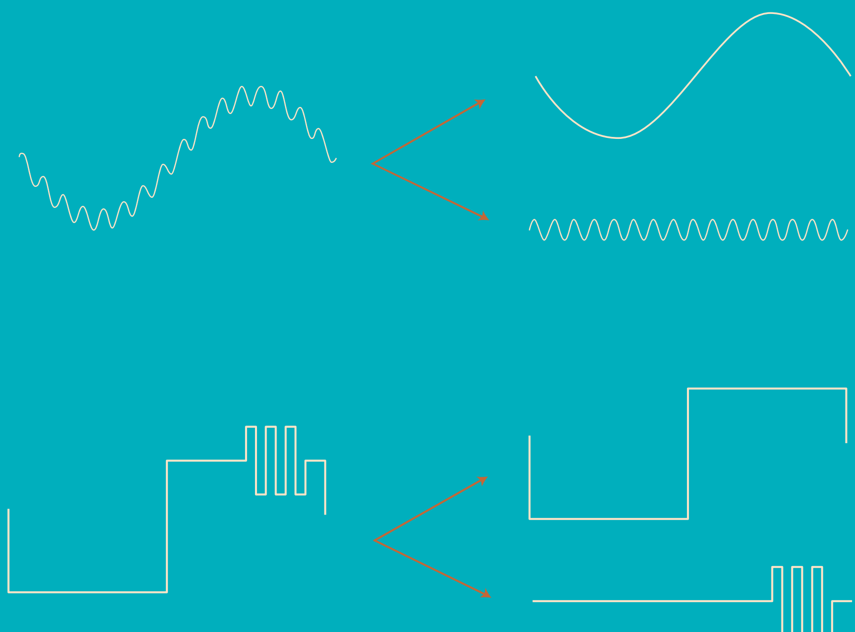
STUDENT MATHEMATICAL LIBRARY
IAS/PARK CITY MATHEMATICAL SUBSERIES
Volume 63

Harmonic Analysis
From Fourier to Wavelets

María Cristina Pereyra
Lesley A. Ward



American Mathematical Society
Institute for Advanced Study

Harmonic Analysis

From Fourier to Wavelets

STUDENT MATHEMATICAL LIBRARY
IAS/PARK CITY MATHEMATICAL SUBSERIES
Volume 63

Harmonic Analysis

From Fourier to Wavelets

María Cristina Pereyra
Lesley A. Ward

American Mathematical Society, Providence, Rhode Island
Institute for Advanced Study, Princeton, New Jersey

2010 *Mathematics Subject Classification.* Primary 42–01;
Secondary 42–02, 42Axx, 42B25, 42C40.

The anteater on the dedication page is by Miguel.
The dragon at the back of the book is by Alexander.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-63

Library of Congress Cataloging-in-Publication Data

Pereyra, María Cristina.
Harmonic analysis : from Fourier to wavelets / María Cristina Pereyra, Lesley A. Ward.
p. cm. — (Student mathematical library ; 63. IAS/Park City mathematical subseries)
Includes bibliographical references and indexes.
ISBN 978-0-8218-7566-7 (alk. paper)
1. Harmonic analysis—Textbooks. I. Ward, Lesley A., 1963– II. Title.

QA403.P44 2012
515′.2433—dc23

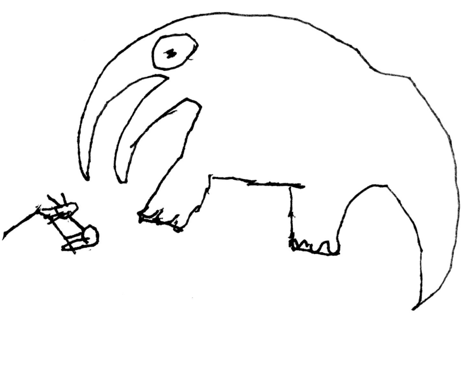
2012001283

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 17 16 15 14 13 12

To Tim, Nicolás, and Miguel

To Jorge and Alexander



Contents

List of figures

List of tables

IAS/Park City Mathematics Institute

The IAS/Park City Mathematics Institute (PCMI) was founded in 1991 as part of the "Regional Geometry Institute" initiative of the National Science Foundation. In mid-1993 the program found an institutional home at the Institute for Advanced Study (IAS) in Princeton, New Jersey. The PCMI continues to hold summer programs in Park City, Utah.

The IAS/Park City Mathematics Institute encourages both research and education in mathematics and fosters interaction between the two. The three-week summer institute offers programs for researchers and postdoctoral scholars, graduate students, undergraduate students, high school teachers, mathematics education researchers, and undergraduate faculty. One of PCMI's main goals is to make all of the participants aware of the total spectrum of activities that occur in mathematics education and research: we wish to involve professional mathematicians in education and to bring modern concepts in mathematics to the attention of educators. To that end the summer institute features general sessions designed to encourage interaction among the various groups. In-year activities at sites around the country form an integral part of the High School Teacher Program.

Each summer a different topic is chosen as the focus of the Research Program and Graduate Summer School. Activities in the Undergraduate Program deal with this topic as well. Lecture notes from the Graduate Summer School are published each year in the IAS/Park City Mathematics Series. Course materials from the Undergraduate Program, such as the current volume, are now being published as part of the IAS/Park City Mathematical Subseries in the Student Mathematical Library. We are happy to make available more of the excellent resources which have been developed as part of the PCMI.

John Polking, Series Editor
February 2012

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 one of the most influential bodies 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 intervals. In the Appendix we review vector, normed, and inner-product spaces, as well as some key concepts from analysis on the real line.

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 notation $A := B$ 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 and 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, one could 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 on dyadic harmonic analysis make the book distinctive, and we would include that material.

The twenty-four 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 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 projects and their next steps, once they have sunk 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.

Some parts of Chapters 10 and 11 are drawn from material in the first author's book [**MP**]. This material originated in lecture notes by the first author for minicourses delivered by her (in Argentina in 2002 and in Mexico in 2006) and by Martin Mohlenkamp (in Honduras in 2004; he kindly filled in for her when her advanced pregnancy made travel difficult).

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 its 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 Wilfredo Urbina, University of New Mexico graduate students Adam Ringler and Elizabeth Kappilof, and Brown University graduate student Constance Liaw. We thank Jorge Aarão for reading and commenting on our almost-final manuscript.

We thank the following students who wrote project reports, gave presentations, and in some cases came up with research topics:

Kourosh Raeen (the project in Section 3.4), Adam Ringler (the project in Section 6.9), Mike Klopfer (the project in Section 10.6), and Bernadette Mendoza-Spencer (the project in Section 12.9.1).

Abbreviated versions of the projects in Sections 3.4, 4.8, and 9.7 were tested as guided one-day projects with fifteen to eighteen 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 go 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 from the US to Australia, and energetic toddlers grew into thriving ten- and eleven-year-old children. Our 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
February 2012

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, the Haar transform and other 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

[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 and 93].

Taylor series (Section 1.2). We introduce and motivate the mathematical definition of Fourier coefficients and series and discuss periodic functions (Section 1.3). We describe the physical problem of heat diffusion in a one-dimensional bar, which 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. (See the two examples shown on the cover.) 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 as 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.

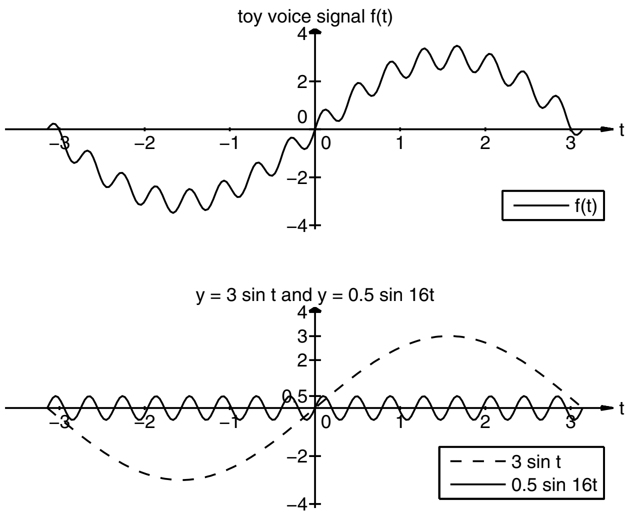


Figure 1.1. Toy voice signal $f(t)$ (upper plot) and the building blocks of the toy voice signal (lower plot). See also the upper figure on the cover.

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. In other words, they complete 1 and 16 full oscillations, respectively, 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 imperfectly 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 and how many building blocks can be thrown away while retaining a recognizable signal. This topic 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. ◊

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

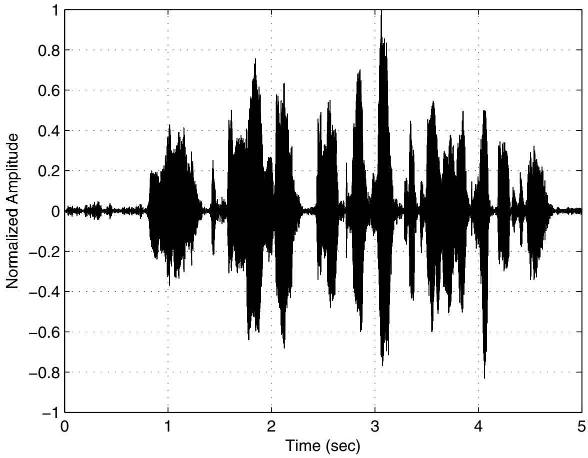


Figure 1.2. Plot of a voice saying, "Hi Mom, it's Amanda, and I can't wait to tell you what happened today." The horizontal axis represents time, in seconds. The vertical axis represents the amplitude (volume) of the signal. The units of amplitude are decibels, but here we have normalized the amplitude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

analysis with physical problems (for instance heat diffusion and the vibrating string) in the early eighteenth century; see Section 1.4. This 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 have used θ instead of t for the independent variable here, and sometimes we will use x instead.

We can rewrite the right-hand side of (1.1) as a linear combination of exponential functions. To do so, we use *Euler's Formula*[2]: $e^{i\theta} = \cos\theta + i\sin\theta$, 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. In general a_n, b_n, and c_n are complex numbers. 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?

[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 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.

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?

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 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 eighteenth 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 differentiable functions. That dream was shattered by Cauchy's[4] discovery in 1821 of a counterexample.

[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).

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 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 $x = 0$. ♢

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

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. ♢

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 $-M \leq n \leq M$, $n \in \mathbb{N}$. ♢

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

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).

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

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. ◊

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

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

[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).

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* with respect to an appropriately normalized inner product.

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. ♢

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. ♢

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 $f(\theta) = 3e^{-2i\theta} - e^{-i\theta} + 1 + e^{i\theta} - \pi e^{4i\theta} + (1/2)e^{7i\theta}$. ♢

---

[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] and [**Tao06b**, Chapters 19].

