics **32**, Amer. Math. Soc., Providence, RI, 2001.

[Bea] J. C. Bean, *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom*, Jossey–Bass, San Francisco, CA, 2001.

[Bec] W. Beckner, *Inequalities in Fourier Analysis on* $\mathbb{R}^n$, Proceedings of the National Academy of Sciences **72** (1975), 638–641.

[BK] C. Bénéteau and D. Khavinson, *The isoperimetric inequality via approximation theory and free boundary problems*, Comput. Methods Funct. Theory **6** No. 2 (2006), 253–274.

[BCR] G. Beylkin, R. Coifman, V. Rokhlin, *Fast wavelet transform and numerical algorithms*, Comm. Pure Appl. Math., **44** (1991), 141–183.

[Bôc] M. Bôcher, *Introduction to the theory of Fourier's series*, Ann. of Math. **7** (1906), 81–152.

[Bre] D. M. Bressoud, *A radical approach to real analysis*, Math. Assoc. of Amer., Washington, DC, 1994.

[Bri95] C. Brislawn, *Fingerprints go digital*, Notices Amer. Math. Soc. **42** No. 11 (1995), 1278–1283.

[Bri02] C. Brislawn, *The FBI Fingerprint Image Compression Standard*, `http://www.c3.lanl.gov/~brislawn/FBI/FBI.html`, last updated 2002.

[Bur] F. E. Burris, *Matrix formulation of the discrete Hilbert transform*, IEEE Transactions on Circuits and Systems **22** No. 10 (1975), 836–838.

[Cal] A. Calderón, *Intermediate spaces and interpolation, the complex method*, Studia Math. **24** (1964), 113–190.

[Can] E. J. Candés, *Compressive sampling*, Int. Congress of Mathematics Vol. III, 1433–1452, Eur. Math. Soc., Zürich, 2006.

[Car] L. Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135–157.

[CG] L. Carleson and T. W. Gamelin, *Complex Dynamics*, Universitext: Tracts in Mathematics, Springer–Verlag, New York, 1993.

[Chu] F. R. K. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics **92**, Amer. Math. Soc., Providence, RI, 1997.

[CMa06] R. R. Coifman and M. Maggioni, *Diffusion wavelets*, Appl. Comput. Harmon. Anal. **21** (2006), 53–94.

[CMa08] R. R. Coifman and M. Maggioni, *Geometry analysis and signal processing on digital data, emergent structures, and knowledge building*, SIAM News **41**, No. 10, December 2008.

[CMe] R. R. Coifman and Y. Meyer, *Remarques sur l'analyse de Fourier à fenêtre*, C. R. Acad. Sci. Paris, Série I **312** (1991), 259–261.

[CW] R. R. Coifman and V. Wickerhauser, *Entropy-based algorithms for best basis selection*, IEEE Trans. Information Theory **38** (1992), 713–718.

[CT] J. W. Cooley and J. W. Tukey, *An algorithm for the machine calculation of complex Fourier series*, Math. Comp. **19** (1965), 297–301.

[Cot] M. Cotlar, *A combinatorial inequality and its applications to L^2 spaces*, Rev. Mat. Cuyana **1** (1955), 41–55.

[CMP] D. Cruz-Uribe, J. M. Martell, C. Perez, *Sharp weighted estimates for approximating dyadic operators*, Electron. Res. Announc. Math. Sci. 17 (2010), 12–19.

[Dav] B. Davis, *On the weak-type* $(1, 1)$ *inequality for conjugate functions*, Proc. Amer. Math. Soc. **44** (1974), 307–311.

[Dau88] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure and Appl. Math. **41** (1988), 909–996.

[Dau92] I. Daubechies, *Ten lectures on wavelets*, CBMS–NSF Lecture Notes **61**, SIAM, Philadelphia, PA, 1992.

[DV] O. Dragičević and A. Volberg, *Sharp estimate of the Ahlfors–Beurling operator via averaging martingale transforms*, Mich. Math. J. **51** (2003), 415–435.

[Duo] J. Duoandikoetxea, *Fourier analysis*, English transl. by D. Cruz-Uribe, SFO, Graduate Studies in Mathematics **29**, Amer. Math. Soc., Providence, RI, 2001.

[DM] H. Dym and H. P. McKean, *Fourier series and integrals*, Probability and Mathematical Statistics **14**, Academic Press, New York–London, 1972.

[Eng] S. Engelberg, *Edge detection using Fourier coefficients*, Amer. Math. Monthly **115** No. 6 (2008), 499–513.

[Fol] G. B. Folland, *Real analysis: Modern techniques and their applications*, Wiley, New York, NY, 1999.

[Fra] M. W. Frazier, *An introduction to wavelets through linear algebra*, Springer–Verlag, New York, NY, 1999.

[Gab] D. Gabor, *Theory of communication*, J. Inst. Electr. Eng., London **93** (III) (1946), 429–457.

[Gar] J. B. Garnett, *Bounded Analytic Functions*, Academic Press, Orlando, FL, 1981.

[Gib1898] J. W. Gibbs, *Fourier's series*, letter in Nature **59** (1898), 200.

[Gib1899] J. W. Gibbs, *Fourier's series*, letter in Nature **59** (1899), 606.

[GS] D. Gottlieb and C.-W. Shu, *On the Gibbs phenomenon and its resolution*, SIAM Rev. **39** No. 4 (1997), 644–668.

[Grab] J. V. Grabiner, *Who gave you the epsilon? Cauchy and the origins of rigorous calculus*, Amer. Math. Monthly **90** No. 3 (1983), 185–194.

[Graf94] L. Grafakos, *An elementary proof of the square summability of the discrete Hilbert transform*, Amer. Math. Monthly **101** No. 5 (1994), 456–458.

[Graf08] L. Grafakos, *Classical Fourier analysis*, second ed., Graduate Texts in Mathematics **249**, Springer, New York, 2008.

[Graf09] L. Grafakos, *Modern Fourier analysis*, second ed., Graduate Texts in Mathematics **250**, Springer, New York, 2009.

[GM] L. Grafakos and S. Montgomery-Smith, *Best constants for uncentered maximal functions*, Bull. London Math. Soc. **29** (1997), 60–64.

[Grat] I. Grattan-Guinness, *The development of the foundations of mathematical analysis from Euler to Riemann*, MIT Press, Cambridge, MA, 1970.

[Gre07b] B. Green, *Long arithmetic progressions of primes*, Gaz. Math. No. 112 (2007), 26–27.

[Haa] A. Haar, *Zur Theorie der orthogonalen Funktionensysteme*, Math. Annal. **69** (1910), 331–371.

[HL] G. H. Hardy and J. E. Littlewood, *A maximal theorem with function theoretic applications*, Acta Math. **54** (1930), 81–116.

[HW] E. Hernandez and G. Weiss, *A first course on wavelets*, CRC Press, Boca Raton, FL, 1996.

[Hig] N. J. Higham, *Handbook of Writing for the Mathematical Sciences*, SIAM, Philadelphia, PA, 1998.

[HT] M. Holschneider and Ph. Tchamitchian, *Pointwise analysis of Riemann's "nondifferentiable" function*, Invent. Math. **105** (1991), 157–175.

[Hun] R. Hunt, *On the convergence of Fourier series*, Orthogonal expansions and their continuous analogues (Proc. Conf., Edwardsville, IL, 1967), Southern Illinois Univ. Press, Carbondale, IL, 1968, 235–255.

[Hyt08] T. Hytönen, *On Petermichl's dyadic shift and the Hilbert transform*, C. R. Math. Acad. Sci. Paris **346** (21–22) (2008), 1133–1136.

[Hyt12] T. Hytönen, *The sharp weighted bound for general Calderón–Zygmund operators*. To appear, Ann. Math (2012).