Example 1.12 (*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

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

It follows that

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

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 $\theta \in (-\pi, \pi)$. Surprisingly, at the endpoints $\theta = \pm\pi$ the series converges to zero (since $\sin(n\pi) = 0$ for all $n \in \mathbb{N}$) and not to the values $f(\pm\pi) = \pm\pi$. Figure 3.1 offers a clue: zero is the midpoint of the jump in height from $+\pi$ to $-\pi$ of the periodic extension of $f(\theta)$. ♢

Exercise 1.13. Use integration by parts to complete the calculation of $\widehat{f}(n)$ in Example 1.12. ♢

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

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

that are complex-valued and that are 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. This periodic extension is called the *periodic ramp function* or *sawtooth function*, since its graph resembles the teeth of a saw.

Exercise 1.14. Use MATLAB to reproduce the plot of the sawtooth function 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 $\theta \in [-\pi, \pi)$. ♢

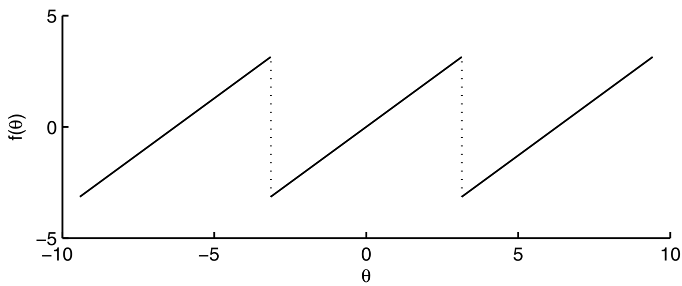


Figure 1.3. Graph of the *sawtooth function* or *periodic ramp function* given by extending $f(\theta) = \theta$ periodically from the interval $[-\pi, \pi)$ to $[-3\pi, 3\pi)$.

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

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

The building blocks we're 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*:

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

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 $\mathbb{R}$, and 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 $0 \le \ell \le k$, where $f^{(l)}$ denotes the ℓ^{th} derivative of f.

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 $a \in \mathbb{R}$. ♢

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 formulas 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. ♢

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 $[a,b)$, or equivalently periodic functions on $\mathbb{R}$, and functions f that are defined on $\mathbb{R}$ but that are not periodic. For nonperiodic functions on $\mathbb{R}$, it turns out that the natural "Fourier quantity" is not a Fourier series but another function, known as the *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, $f(\theta) = \sum_{n=-M}^{M} a_n e^{2\pi in\theta/L}$. Verify that f coincides with its L-Fourier series. ♢

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 mod-*

> *ern 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 eighteenth, nineteenth, and early twentieth centuries: Brook Taylor (1685–1731), Daniel Bernoulli (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

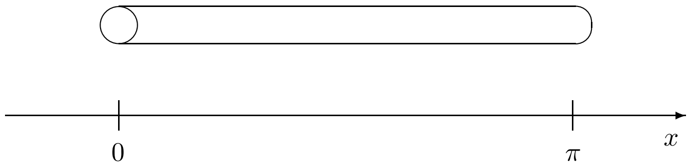


Figure 1.4. Sketch of a one-dimensional bar.

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. (For example, in Table 1.1 we consider a bar with insulated ends, so that $u_x(0,t) = u_x(\pi,t) = 0$.)

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 $\alpha(x)\beta'(t) = \alpha''(x)\beta(t)$. Then

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

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

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, in the subsequent projects we omit mention of these expectations.

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). Here are 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		
$u_{tt} = u_{xx}$	$u(0,t) = u(\pi,t) = 0,$ $u(x,0) = \sum_n b_n \sin(nx)$	$u(x,t) = \sum_n b_n e^{int} \sin(nx)$
Temperature in a bar with insulated ends		
$u_t = u_{xx}$	$u_x(0,t) = u_x(\pi,t) = 0,$ $u(x,0) = \sum_n c_n \cos(nx)$	$u(x,t) = \sum_n c_n e^{-n^2 t} \cos(nx)$
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		
$u_{xx} + u_{yy} = 0$	$u(x,0) = \sum_k A_k \sin(kx),$ $u(x,1) = \sum_k B_k \sin(kx)$	$u(x,y) = \sum_k \left(\frac{\sinh[k(1-y)]}{\sinh k} A_k + \frac{\sinh(ky)}{\sinh k} B_k\right) \sin(kx)$

Chapter 2

Interlude: Analysis concepts

This chapter is a compendium of background material.We omit the proofs, instead giving pointers to where they may be found in the literature.

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 or recurrent theme 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 $I = \bigcup_{k=1}^n J_k$. The intervals J_k are contained in I and are necessarily bounded. $\Diamond$

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:

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

The building blocks in 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. $\Diamond$

[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 is the collection of disjoint 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 $[-5, 7)$, he or 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 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. $\diamondsuit$

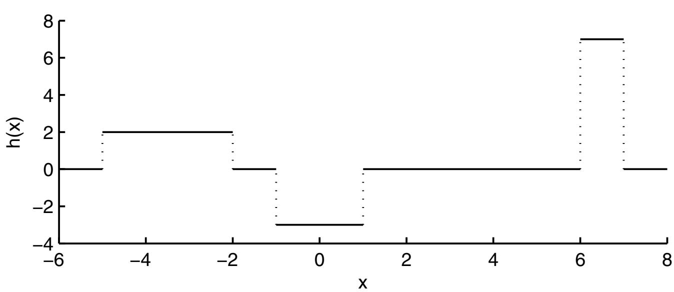


Figure 2.1. Graph of the step function in Example 2.4: $h(x) = 2\chi_{[-5,-2)}(x) - 3\chi_{[-1,1)}(x) + 7\chi_{[6,7)}(x)$.

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 $\mathcal{R}(I)$. $\Diamond$

Exercise 2.9. Show that step functions are Riemann integrable. $\Diamond$

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*

$$\left|\int_I f\right| \leq \int_I |f|$$

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

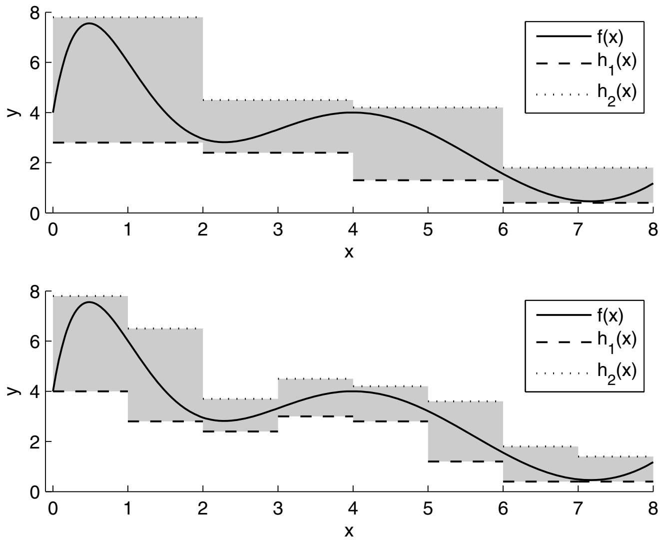


Figure 2.2. This figure illustrates the definition of the Riemann integral. We see the approximation of a function f from below by a step function h_1 and from above by a step function h_2, at a coarse scale (upper plot) and at a finer scale (lower plot). The upper and lower plots use different partitions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

Exercise 2.10 (*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 be different. Try to find a partition that works for both (a *common refinement*). $\Diamond$

Exercise 2.11 (*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. ♢

One can deduce the following useful proposition from Definition 2.8.

Proposition 2.12. *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 2.13. 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. ♢

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

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

[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) Riemann 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.

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.

Here are some examples of familiar functions that are Riemann integrable and for which we can define Fourier coefficients. See Theorem 2.33 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]. ♢

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

Exercise 2.16. 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. ♢

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

Example 2.17 (*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. ♢

Exercise 2.18. Show that the Dirichlet function is not Riemann integrable. ♢

[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).

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].

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 2.17 is Lebesgue integrable but not Riemann integrable, with Lebesgue integral equal to zero; see Remark 2.38. 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 nonmeasurable, 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

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

[7] For a brief introduction see [**Tao06b**, Chapters 18 and 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**].

[8] In statistics and probability the characteristic function of a set A is called the *indicator function* of the set and is 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).

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.

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 2.34 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 2.38) with the L^p

[9]See the Appendix for the definition of a normed space.

norm defined by

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

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

In Definition 2.35 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 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 2.19. *For each p with $1 \le p \le \infty$, 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 in the L^p sense to a function in $L^p(I)$, and the set $\mathcal{R}(I)$ is* dense[12] *in $L^p(I)$.*

Aside 2.20. 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]. ♢

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

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

---

[10] Named after the Polish mathematician Stefan Banach (1892–1945). See the Appendix 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**].

[11] Named after the same Cauchy as the counterexample in Example 1.3. 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 $\|f - g\|_{L^p(I)} \le \epsilon$. See Section 2.4.

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 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 2.21. 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 on page 33.

---

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

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

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.

Exercise 2.22. 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 $x = 0$. ♢

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 $[-\pi, \pi)$. 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, $\mathcal{P}_N(\mathbb{T}) := \left\{ \sum_{|n| \leq N} a_n e^{in\theta} : a_n \in \mathbb{C},\ |n| \leq N \right\}$.

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

$\mathcal{M}(\mathbb{T})$	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
$\cup$	
$\square$	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', $\dots$, $f^{(k)}$ exist and are continuous
$\cup$	
$\vdots$	$\vdots$
$\cup$	
$C^\infty(\mathbb{T})$	infinitely differentiable: f, f', f'', $\dots$ 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 2.23.

all the way to $C^\infty(\mathbb{T})$, by including the nested spaces of trigonometric polynomials. See the lowest part of the ladder in Figure 2.3.

Remark 2.23. 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 2.27). $\diamondsuit$

Three 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 $1 \leq p \leq q \leq \infty$, $L^p(X)$ 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, continuous functions on $\mathbb{R}$ are not always bounded. Furthermore, unbounded continuous functions on $\mathbb{R}$ are not integrable in any plausible sense. Similarly, continuous functions on an open or half-open bounded interval I are not necessarily bounded. Unbounded continuous functions on I are not Riemann integrable, and possibly not in $L^1(I)$. However, if $I = [a, b]$ is a closed, bounded interval, then the ladder in Figure 2.3, with $\mathbb{T}$ replaced by I, holds true. Third, 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 $1 \leq p \leq \infty$. Our ladder has a second branch:

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

Exercise 2.24. 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 $A \subset I$. ♢

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.

Example 2.25. 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 $\|f\|_{L^q(I)}^q \leq \|f\|_{L^\infty(I)}^{q-p} \|f\|_{L^p(I)}^p$, since $\|f\|_{L^\infty(I)}$ is the infimum of the numbers M that bound f. ♢

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

Lemma 2.26 (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*

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

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

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

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 2.27. 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 $h \in L^q(I)$. ◊

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*.

Definition 2.28. 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 2.30.

---

[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 independently twenty-five years earlier by the Ukranian mathematician Viktor Yakovlevich Bunyakovsky (1804–1889), and it is also known as the Cauchy–Bunyakovsky–Schwarz inequality.

[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).

Exercise 2.29. Show that finite subsets of $\mathbb{R}$ have measure zero. Show that countable subsets of $\mathbb{R}$ have measure zero. In particular show that the set $\mathbb{Q}$ of rational numbers has measure zero. $\Diamond$

Example 2.30 (*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]. $\Diamond$

Exercise 2.31. Show that the Cantor set C is closed. Use Definition 2.28 to show that C has measure zero. Alternatively, show that the measure of its complement is 1. Show that C is uncountable. $\Diamond$

Aside 2.32. We assume the reader is familiar with the 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 2.33 (Lebesgue's Theorem, 1901)**.** *A bounded function $f : I \to \mathbb{C}$ on a closed bounded interval $I = [a,b]$ 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.

[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, $m(X \setminus A) = m(X) - m(A)$.

Definition 2.34. 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 2.35. 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 2.36. 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 2.37. Verify that Exercise 2.27 holds true when $q = \infty$. $\Diamond$

Remark 2.38. 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 2.17 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 on page 42 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 2.39. 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 $n > N$,

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

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

Definition 2.40. 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 $n > N$,

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

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 2.41. 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 $n > N$,

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

Here is a slightly weaker version of uniform convergence.

Definition 2.42. 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 $L^\infty(X)$. $\lozenge$

Remark 2.43. 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 2.41 are continuous and converge uniformly, then the limiting function is itself continuous; see Theorem 2.59. $\lozenge$

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 2.44. 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 $L^1(I)$. $\lozenge$

Definition 2.45. 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 $L^2(I)$. ◊

Convergence in $L^1(I)$ and convergence in $L^2(I)$ are particular cases of *convergence in* $L^p(I)$, defined replacing $L^2(I)$ by $L^p(I)$ in Definition 2.45.

We have already noted that the L^p spaces are complete normed spaces, also known as Banach spaces. Therefore in Definitions 2.44 and 2.45 the limit functions must be in $L^1(I)$ and $L^2(I)$, respectively.

Exercise 2.46. Verify that uniform convergence on a bounded interval I implies the other modes of convergence on I. ◊

Exercise 2.47. Show that for $1 \le q < p \le \infty$ convergence in $L^p(I)$ implies convergence in $L^q(I)$. **Hint:** Use Hölder's Inequality (inequality (2.6)). ◊

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 2.48 (*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 $0 \le x < 1$ and $f(1) = 1$, but it does not converge uniformly to f, nor to any other function. See Figure 2.5. ◊

Example 2.49 (*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. ◊

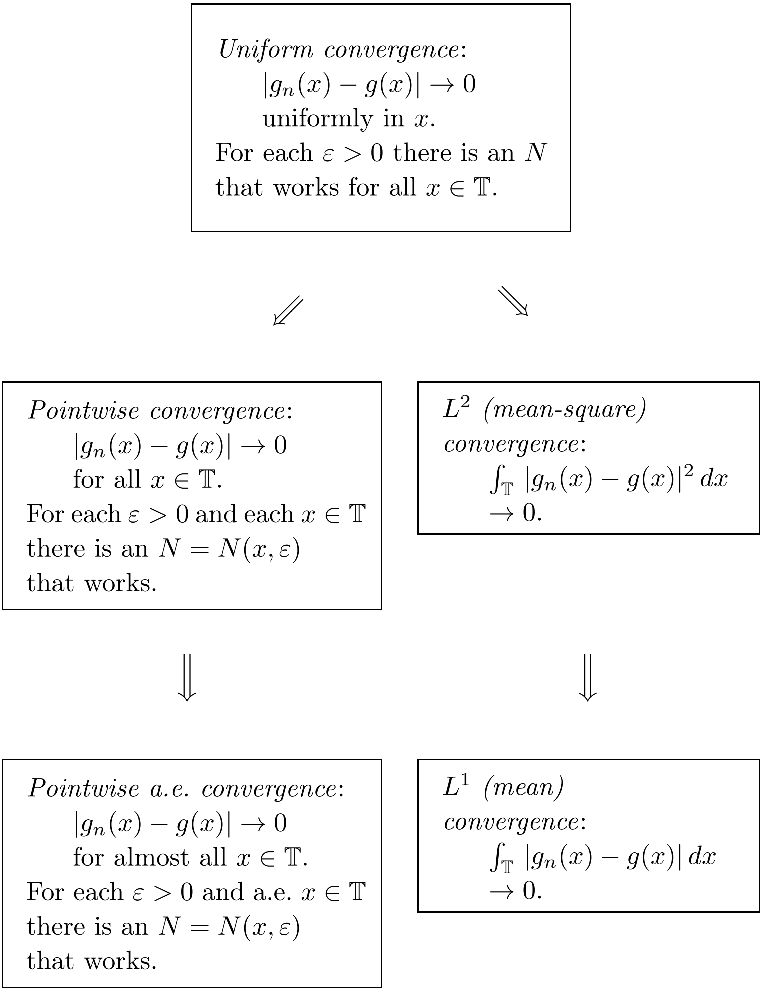


Figure 2.4. Relations between five of the six modes of convergence discussed in Section 2.2, for functions g_n, $g : \mathbb{T} \to \mathbb{C}$, showing which modes imply which other modes. All potential implications between these five modes other than those shown here are false.

Example 2.50 (*In the Mean But Not Pointwise*)**.** Given a positive integer n, let m, k be the unique nonnegative integers such that $n =$

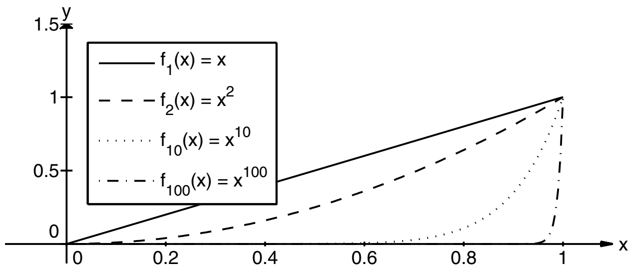


Figure 2.5. The functions $f_n(x) = x^n$ converge pointwise, but not uniformly, on $[0, 1]$. Their limit is the function $f(x)$ that takes the value 0 for $x \in [0, 1)$ and 1 at $x = 1$. Graphs of f_n are shown for $n = 1$, 2, 10, and 100.

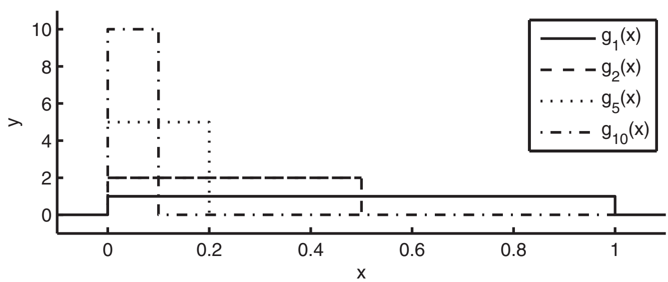


Figure 2.6. The functions g_n defined in Example 2.49 converge pointwise on $(0, 1]$, but not in the mean, in L^2, or uniformly. Their limit is the function $g(x)$ that takes the value 0 for $x \in (0, 1]$; g is undefined at $x = 0$. Graphs of g_n are shown for $n = 1$, 2, 5, and 10.

$2^k + m$ 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. $\Diamond$

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

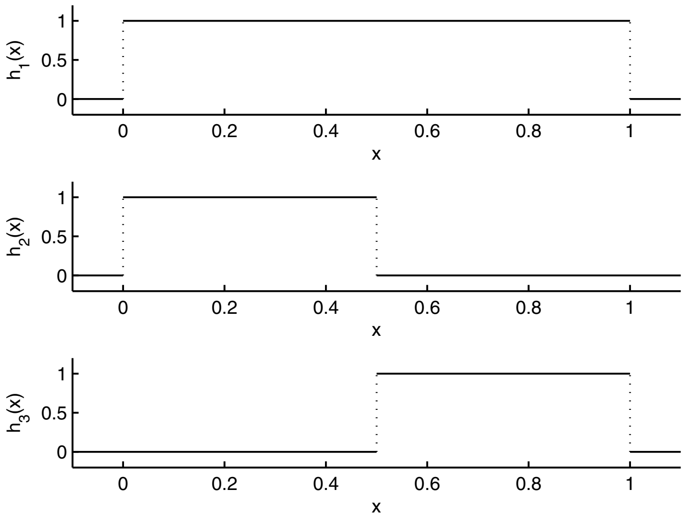


Figure 2.7. The first three walking functions of Example 2.50. The functions $h_n(x)$ converge in mean but do not converge pointwise at any x in $[0, 1)$.

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

Exercise 2.52. Where would uniform convergence outside a set of measure zero and L^p convergence for a given p with $1 \le p \le \infty$ 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. ♢

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 2.53 (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 2.49 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 2.54. 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 2.55 (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 2.56 (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,*

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

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

Exercise 2.57. Show by exhibiting counterexamples that each hypothesis in Theorem 2.56 is necessary. ◊

Exercise 2.58. 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 $\left(\sum_{n=0}^{\infty} g_n(x)\right)' = \sum_{n=0}^{\infty} (g_n)'(x)$? ◊

Here is another useful result in the same vein, already mentioned in Remark 2.43.

Theorem 2.59. *The uniform limit of continuous functions on a subset X of the real numbers is continuous on X.*

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

Theorem 2.59 ensures 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 2.48 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 2.59. It says that the pointwise limit of a decreasing sequence of functions is also their uniform limit.

Theorem 2.60 (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 2.61. Prove Dini's Theorem, and show that hypothesis (ii) is essential. A continuous version of Example 2.49, where the *steps* are replaced by *tents*, might help. ♢

In Weierstrass's M-Test, Theorem 2.55, 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.

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

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

Theorem 2.62 (Fubini's Theorem)**.** *Let I and J be closed intervals, and let $F : I \times J \to \mathbb{C}$ 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 2.62 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; sometimes it is invalid. We illustrate throughout the book settings where this interchange is allowed, such as the ones just described involving uniform convergence, and settings where it is illegal.