[Jaf95] S. Jaffard, *Local behavior of Riemann's function*, Contemp. Mat. **189** (1995), 278–307.

[Jaf97] S. Jaffard, *Old friends revisited: The multifractal nature of some classical functions*. J. Fourier Anal. Appl. **3** (1997), 1–22.

[JMR] S. Jaffard, Y. Meyer, and R. D. Ryan, *Wavelets: Tools for science and technology*, revised ed., SIAM, Philadelphia, PA, 2001.

[KK] J.-P. Kahane and Y. Katznelson, *Sur les ensembles de divergence des séries trigonométriques*, Studia Math. **26** (1966), 305–306.

[KL] J.-P. Kahane and P.-G. Lemarié-Rieusset, *Fourier series and wavelets*, Gordon and Breach Publishers, 1995.

[Kat] Y. Katznelson, *Introduction to harmonic analysis*, Dover Publications Inc., 1976.

[Kol] A. Kolmogorov, *Une série de Fourier–Lebesgue divergente partout*, C. R. Acad. Sci. Paris **183** (1926), 1327–1328.

[Koo] P. Koosis, *Introduction to the H_p spaces*, Cambridge University Press, Cambridge, 1980.

[Kör] T. W. Körner, *Fourier analysis*, Cambridge University Press, Cambridge, 1988.

[Kra] B. Kra, *The Green–Tao theorem on arithmetic progressions in the primes: an ergodic point of view*, Bull. Amer. Math. Soc. (N.S.) **43** No. 1 (2006), 3–23.

[Kran] S. G. Krantz, *A panorama of harmonic analysis*, The Carus Mathematical Monographs **27**, Math. Assoc. Amer., Providence, RI, 1999.

[KRWW] I. A. Kryshtal, B. D. Robinson, G. L. Weiss, and E. N. Wilson, *Some simple Haar-type wavelets in higher dimensions*, J. Geom. Anal. **17**, No. 1 (2007), 87–96.

[Lac] M. Lacey, *Haar Shifts, Commutators, and Hankel Operators*, Rev. Un. Mat. Argentina **50** No. 2 (2009), 1–13.

[LPR] M. Lacey, S. Petermichl, and M. Reguera-Rodriguez, *Sharp A_2 inequality for Haar shift operators*, Math. Ann. **348** (2010), no. 1, 127–141.

[Ler] A. Lerner, *An elementary approach to several results on the Hardy–Littlewood maximal operator*, Proc. Amer. Math. Soc. **136** No. 8 (2008), 2829–2833.

[LL] E. Lieb and M. Loss, *Analysis*, Graduate Studies in Mathematics **14**, Amer. Math. Soc., Providence, RI, 2001.

[LC] G. M. Livadas and A. G. Constantinides, *Image edge detection and segmentation based on the Hilbert transform*, Proceedings of ICASSP-88 conference, vol. 2 (1988), 1152–1155.

[LDPU] M. Lyatti, Y. Divin, U. Poppe and K. Urban, *Liquid identification by Hilbert spectroscopy*, Supercond. Sci. Technol. **22** (2009), 114005, 8pp.

[Mall89] S. Mallat, *Multiresolution approximations and wavelet orthonormal bases for $L^2(\mathbb{R})$*, Trans. Amer. Math. Soc. **315** (1989), 69–87.

[Mall98] S. Mallat, *A wavelet tour of signal processing*, Academic Press, San Diego, CA, 1998.

[Malv] H. Malvar, *Lapped transforms for efficient transform/subband coding*, IEEE Trans. Acoust. Speech Signal Process. **38** (1990), 969–978.

[Mel] A. Melas, *The best constant for the centered Hardy–Littlewood maximal inequality*, Ann. of Math. **157** (2003), 647–688.

[Mey] Y. Meyer, *Wavelets: Algorithms & applications*, translated from the French and with a foreword by Robert D. Ryan, SIAM, Philadelphia, PA, 1993.