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

Theorem 2.63 (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 $|f_n(x)| \le g(x)$ a.e. for all $n > 0$. Then*

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

Exercise 2.64. Verify that the functions in Example 2.49 cannot be dominated by an integrable function g. ♢

Remark 2.65. The Lebesgue Dominated Convergence Theorem holds on $\mathbb{R}$ (see the Appendix). ♢

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 2.66. 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

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

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 2.12. We state this result as a theorem, for future reference.

Theorem 2.67. *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 2.68. We can choose the step functions that approximate a given Riemann integrable function to lie entirely below the function or entirely above it. $\lozenge$

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

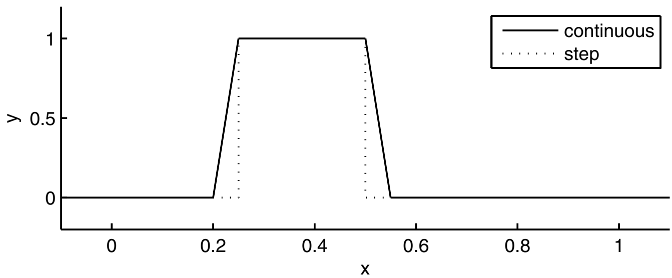


Figure 2.8. Graph of a continuous function that approximates a step function, both pointwise a.e. and in L^p norm.

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, as well as in the L^p spaces.

Theorem 2.69. *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 2.70. 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 2.69) 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. ◊

Exercise 2.71. Show that the function g_0 defined in Remark 2.70 satisfies the properties claimed there. ◊

Exercise 2.72. 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). ◊

Exercise 2.73 (*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). ◊

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 that 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 2.27), 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 2.74. Use Exercise 2.27 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 $\|f - g\|_{L^q(I)} < \delta$. $\Diamond$

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 2.75. *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 2.76. 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 $\|h(x) - f(x)\|_{L^p(I)} < \varepsilon$. $\Diamond$

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 Holschneider and Tchamitchian 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) Consider the following statement: 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[25] wavelets (see Chapter 9–11). Write an essay supporting this statement, choose one of these examples, and investigate its definition, history, and properties.

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

[25] Belgian-born American mathematician and physicist Ingrid Daubechies (born 1954).

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 reviewed 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 sawtooth 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 could we perhaps 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 $|n| \le N$. Instead of the full Fourier series, we obtain the N^{th} *partial Fourier sum* $S_N f$ *of* f, given by

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

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 $N \to \infty$ and 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 sawtooth function f in Example 1.12, the partial Fourier sums are

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

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 *square wave function* that arises as the periodic extension of the

[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 (see [**Gib1899**]), correcting his earlier letter in which he had overlooked the phenomenon (see [**Gib1898**]). Unbeknownst to Gibbs, the English mathematician Henry Wilbraham (1825–1883) had discovered the same phenomenon fifty years earlier (see [**Wil**]). 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%.*

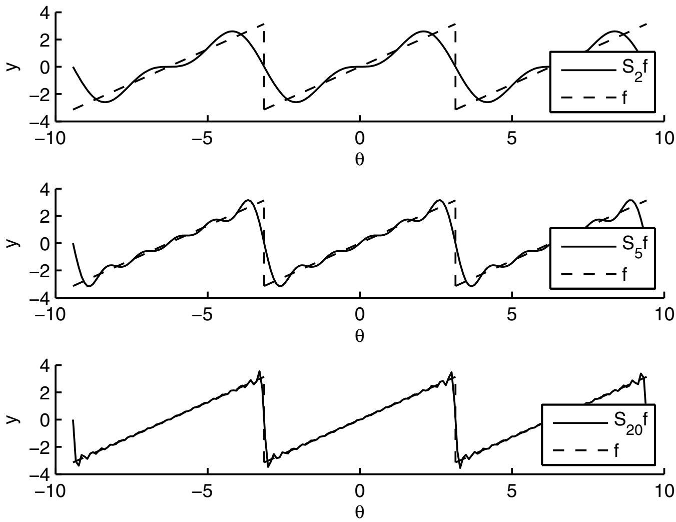


Figure 3.1. Graphs of the sawtooth function f and its partial Fourier sums $S_N f$, for $N = 2$, 5, and 20, extended periodically from $[-\pi, \pi)$ to $\mathbb{R}$. Compare with Figure 3.2.

following step function $g(\theta)$:

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

The project in Section 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). ♢

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 *Cesàro*[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 Cesàro 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 $f = 0$.*

---

[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 Cesàro (1859–1906).

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

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

The Cesàro 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).

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]. ◊

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,

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

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 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 2.55).

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 $n \neq 0$,

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

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 exists $C > 0$ 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 2.55) 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. ♢

In Section 4.4 we show that for continuous functions f on $\mathbb{T}$, the Cesàro 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 Cesàro 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 of partial sums $\{S_N f(\theta)\}_{N\in\mathbb{N}}$ for $f \in C^1(\mathbb{T})$. But once the convergence of the series is guaranteed, then the same result about convergence of Cesàro 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 $n \neq 0$,

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

(Check!) 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 have just 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

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

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 a 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. ♢

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. □

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 2.56 may be useful.

Now prove Theorem 3.16 by induction. ◊

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

> *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

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

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

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

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,

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

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 2.53), obtaining

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

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 $g_n(\theta) := |f(\theta) - f\left(\theta + \pi/n\right)|$ converge uniformly to zero. ♢

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

Exercise 3.22. (*Decay of Fourier Coefficients of Lipschitz Functions*). A function $f : \mathbb{T} \to \mathbb{C}$ is called *Lipschitz* if there is a constant k such that $|f(\theta_1) - f(\theta_2)| \leq k|\theta_1 - \theta_2|$ for all $\theta_1, \theta_2 \in \mathbb{T}$. Show that if f is Lipschitz, then

$$|\widehat{f}(n)| \leq \frac{C}{n}$$

for some constant C independent of $n \in \mathbb{Z}$. ♢

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 3.23. *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**, Theorem 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 3.24 for $k = 2$. It is not difficult to adapt that example to the case $k > 2$.

Example 3.24 (*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:

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

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

Exercise 3.25. Compute the Fourier coefficients of the plucked string function in Example 3.24. 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

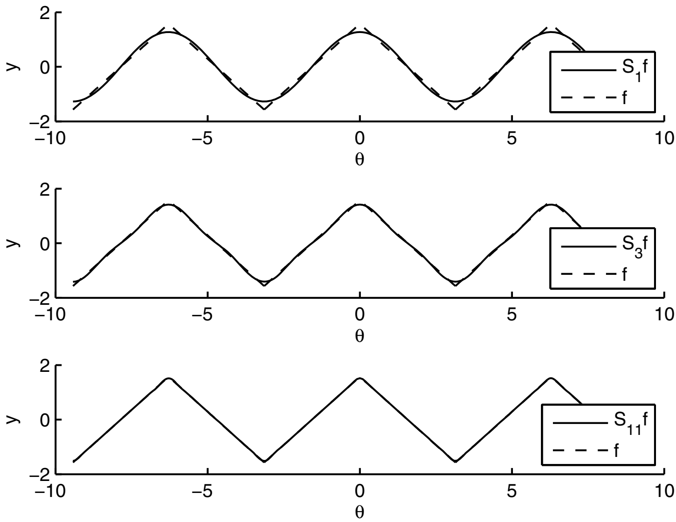


Figure 3.2. Graphs of the plucked string function $f(\theta) = \pi/2 - |\theta|$ from Example 3.24 and its partial Fourier sums $S_N f$ for $N = 1$, 3, and 9, extended periodically from $(-\pi, \pi)$ to $\mathbb{R}$. Compare the convergence of $S_N f$ with that shown for the ramp function in Figure 3.1.

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]. ◊

Exercise 3.26. 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 3.27 (Dirichlet, 1829). *Suppose that $f : \mathbb{T} \to \mathbb{C}$ is continuous except possibly at finitely many points at which the one-sided limits $f(\theta^+)$ and $f(\theta^-)$ exist, and that its derivative f' is continuous and bounded except at such 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 is shattered by the following result.

Theorem 3.28 (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 the project in Section 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 3.29 (Jordan, 1881). *If f is a function of* bounded variation[8] *on $\mathbb{T}$, then the Fourier series of f converges to $[f(\theta^+)+f(\theta^-)]/2$*

[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.

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 3.27 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 3.30 (Dini, 1878). *Let* f *be a* 2π*-periodic function. Suppose there is* $\delta > 0$ *such that* $\int_{|t|<\delta} \big|f(\theta+t) - f(\theta)\big|/|t|\,dt < \infty$ *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 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 and 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 published only after his death in 1867. See [**KL**, Chapter 5, Sections 1 and 4].

Theorem 3.31 (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* $f(\theta)$ *depends only 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 the following: *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 3.32 (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 3.33 (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 3.34 (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 3.34 can be found in Katznelson's book *An introduction to harmonic analysis* [**Kat**, Chapter II, Section 3]. The same chapter includes a construction of a Kolmogorov-type example.

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

Theorem 3.35 (Hunt[11])**.** *Carleson's Theorem holds for all $f \in L^p(\mathbb{T})$ and for all $p > 1$.*

It took almost forty 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

[9] Jean-Pierre Kahane, French mathematician (born 1926).
[10] Yitzhak Katznelson, Israeli mathematician (born 1934).
[11] Richard Allen Hunt, American mathematician (1937–2009).

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.

(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 and 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's 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 Gibbs's phenomenon? This question will make more sense once you get acquainted with Chapters 9–11.

3.5. Project: Monsters, Take II

As in the project in Section 2.5, we explore several unusual functions.

(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 3.28.) Write an explanation of the construction. Your intended audience is another student at the same stage of 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) Repeat 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].

Chapter 4

Summability methods

Pointwise convergence of partial Fourier sums for continuous functions was ruled out by the du Bois-Reymond example. However, in an attempt to obtain a convergence result for all continuous functions, mathematicians devised various averaging or *summability methods.* These methods require only knowledge of the Fourier coefficients in order to recover a continuous function as the sum of an appropriate trigonometric series. Along the way, a number of very important approximation techniques in analysis were developed, in particular *convolutions* (Section 4.2) and *approximations of the identity*, also known as *good kernels* (Section 4.3). We discuss these techniques in detail in the context of 2π-periodic functions. Convolution of a continuous function with an approximation of the identity approximates uniformly the function and also in the L^p norm (Section 4.6).

We describe several kernels that arise naturally in the theory of Fourier series. The Fejér and Poisson[1] kernels (Sections 4.4 and 4.5) are good kernels that generate approximations of the identity; the Dirichlet kernel[2] (Section 4.1) is equally important but not good in this sense. As we will see, the fact that the Dirichlet kernel is not a good kernel accounts for the difficulties in achieving pointwise convergence for continuous functions.

---

[1] Named after the French mathematician Siméon-Denis Poisson (1781–1840).
[2] Named after the same Dirichlet as is the function in Example 2.17.

4.1. Partial Fourier sums and the Dirichlet kernel

We compute an integral representation for $S_N f$, the N^{th} partial Fourier sum of a suitable function f. Note that in this calculation we may interchange the summation and the integral, since there are only finitely many terms in the sum. The notation D_N and $*$ will be explained below. Recall that the N^{th} partial Fourier sum of an integrable function $f : \mathbb{T} \to \mathbb{C}$ is given by

$$S_N f(\theta) = \sum_{|n|\le N} \widehat{f}(n) e^{in\theta}, \quad \widehat{f}(n) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) e^{-iny} dy.$$

Inserting the formula for $\widehat{f}(n)$ into the right-hand side and using the linearity of the integral and the properties of the exponential, we obtain a formula for $S_N f$ as an integral involving $f(y)$, and we are led to define the 2π-periodic function $D_N : \mathbb{T} \to \mathbb{C}$,

$$\begin{aligned} S_N f(\theta) &= \sum_{|n|\le N} \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) e^{-iny} dy \right) e^{in\theta} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) \sum_{|n|\le N} e^{in(\theta - y)}\, dy \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) D_N(\theta - y)\, dy. \end{aligned}$$

In the last line of our calculation we used the idea of convolving two functions. If f, $g : \mathbb{T} \to \mathbb{C}$ are integrable, their *convolution* is the new integrable[3] function $f * g : \mathbb{T} \to \mathbb{C}$ given by

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

With this notation we have shown that given an integrable function $f : \mathbb{T} \to \mathbb{C}$, then for all $N \in \mathbb{N}$ and $\theta \in \mathbb{T}$,

$$S_N f(\theta) = (D_N * f)(\theta). \tag{4.1}$$

[3]Riemann integrable functions on bounded intervals are closed under translations, reflections, and products. We will see in Section 4.2 that $\mathcal{R}(\mathbb{T})$ is closed under convolutions. On the other hand, the set of Lebesgue integrable functions on $\mathbb{T}$ is not closed under products; however, it is closed under convolutions.

Definition 4.1. The *Dirichlet kernel* D_N is defined by

$$D_N(\theta) := \sum_{|n|\leq N} e^{in\theta}. \qquad \Diamond$$

Notice that D_N does not depend on f. Also, D_N is a trigonometric polynomial of degree N, with coefficients equal to 1 for $-N \leq n \leq N$ and 0 for all other values of n.

The Dirichlet kernel and convolution are essential in what follows. We discuss their properties, beginning with the Dirichlet kernel.

Aside 4.2. In this chapter we use the term *integrable* without specifying whether we mean Riemann integrable or Lebesgue integrable. Most results do hold for both types of integrability. Some results we have stated and proved specifically for integrable and bounded functions. If the functions are Riemann integrable, then they are bounded by definition, but this is not true for Lebesgue integrable functions. Nevertheless the results hold for Lebesgue integrable functions, but with different proofs than the ones presented. $\Diamond$

4.1.1. Formulas for the Dirichlet kernel. We begin with a formula for the Dirichlet kernel in terms of cosines. Using the formula $2\cos x = e^{ix} + e^{-ix}$, we see that

$$\begin{aligned} D_N(\theta) &= e^{-iN\theta} + \cdots + e^{-i\theta} + e^0 + e^{i\theta} + \cdots + e^{iN\theta} \\ &= 1 + 2\sum_{1\leq n\leq N} \cos(n\theta). \end{aligned} \tag{4.2}$$

In particular, the Dirichlet kernel D_N is real-valued and even.

Next, we obtain a formula for D_N as a quotient of two sines:

$$D_N(\theta) = \sum_{|n|\leq N} e^{in\theta} = e^{-iN\theta} \sum_{0\leq n\leq 2N} (e^{i\theta})^n.$$

Evaluating this partial geometric sum, we find that for $\theta \neq 0$,

$$D_N(\theta) = e^{-iN\theta}\,\frac{(e^{i\theta})^{2N+1} - 1}{e^{i\theta} - 1} = \frac{e^{i(2N+1)\theta/2} - e^{-i(2N+1)\theta/2}}{e^{i\theta/2} - e^{-i\theta/2}}.$$

Here we have used the typical analysis trick of multiplying and dividing by the same quantity, in this case $e^{-i\theta/2}$. Now we use the formula

$2i\sin\theta = e^{i\theta} - e^{-i\theta}$ to conclude that

$$D_N(\theta) = \sin\big[(2N+1)\theta/2\big]/\sin(\theta/2)\,. \tag{4.3}$$

The quotient of sines has a removable singularity at $\theta = 0$. We fill in the appropriate value $D_N(0) = 2N+1$ and use it without further comment. The kernel D_N has no other singularities on $[-\pi, \pi]$, and in fact D_N is C^∞ on the unit circle $\mathbb{T}$.

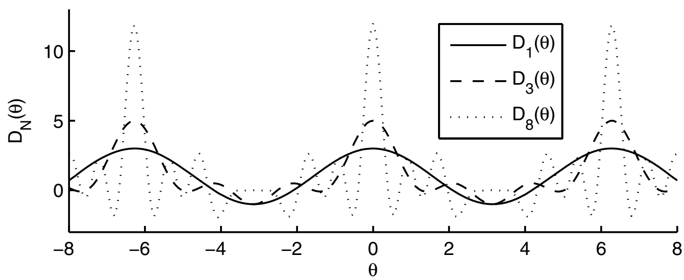


Figure 4.1. Graphs of Dirichlet kernels $D_N(\theta)$ for $N = 1$, 3, and 8, extended periodically from $[-\pi, \pi)$ to $\mathbb{R}$.

In Figure 4.1 we see that as $\sin[(2N+1)\theta/2]$ varies between 1 and -1, the oscillations of D_N are bounded above and below by $\pm 1/\sin(\theta/2)$. Also, D_N takes on both positive and negative values, D_N is even, and D_N achieves its maximum value of $2N+1$ at $\theta = 0$.

4.1.2. Properties of the Dirichlet kernel. We list the basic properties of the Dirichlet kernel.

Theorem 4.3. *The Dirichlet kernel has the following properties.*

(i) $(f * D_N)(\theta) = S_N f(\theta)$, *for* $\theta \in \mathbb{T}$ *and integrable* $f : \mathbb{T} \to \mathbb{C}$.

(ii) D_N *is an even function of* θ.

(iii) D_N *has mean value* 1 *for all* N: $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(\theta)\,d\theta = 1$.

(iv) *However, the integral of the absolute value of* D_N *depends on* N, *and in fact* $\frac{1}{2\pi}\int_{-\pi}^{\pi} |D_N(\theta)|\,d\theta \approx \log N$.

We are using the notation $A_N \approx B_N$ (to be read A_N *is comparable to* B_N) if there exist constants c, $C > 0$ such that $cB_N \le A_N \le CB_N$ for all N.

It may seem surprising that the averages of D_N can be uniformly bounded while the averages of $|D_N|$ grow like $\log N$; see Figure 4.2. This growth is possible because of the increasing oscillation of D_N with N, which allows the total area between the graph of D_N and the θ-axis to grow without bound as $N \to \infty$, while cancellation of the positive and negative parts of this area keeps the integral of D_N constant.

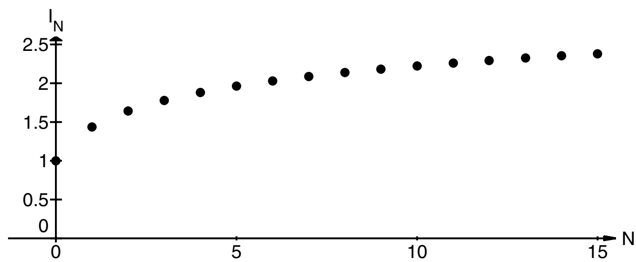


Figure 4.2. The mean value $I_N = 1/(2\pi)\int_{-\pi}^{\pi} |D_N(\theta)|\,d\theta$, over the interval $[-\pi, \pi)$, of the absolute value of the Dirichlet kernel D_N, for $0 \leq N \leq 15$. Note the logarithmic growth.

Exercise 4.4. Show that $D_N(\theta)$ does not converge to zero at any point. $\diamondsuit$

Proof of Theorem 4.3. We have already established properties (i) and (ii). For property (iii), note that

$$\int_{-\pi}^{\pi} D_N(\theta)\,d\theta = \int_{-\pi}^{\pi} \Big[\sum_{|n|\leq N} e^{in\theta}\Big]\,d\theta = \sum_{|n|\leq N} \Big[\int_{-\pi}^{\pi} e^{in\theta}\,d\theta\Big] = 2\pi,$$

since the integral of $e^{in\theta}$ over $[-\pi, \pi)$ is zero for $n \neq 0$, and 2π for $n = 0$. Verifying property (iv) is left to the reader; see Exercise 4.5. See also Figure 4.2 for numerical evidence. $\square$

Exercise 4.5. Justify Theorem 4.3(iv) analytically. It may be useful to remember that $\sum_{n=1}^{N}(1/n) \approx \int_1^N (1/x)\,dx = \log N$. $\diamondsuit$

4.2. Convolution

Convolution can be thought of as a way to create new functions from old ones, alongside the familiar methods of addition, multiplication, and composition of functions. We recall the definition.

Definition 4.6. Given two integrable, 2π-periodic functions f, $g : \mathbb{T} \to \mathbb{C}$, their *convolution on* $\mathbb{T}$ is the new function $f * g : \mathbb{T} \to \mathbb{C}$ given by

$$(f * g)(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) g(\theta - y)\, dy. \qquad (4.4) \qquad \diamond$$

When convolving two functions, we integrate over the dummy variable y, leaving a function of θ. The point $\theta - y$ will not always lie in $[-\pi, \pi)$, but since the integrand is 2π-periodic ($g(\theta) = g(\theta + 2\pi k)$ for all integers k), we still know the values of the integrand everywhere on $\mathbb{R}$. Pictorially, we compute $f * g$ at a point θ by reflecting the graph of $g(y)$ in the vertical axis and translating it to the left by θ units to obtain $g(\theta - y)$, multiplying by $f(y)$, then finding the area under the graph of the resulting product $f(y)g(\theta - y)$ over $y \in [-\pi, \pi)$ and dividing by the normalizing constant 2π.