[Mic] A. A. Michelson, *Fourier's series*, Letter in Nature **58** (1898), 544–545.

[MS] A. A. Michelson and S. W. Stratton, *A new harmonic analyser*, Philosophical Magazine **45** (1898), 85–91.

[MP] M. Mohlenkamp and M. C. Pereyra, *Wavelets, their friends, and what they can do for you*, EMS Series of Lectures in Mathematics, European Math. Soc. Publishing House, Zurich, 2008.

[Mun] J. R. Munkres, *Topology: A first course*, Prentice Hall, Englewood Cliffs, NJ, 1975.

[Nat] M. Nathanson, *Elementary Methods in Number Theory*, Graduate Texts in Mathematics **195**, Springer, New York, NY, 2000.

[NTV97] F. Nazarov, S. Treil, and A. Volberg, *Cauchy integral and Calderón–Zygmund operators on nonhomogeneous spaces*, Int. Math. Res. Not. (1997), 703–726.

[NTV08] F. Nazarov, S. Treil, and A. Volberg, *Two weight inequalities for individual Haar multipliers and other well localized operators*, Math. Res. Lett. **15** No. 3 (2008), 583–597.

[Per] M. C. Pereyra, *Lecture notes on dyadic harmonic analysis*, Contemp. Math. **289** (2001), Amer. Math. Soc., Ch. I, 1–61.

[Pet] S. Petermichl, *Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol*, C.R. Acad. Sci. Paris, Sér. I Math. **330** (2000), 455–460.

[PTV] S. Petermichl, S. Treil, and A. Volberg, *Why the Riesz transforms are averages of the dyadic shifts*, Publ. Mat. Barc., Spec. Vol. (2002), 209–228.

[Pin] M. A. Pinsky, *Introduction to Fourier analysis and wavelets*, The Brooks/Cole Series in Advanced Mathematics (Paul J. Sally, Jr, ed.), Brooks/Cole, 2001.

[Pre] E. Prestini, *The evolution of applied harmonic analysis: Models of the real world*, Birkhäuser, Boston, MA, 2004.

[RS] M. Reed and B. Simon, *Methods of modern mathematical physics, Vol. II: Fourier analysis, self-adjointness*, Academic Press, 1975.

[Roy] H. L. Royden, *Real analysis*, 2nd ed., MacMillan, New York, NY, 1968.

[Rud] W. Rudin, *Principles of mathematical analysis*, McGraw–Hill, New York, NY, 1976.

[Sad] C. Sadosky, *Interpolation of operators and singular integrals*, Marcel Dekker, Inc., New York, NY, 1979.

[Sch] M. Schechter, *Principles of functional analysis*, Graduate Studies in Mathematics **36**, 2nd ed., Amer. Math. Soc., Providence, RI, 2002.

[Sha] C. E. Shannon, *Communications in the presence of noise*, Proc. Inst. Radio Eng. **37** (1949), 10–21.

[Spi] M. Spiegel, *Schaum's outline of Fourier analysis with applications to boundary value problems*, McGraw–Hill, New York, NY, 1974.

[Ste] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Univ. Press, Princeton, NJ, 1970.

[SS03] E. M. Stein and R. Shakarchi, *Fourier analysis: an introduction*, Princeton Lectures in Analysis I, Princeton Univ. Press, Princeton, NJ, 2003.

[SS05] E. M. Stein and R. Shakarchi, *Real analysis: Measure theory, integration, and Hilbert spaces*, Princeton Lectures in Analysis III, Princeton Univ. Press, Princeton, NJ, 2005.

[SW59] E. M. Stein and G. Weiss, *An extension of a theorem of Marcinkiewicz and some of its applications*, J. Math. Mech. **8** (1959), 263–264.

[SW71] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton Univ. Press, Princeton, NJ, 1971.

[Stra88] G. Strang, *Linear algebra and its applications*, Saunders College Publishing, third ed., 1988.