Notice that if the periodic function g is integrable on $\mathbb{T}$, then so is the new periodic function $h(y) := g(\theta - y)$. Also, the product of two Riemann integrable periodic functions on $\mathbb{T}$ is also periodic and integrable (Section 2.1.1). This last statement is not always true for Lebesgue integrable functions. Nevertheless the convolution of two functions f, $g \in L^1(\mathbb{T})$ is always a function in $L^1(\mathbb{T})$, and the following inequality[4] holds (see Exercise 4.13):

$$\|f * g\|_{L^1(\mathbb{T})} \leq \|f\|_{L^1(\mathbb{T})} \|g\|_{L^1(\mathbb{T})}. \qquad (4.5)$$

The convolution of two integrable functions is always integrable, and so we can calculate its Fourier coefficients.

Example 4.7 (*Convolving with a Characteristic Function*). We compute a convolution in the special case where one of the convolution factors is the characteristic function $\chi_{[a,b]}(\theta)$ of a closed interval $[a, b]$ in $\mathbb{T}$. Recall that the characteristic function is defined by $\chi_{[a,b]}(\theta) = 1$

[4]This is an instance of Young's inequality; see Chapters 7 and 12.

if $\theta \in [a, b]$ and $\chi_{[a,b]}(\theta) = 0$ otherwise. Using the observation that $a \leq \theta - y \leq b$ if and only if $\theta - b \leq y \leq \theta - a$, we see that

$$\begin{aligned}(f * \chi_{[a,b]})(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)\chi_{[a,b]}(\theta - y)\, dy \\ &= \frac{1}{2\pi} \int_{\theta-b}^{\theta-a} f(y)\, dy = \frac{b-a}{2\pi} \frac{1}{b-a} \int_{\theta-b}^{\theta-a} f(y)\, dy.\end{aligned}$$

So the convolution of f with $\chi_{[a,b]}$, evaluated at θ, turns out to be $(b - a)/(2\pi)$ times the average value of f on the reflected and translated interval $[\theta - b, \theta - a]$. ♢

Exercise 4.8 (*Convolving Again and Again*)**.** Let $f = \chi_{[0,1]}$. Compute $f * f$ and $f * f * f$ by hand. Use MATLAB to compute and plot the functions f, $f * f$, and $f * f * f$ over a suitable interval. Notice how the smoothness improves as we take more convolutions. ♢

This exercise gives an example of the so-called *smoothing properties* of convolution. See also Section 4.2.2.

4.2.1. Properties of convolution. We summarize the main properties of convolution.

Theorem 4.9. *Let f, g, $h : \mathbb{T} \to \mathbb{C}$ be 2π-periodic integrable functions, and let $c \in \mathbb{C}$ be a constant. Then*

(i) $f * g = g * f$ (*commutative*);

(ii) $f * (g + h) = (f * g) + (f * h)$ (*distributive*);

(iii) $(cf) * g = c(f * g) = f * (cg)$ (*homogeneous*);

(iv) $(f * g) * h = f * (g * h)$ (*associative*);

(v) $\widehat{f * g}\,(n) = \widehat{f}\,(n)\,\widehat{g}\,(n)$ (*Fourier process converts convolution to multiplication*).

The first four properties are obtained by manipulating the integrals (change of variables, interchanging integrals, etc.). The last property requires some work. The idea is first to prove these properties for *continuous* functions and then to extend to *integrable* functions by approximating them with continuous functions. This is an example of the usefulness of approximating functions by nicer functions[5]. We will use the following approximation lemmas.

Lemma 4.10. *Suppose* $f : \mathbb{T} \to \mathbb{C}$ *is an integrable and bounded function. Let* $B > 0$ *be such that* $|f(\theta)| \leq B$ *for all* $\theta \in \mathbb{T}$*. Then there is a sequence* $\{f_k\}_{k=1}^{\infty}$ *of continuous functions* $f_k : \mathbb{T} \to \mathbb{C}$ *such that*

(i) $\sup_{\theta\in\mathbb{T}} |f_k(\theta)| \leq B$ *for all* $k \in \mathbb{N}$ *and*

(ii) $\int_{-\pi}^{\pi} |f(\theta) - f_k(\theta)|\, d\theta \to 0$ *as* $k \to \infty$.

Part (i) of the lemma says that the f_k are also bounded by B, and part (ii) says, in language we have already met in Chapter 2, that the f_k converge to f in $L^1(\mathbb{T})$. If $f \in \mathcal{R}(\mathbb{T})$, then it is bounded, and the lemma says that *the continuous functions on* $\mathbb{T}$ *are dense in the set of Riemann integrable functions on* $\mathbb{T}$. See Theorem 2.69, and in particular Remark 2.70.

Lemma 4.11. *Given* $f, g : \mathbb{T} \to \mathbb{C}$ *bounded and integrable functions (bounded by* B*), let* $\{f_k, g_k\}_{k=1}^{\infty}$ *be sequences of continuous functions (bounded by* B*) approximating* f *and* g*, respectively, in the* L^1 *norm, as in Lemma* 4.10*. Then* $f_k * g_k$ *converges uniformly to* $f * g$.

These two lemmas work for functions that are assumed to be Riemann integrable (hence bounded) or for bounded Lebesgue integrable functions. We will prove Lemma 4.11 after we prove Theorem 4.9.

Proof of Theorem 4.9. Property (i), rewritten slightly, says that

$$(4.6)\quad (f*g)(\theta) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)g(\theta-y)\,dy = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta-y)g(y)\,dy,$$

so that "the $\theta - y$ can go with g *or* with f". Equation (4.6) follows from the change of variables $y' = \theta - y$ and the 2π-periodicity of the integrand.

Properties (ii) and (iii) show that convolution is linear in the second variable. By the commutativity property (i), it follows that convolution is linear in the first variable as well. Hence convolution

[5] Approximation by nice or simpler functions, such as continuous functions, trigonometric polynomials, or step functions, is a recurrent theme in analysis and in this book. We encountered this theme in Chapter 2 and we will encounter it again in the context of mean-square convergence of Fourier series, Fourier integrals, Weierstrass's Theorem, and orthogonal bases, in particular the Haar and wavelet bases.

is a bilinear operation. Properties (ii) and (iii) are consequences of the linearity of the integral. Property (iv), associativity, is left as an exercise for the reader.

We prove property (v), that the n^{th} Fourier coefficient of the convolution $f*g$ is the product of the n^{th} coefficients of f and of g. For *continuous* f and g, we may interchange the order of integration by Fubini's Theorem (Theorem 2.62) for continuous functions, obtaining

$$\begin{aligned}
\widehat{f*g}\,(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}(f*g)(\theta)\,e^{-in\theta}\,d\theta \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)g(\theta-y)\,dy\right]e^{-in\theta}\,d\theta \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} g(\theta-y)e^{-in\theta}\,d\theta\right]dy.
\end{aligned}$$

We now multiply the integrand by $1=e^{-iny}e^{iny}$, distribute it appropriately, and change variables to make the Fourier transform of g in the integral with respect to θ appear:

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

We have established property (v) for continuous functions. Now suppose f and g are integrable and bounded. Take sequences $\{f_k\}_{k\in\mathbb{N}}$ and $\{g_k\}_{k\in\mathbb{N}}$ of continuous functions such that $f_k\to f$, $g_k\to g$, as in the Approximation Lemma (Lemma 4.10).

First, $\widehat{f_k}\,(n)\to\widehat{f}\,(n)$ for each $n\in\mathbb{Z}$, since by Lemma 4.10(ii)

$$|\widehat{f_k}\,(n)-\widehat{f}\,(n)|\le\frac{1}{2\pi}\int_{-\pi}^{\pi}|f_k(\theta)-f(\theta)|\,d\theta\to 0$$

as $k\to\infty$. Similarly $\widehat{g_k}\,(n)\to\widehat{g}\,(n)$ for each $n\in\mathbb{Z}$.

Second, by property (v) for continuous functions, for each $n\in\mathbb{Z}$ and $k\in\mathbb{N}$,

$$\widehat{f_k*g_k}\,(n)=\widehat{f_k}\,(n)\,\widehat{g_k}\,(n).$$

Third, we show that $\widehat{f_k * g_k}(n) \to \widehat{f * g}(n)$ as $k \to \infty$ for each $n \in \mathbb{Z}$. In fact,

$$\begin{aligned} \lim_{k\to\infty} \widehat{f_k * g_k}(n) &= \lim_{k\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} (f_k * g_k)(\theta)\, e^{-in\theta}\, d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \lim_{k\to\infty} (f_k * g_k)(\theta)\, e^{-in\theta}\, d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \big[(f * g)(\theta)\, e^{-in\theta}\big]\, d\theta \;=\; \widehat{f * g}(n). \end{aligned}$$

Here we can move the limit inside the integral since, by Lemma 4.11, the integrands converge uniformly on $\mathbb{T}$ (see Theorem 2.53). Hence

$$\begin{aligned} \widehat{f * g}(n) &= \lim_{k\to\infty} \widehat{f_k * g_k}(n) = \lim_{k\to\infty} \widehat{f_k}(n)\, \widehat{g_k}(n) \\ &= \big(\lim_{k\to\infty} \widehat{f_k}(n)\big)\big(\lim_{k\to\infty} \widehat{g_k}(n)\big) = \widehat{f}(n)\,\widehat{g}(n). \end{aligned}$$

For Lebesgue integrable functions, the L^1 version of Fubini's Theorem is needed to carry out the calculation we did for continuous functions. Then we do not need to use the approximation argument explicitly, because it is used implicitly in the proof of Fubini's Theorem for L^1 functions. □

Exercise 4.12 (*Convolution Is Associative*)**.** Prove property (iv) of Theorem 4.9. ♢

Proof of Lemma 4.11. We show that $f_k * g_k \to f * g$ uniformly on $\mathbb{T}$. We can estimate their pointwise difference. First, by the Triangle Inequality for integrals,

$$|(f_k * g_k)(\theta) - (f * g)(\theta)| \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |f_k(y)g_k(\theta - y) - f(y)g(\theta - y)|\, dy.$$

Second, adding and subtracting $f_k(y)g(\theta - y)$ to the integrand,

$$\begin{aligned} |(f_k * g_k)(\theta) - (f * g)(\theta)| \le{}& \frac{1}{2\pi}\int_{-\pi}^{\pi} |f_k(y)|\,|g_k(\theta - y) - g(\theta - y)|\, dy \\ &+ \frac{1}{2\pi}\int_{-\pi}^{\pi} |f_k(y) - f(y)|\,|g(\theta - y)|\, dy. \end{aligned}$$

The inequality is a consequence of the *Triangle Inequality*[6] for complex numbers and the *additivity*[7] of the integral.

Third, recall that f_k, g_k and f, g are all periodic functions bounded by $B > 0$ on $\mathbb{T}$. We conclude that the difference $|(f_k * g_k)(\theta) - (f * g)(\theta)|$ is bounded by $\frac{B}{2\pi}\big(\int_{-\pi}^{\pi} |g_k(\theta-y)-g(\theta-y)|\,dy+\int_{-\pi}^{\pi} |f_k(y)-f(y)|\,dy\big)$. This quantity is independent of θ because of the periodicity of g and g_k, more precisely, $\int_{-\pi}^{\pi} |g_k(\theta - y) - g(\theta - y)|\,dy = \int_{-\pi}^{\pi} |g_k(y) - g(y)|\,dy$. Finally, by hypothesis, $\int_{-\pi}^{\pi} |f_k - f| \to 0$ and $\int_{-\pi}^{\pi} |g_k - g| \to 0$ as $k \to \infty$ by Lemma 4.10. Hence, $|(f_k * g_k)(\theta) - (f * g)(\theta)| \to 0$ as $k \to \infty$, and the convergence is uniform. □

Exercise 4.13. Use property (v) of Theorem 4.9, with $n = 0$, to prove inequality (4.5). ◊

4.2.2. Convolution is a smoothing operation. Convolution is a so-called *smoothing operation*. We will see that the smoothness of the convolution is the combined smoothness of the convolved functions (see Exercise 4.16). (This is how in theory you would like marriages to be: the relationship combines the smoothness of each partner to produce an even smoother couple.)

The following result is a precursor of that principle. We start with two functions, at least one of which is continuous; then their convolution preserves continuity. It is not apparent that there has been any improvement in the smoothness of the convolution compared to the smoothness of the convolved functions. However an approximation argument allows us to start with a pair of integrable and bounded functions that are not necessarily continuous, and their convolution will be continuous.

Lemma 4.14. *If f, $g \in C(\mathbb{T})$, then $f*g \in C(\mathbb{T})$. In fact, if $f \in C(\mathbb{T})$ and g is integrable, then $f * g \in C(\mathbb{T})$.*

Proof. Let $f \in C(\mathbb{T})$, and let g be integrable on $\mathbb{T}$. First, f is uniformly continuous (as it is continuous on a closed and bounded interval). That is, given $\varepsilon > 0$, there exists $\delta > 0$ such that if $|h| < \delta$,

[6]Namely, for a, $b \in \mathbb{C}$, $|a + b| \le |a| + |b|$.
[7]Additivity means that if f, $g \in L^1(\mathbb{T})$, then $\int_{-\pi}^{\pi}(f + g) = \int_{-\pi}^{\pi} f + \int_{-\pi}^{\pi} g$.

then $|f(\alpha+h)-f(\alpha)|<\varepsilon$ for all $\alpha\in\mathbb{T}$ and by periodicity for all $\alpha\in\mathbb{R}$.

Second, by the linearity of the integral,

$$(f*g)(\theta+h)-(f*g)(\theta)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\big(f(\theta+h-y)-f(\theta-y)\big)g(y)\,dy.$$

Third, using the Triangle Inequality for integrals[8] and the uniform continuity of f, we conclude that $|(f*g)(\theta+h)-(f*g)(\theta)|$ is bounded above for all $|h|<\delta$ by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\big|f(\theta+h-y)-f(\theta-y)\big||g(y)|\,dy\leq\frac{\varepsilon}{2\pi}\int_{-\pi}^{\pi}|g(y)|=\frac{\varepsilon}{2\pi}\|g\|_{L^1(\mathbb{T})}.$$

Therefore the convolution of a continuous function and an integrable function is continuous. In particular the convolution of two continuous functions is continuous. □

Aside 4.15. Let f and g be integrable and bounded functions, and let f_k and g_k be continuous and bounded functions approximating f and g as in Lemma 4.10. Then f_k*g_k is continuous (by Lemma 4.14), and $f*g$ is the uniform limit of continuous functions (by Lemma 4.11). We conclude by Theorem 2.59 that $f*g$ is continuous. ♢

We have shown (see Aside 4.15) that *the convolution of two integrable and bounded functions on* $\mathbb{T}$ *is continuous.* This is an example of how *convolution improves smoothness.* Even more is true: if the convolved functions are smooth, then the convolution absorbs the smoothness from each of them, as the following exercise illustrates.

Exercise 4.16 (*Convolution Improves Smoothness*)**.** Suppose f, $g:\mathbb{T}\to\mathbb{C}$, $f\in C^k(\mathbb{T})$, $g\in C^m(\mathbb{T})$. Show that $f*g\in C^{k+m}(\mathbb{T})$. Furthermore the following formula holds: $(f*g)^{(k+m)}=f^{(k)}*g^{(m)}$. **Hint:** Check first for $k=1$, $m=0$; then by induction on k check for all $k\geq 0$ and $m=0$. Now use commutativity of the convolution to verify the statement for $k=0$ and any $m\geq 0$. Finally put all these facts together to get the desired conclusion. (It suffices to assume that the functions $f^{(k)}$ and $g^{(m)}$ are bounded and integrable to conclude that $f*g\in C^{k+m}(\mathbb{T})$.) ♢

[8]That is, the fact that $|\int_{\mathbb{T}} f|\leq\int_{\mathbb{T}}|f|$.

Another instance of how convolution keeps the best features from each function is the content of the following exercise.

Exercise 4.17 (*Convolution with a Trigonometric Polynomial Yields a Trigonometric Polynomial*)**.** Show that if f is integrable and $P \in \mathcal{P}_N$, then $f * P \in \mathcal{P}_N$. $\Diamond$

Property (v) in Theorem 4.9 is another instance of the time–frequency dictionary. Earlier, with equations (3.3) and (3.4), we showed that *differentiation is transformed into polynomial multiplication.* Here we show that *convolution is transformed into the ordinary product.* In the next exercise we have yet another instance of this interplay: *translations are transformed into modulations.*

Exercise 4.18 (*Translation Corresponds to Modulation*)**.** Prove that if f is 2π-periodic and integrable on $\mathbb{T}$ and the *translation operator* τ_h is defined by $\tau_h f(\theta) := f(\theta - h)$ for $h \in \mathbb{R}$, then $\tau_h f$ is also 2π-periodic and integrable on $\mathbb{T}$. Moreover, $\widehat{\tau_h f}(n) = e^{-inh}\widehat{f}(n)$. So translation of f by the amount h has the effect of multiplying the n^{th} Fourier coefficient of f by e^{inh}, for $n \in \mathbb{Z}$. $\Diamond$

We summarize the *time–frequency dictionary* for Fourier coefficients of periodic functions in Table 4.1. Note that f is treated as a 2π-periodic function on $\mathbb{R}$.

Table 4.1. The time–frequency dictionary for Fourier series.

Time/Space $\theta \in \mathbb{T}$	**Frequency** $n \in \mathbb{Z}$
derivative $f'(\theta)$	polynomial $\widehat{f'}(n) = in\widehat{f}(n)$
circular convolution $(f * g)(\theta)$	product $\widehat{f * g}(n) = \widehat{f}(n)\widehat{g}(n)$
translation/shift $\tau_h f(\theta) = f(\theta - h)$	modulation $\widehat{\tau_h f}(n) = e^{-ihn}\widehat{f}(n)$

We will study other versions of convolution. For convolution of vectors in $\mathbb{C}^n$, see Chapter 6, and for convolution of integrable functions defined on the whole real line and not necessarily periodic, see Section 7.5. In these contexts there is a Fourier theory and the same phenomenon occurs: the Fourier transform converts convolution to multiplication.

4.3. Good kernels, or approximations of the identity

Roughly speaking, a family of good kernels is a sequence of functions whose mass is concentrating near the origin (very much like a delta function[9]). A precise definition is given below.

Definition 4.19. A family $\{K_n\}_{n=1}^{\infty}$ of real-valued integrable functions on the circle, $K_n : \mathbb{T} \to \mathbb{R}$, is a *family of good kernels* if it satisfies these three properties:

(a) For all $n \in \mathbb{N}$, $\frac{1}{2\pi} \int_{-\pi}^{\pi} K_n(\theta)\, d\theta = 1$.

(b) There exists $M > 0$ so that $\int_{-\pi}^{\pi} |K_n(\theta)|\, d\theta \leq M$ for all $n \in \mathbb{N}$.

(c) For each $\delta > 0$, $\int_{\delta \leq |\theta| \leq \pi} |K_n(\theta)|\, d\theta \to 0$ as $n \to \infty$. $\diamondsuit$

For brevity, we may simply say that K_n is a good kernel, without explicitly mentioning the family $\{K_n\}_{n=1}^{\infty}$ of kernels.

A family of good kernels is often called an *approximation of the identity*. After Theorem 4.23 below, we explain why.

Condition (a) says that the K_n all have mean value 1. Condition (b) says that the integrals of the absolute value $|K_n|$ are uniformly bounded. Condition (c) says that for each fixed positive δ, the total (unsigned) area between the graph of K_n and the θ-axis, more than distance δ from the origin, tends to zero. A less precise way to say this is that as $n \to \infty$, most of the *mass* of K_n concentrates as near to $\theta = 0$ as we please.

Aside 4.20. Some authors use a more restrictive definition of approximations of the identity, replacing conditions (b) and (c) by

[9]The so-called delta function is not a function at all, but rather an example of the kind of generalized function known as a distribution. We define it in Section 8.4.

(b′) K_n is positive: $K_n(\theta) \geq 0$ for all $\theta \in \mathbb{T}$, for all $n \in \mathbb{N}$.

(c′) For each $\delta > 0$, $K_n(\theta) \to 0$ uniformly on $\delta \leq |\theta| \leq \pi$. ♢

Exercise 4.21. Show that a family of kernels satisfying (a), (b′), and (c′) is a family of good kernels according to our definition. Show that conditions (a) and (b′) imply condition (b) and that condition (c′) implies condition (c). ♢

A canonical way to generate a good family of kernels from one function and its dilations is described in the following exercise.

Exercise 4.22. Suppose K is a continuous function on $\mathbb{R}$ that is zero for all $|\theta| \geq \pi$. Assume $\int_{-\pi}^{\pi} |K(\theta)|\, d\theta = 2\pi$. Let $K_n(\theta) := nK(n\theta)$ for $-\pi \leq \theta \leq \pi$. Verify that $\{K_n\}_{n\geq 1}$ is a family of good kernels in $\mathbb{T}$. ♢

Why is a family of kernels satisfying conditions (a), (b), and (c) called *good*? In the context of Fourier series, it is because such a family allows us to recover the values of a continuous function $f : \mathbb{T} \to \mathbb{C}$ from its Fourier coefficients. This mysterious statement will become clear in Section 4.4. In the present Fourier-free context, the kernels are called good because they provide a constructive way of producing approximating functions that are smoother than the limit function.