[Stra94] G. Strang, *Wavelets*, American Scientist **82**, April 1994, 250–255, reprinted in Appendix 1 of [**SN**].

[SN] G. Strang and T. Nguyen, *Wavelets and filter banks*, Wellesley–Cambridge Press, Wellesley, MA, 1996.

[SSZ] G. Strang, V. Strela and D.-X. Zhou, *Compactly supported refinable functions with infinite masks*, in The functional and harmonic analysis of wavelets and frames, L. Baggett and D. Larson, eds., Contemp. Math. **247** (1999), Amer. Math. Soc., 285–296.

[Stri94] R. S. Strichartz, *A guide to distribution theory and Fourier transforms*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1994.

[Stri00a] R. S. Strichartz, *The way of analysis*, Jones & Bartlett Publishers, revised ed., 2000.

[Stri00b] R. S. Strichartz, *Gibbs' phenomenon and arclength*, J. Fourier Anal. and Appl. **6** Issue 5 (2000), 533–536.

[Tao05] T. Tao, *An uncertainty principle for cyclic groups of prime order*, Math. Research Letters **12** (2005), 121–127.

[Tao06a] T. Tao, *Analysis I*, Text and Readings in Mathematics **37**, Hindustan Book Agency, India, 2006.

[Tao06b] T. Tao, *Analysis II*, Text and Readings in Mathematics **37**, Hindustan Book Agency, India, 2006.

[Tao06c] T. Tao, *Arithmetic progressions and the primes*, Collect. Math., Vol. Extra (2006), 37–88.

[Ter] A. Terras, *Fourier analysis on finite groups and applications*, London Math. Soc. Student Texts **43**, Cambridge University Press, Cambridge, 1999.

[Tor] A. Torchinsky, *Real-variable methods in harmonic analysis*, Pure and Applied Mathematics **123**, Academic press, Inc., San Diego, CA, 1986.

[Tre] A. Treibergs, *Inequalities that imply the isoperimetric inequality*, `http://www.math.utah.edu/~treiberg/isoperim/isop.pdf`.

[Vag] A. Vagharshakyan, *Recovering singular integrals from Haar shifts*, Proc. Amer. Math. Soc. **138** (2010), no. 12, 4303–4309.

[VK] M. Vetterli and J. Kovačević, *Wavelets and subband coding*, Prentice Hall, Upper Saddle River, NJ, 1995.

[Wil] H. Wilbraham, *On a certain periodic function*, Cambridge & Dublin Math. J. **3** (1848), 198–201.

[Wic] V. Wickerhauser, *Adapted wavelet analysis from theory to software*, A. K. Peters, Wellesley, 1994.

[Wien] N. Wiener, *The ergodic theorem*, Duke Math. J. **5** (1939), 1–18.

[Wieg] J. Wiegert, *The Sagacity of Circles: A History of the Isoperimetric Problem*, `http://mathdl.maa.org/mathDL/23/` (keyword: sagacity).

[Woj91] P. Wojtaszczyk, *Banach spaces for analysts*, Cambridge Studies in Advanced Mathematics **25**, Cambridge Univ. Press, Cambridge, 1991.

[Woj03] P. Wojtaszczyk, *A mathematical introduction to wavelets*, London Math. Soc. Student Texts **37**, Cambridge Univ. Press, Cambridge, 2003.

[Zyg59] A. Zygmund, *Trigonometric Series*, 2nd ed., Cambridge Univ. Press, Cambridge, 1959.

[Zyg71] A. Zygmund, *Intégrales Singuliéres*, Lecture Notes in Math. **204**, Springer–Verlag, Berlin, 1971.

Name Index

Subject Index

Page numbers are printed in **boldface** when they represent the definition of, or the main source of information about, the topic being indexed. For brevity, we sometimes use "F." to denote "Fourier", "H. I." for "Hölder Inequality", "M. I." for "Minkowski Inequality", and "trig." for "trigonometric".