Theorem 4.23. *Let $\{K_n(\theta)\}_{n=1}^{\infty}$ be a family of good kernels, and let $f : \mathbb{T} \to \mathbb{C}$ be integrable and bounded. Then*

$$\lim_{n\to\infty} (f * K_n)(\theta) = f(\theta)$$

at all points of continuity of f. Furthermore, if f is continuous on the whole circle $\mathbb{T}$, then $f * K_n \to f$ uniformly on $\mathbb{T}$.

This result explains the use of the term *approximation of the identity* for a family of good kernels K_n, for it says that the convolutions of these kernels K_n with f converge, as $n \to \infty$, to the function f again. Moreover, a priori, $f * K_n$ is at the very least continuous. If we can locate a kernel that is k times differentiable, then $f * K_n$ will be at least C^k; see Exercise 4.16. The second part of Theorem 4.23 will say that we can approximate uniformly continuous functions by k times continuously differentiable functions. If we could find a kernel that is a trigonometric polynomial, then $f * K_n$

will be itself a trigonometric polynomial; see Exercise 4.17. The second part of Theorem 4.23 will say that we can approximate uniformly continuous functions by trigonometric polynomials, and Weierstrass's Theorem 3.4 will be proved.

Aside 4.24. Before proving Theorem 4.23, let us indulge in a flight of fancy. Suppose for a moment that the Dirichlet kernel D_N were a good kernel. (The subscript change from n in K_n to N in D_N is just to echo the notation in our earlier discussion of the Dirichlet kernel.) Then Theorem 4.23 would imply that the Fourier partial sums $S_N f$ of a continuous function f on $\mathbb{T}$ would converge pointwise and uniformly to f. We know that this dream cannot be real, since $S_N f(0) \not\to f(0)$ for du Bois-Reymond's function, which is continuous on $\mathbb{T}$. However, we will see below that we can salvage something from what we learn about good kernels. In particular, we *can* recover $f(\theta)$ at points of continuity θ of f, using a modification of the Fourier partial sums. (The only information we need is the Fourier coefficients.) This is Fejér's Theorem about Cesàro means, discussed in Section 4.4 below, and already mentioned in Chapter 3. $\diamondsuit$

Exercise 4.25 (*The Dirichlet Kernel Is a Bad Kernel*)**.** Determine which of the properties of a good kernel fail for the Dirichlet kernel D_N. $\diamondsuit$

In preparation, let us continue our study of good kernels. We begin with the question "Why should $f * K_n(\theta)$ converge to $f(\theta)$ anyway?" As n increases, the mass of $K_n(y)$ becomes concentrated near $y = 0$, and so the mass of $K_n(\theta - y)$ becomes concentrated near $y = \theta$. So the integral in $f * K_n(\theta)$ only "sees" the part of f near $y = \theta$. See Figure 4.3. Here is a model example in which this averaging is exact.

Example 4.26 (*A Good Kernel Made Up of Characteristic Functions*)**.** Consider the kernels

$$K_n(\theta) = n\chi_{[-\pi/n,\pi/n]}(\theta), \qquad n \in \mathbb{N}.$$

These rectangular kernels K_n are pictured in Figure 4.4, for several values of n. It is straightforward to check that these functions K_n are

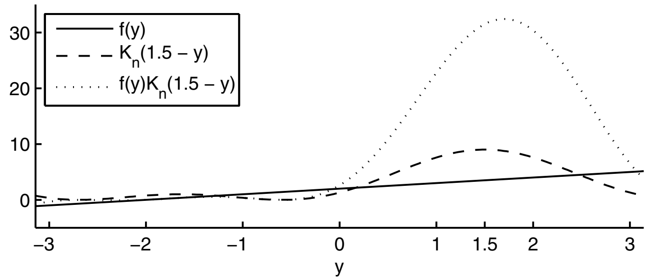


Figure 4.3. Convolution of a function f with a good kernel. The convolution $f * K_n(\theta)$ at $\theta = 1.5$ is given by the integral representing the area under the dotted curve $f(y)K_n(1.5 - \theta)$. The values of $f(y)$ for y near 1.5 contribute most to this integral. We have used the function $f(y) = y+2$, the Fejér kernel $K_n(y) = F_2(y)$, and the point $\theta = 1.5$.

a family of good kernels. Further, for a fixed $\theta \in \mathbb{T}$,

$$\begin{aligned}(f * K_n)(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)\, n\chi_{[-\pi/n,\pi/n]}(\theta - y)\, dy \\ &= \frac{n}{2\pi}\int_{\theta-(\pi/n)}^{\theta+(\pi/n)} f(y)\, dy,\end{aligned}$$

which is the integral average value of f on $[\theta - \pi/n, \theta + \pi/n]$, as for convolution with the characteristic function in Example 4.7. We expect this average value over a tiny interval centered at θ to converge to the value $f(\theta)$ as $n \to \infty$. This is certainly true if f is continuous at θ, by an application of the Fundamental Theorem of Calculus. In fact, if f is continuous, let $F(\theta) := \int_{-\pi}^{\theta} f(y)\, dy$. Then $F'(\theta) = f(\theta)$, and by the Fundamental Theorem of Calculus,

$$\frac{n}{2\pi}\int_{\theta-(\pi/n)}^{\theta+(\pi/n)} f(y)\, dy = \frac{n}{2\pi}\left[F\big(\theta + (\pi/n)\big) - F\big(\theta - (\pi/n)\big)\right].$$

It is not hard to see that the limit as $n \to \infty$ of the right-hand side is exactly $F'(\theta) = f(\theta)$. ◊

Exercise 4.27. Show that if f is continuous at a point x, then the integral averages over intervals that are shrinking to the point x must converge to $f(x)$. Do not use the Fundamental Theorem of Calculus.

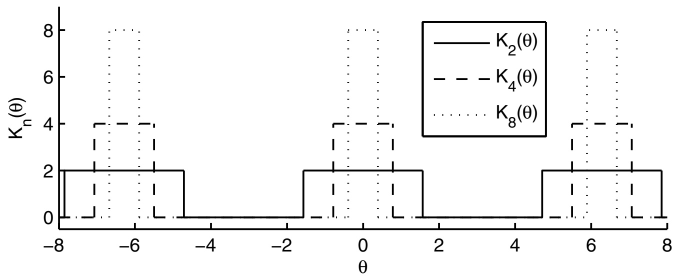


Figure 4.4. Graphs of rectangular kernels $K_n(\theta)$, for $n = 2$, 4, and 8, extended periodically from $[-\pi, \pi)$ to $\mathbb{R}$.

Instead use the observation that the integral average must be trapped between the minimum and maximum values of f on the interval and then use continuity to conclude that as the interval shrinks to x, the maximum and minimum values converge to $f(x)$. ◊

We are now ready to return to the approximation property enjoyed by good kernels.

Proof of Theorem 4.23. First $f : \mathbb{T} \to \mathbb{C}$ is assumed to be integrable and bounded. Let B be a bound for f on $\mathbb{T}$. Suppose f is continuous at some point $\theta \in \mathbb{T}$. Fix $\varepsilon > 0$. Choose δ such that

$$|f(\theta - y) - f(\theta)| < \varepsilon \qquad \text{whenever } |y| < \delta. \tag{4.7}$$

Now we can write an integral for the pointwise difference between $K_n * f$ and f:

$$\begin{aligned}(K_n * f)(\theta) - f(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y) f(\theta - y)\,dy - f(\theta) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y) f(\theta - y)\,dy - f(\theta)\,\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)\,dy \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)\,[f(\theta - y) - f(\theta)]\,dy.\end{aligned}$$

In the second line we were able to multiply $f(\theta)$ by $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)\,dy$ (another typical analysis trick), since this quantity is 1 by condition (a) on good kernels. We can control the absolute value of the integral by the integral of the absolute values:

$$|(K_n * f)(\theta) - f(\theta)| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |K_n(y)|\,|f(\theta - y) - f(\theta)|\, dy.$$

Yet another common technique in analysis is to estimate an integral (that is, to find an upper bound for the absolute value of the integral) by splitting the domain of integration into two regions, typically one region where some quantity is small and one where it is large, and controlling the two resulting integrals by different methods. We use this technique now, splitting the domain $\mathbb{T}$ into the regions where $|y| \leq \delta$ and $\delta \leq |y| \leq \pi$, so that $\int_{-\pi}^{\pi} \cdots = \int_{|y|\leq\delta} \cdots + \int_{\delta\leq|y|\leq\pi} \cdots$.

For $y \leq \delta$ we have $|f(\theta - y) - f(\theta)| < \varepsilon$, and so

$$\begin{aligned} &\frac{1}{2\pi} \int_{|y|\leq\delta} |K_n(y)||f(\theta - y) - f(\theta)|\, dy \\ &\qquad \leq \frac{\varepsilon}{2\pi} \int_{|y|\leq\delta} |K_n(y)|\, dy \leq \frac{\varepsilon}{2\pi} \int_{-\pi}^{\pi} |K_n(y)|\, dy \leq \frac{\varepsilon M}{2\pi}. \end{aligned}$$

We have used condition (b) on good kernels in the last inequality.

Condition (c) on good kernels implies that for our δ, as $n \to \infty$, $\int_{\delta\leq|y|\leq\pi} |K_n(y)|\, dy \to 0$. So there is some N such that for all $n \geq N$, $\int_{\delta\leq|y|\leq\pi} |K_n(y)|\, dy \leq \varepsilon$. Also, $|f(\theta - y) - f(\theta)| \leq 2B$. So we can estimate the integral over $\delta \leq |y| \leq \pi$ and for $n \geq N$ by

$$\frac{1}{2\pi} \int_{\delta\leq|y|\leq\pi} |K_n(y)|\,|f(\theta - y) - f(\theta)|\, dy \leq \frac{2B\varepsilon}{2\pi}.$$

Putting together the contributions from both pieces, we conclude that for $n \geq N$ we have

$$(4.8) \qquad |(K_n * f)(\theta) - f(\theta)| \leq (M + 2B)\,\varepsilon/(2\pi).$$

This proves the first part of the theorem.

Now suppose that f is continuous on $\mathbb{T}$. Then it is uniformly continuous. This means that for fixed $\varepsilon > 0$ there is some $\delta > 0$ such that inequality (4.7) holds for all $\theta \in \mathbb{T}$ and $|y| < \delta$. We can repeat verbatim the previous argument and conclude that the estimate (4.8) holds for all $\theta \in \mathbb{T}$; in other words, $K_n * f \to f$ uniformly. □

Exercise 4.28. Check that if F is continuously differentiable at θ and $h > 0$, then $\lim_{h\to 0} \big[F(\theta + h) - F(\theta - h)\big]/2h = F'(\theta)$. ♢

For an arbitrary integrable function f, the limit as n goes to infinity of the convolution with the good kernel K_n discussed in Example 4.26 does pick out the value of f at all points θ where f is continuous. The L^1 version of this averaging limiting result is called the *Lebesgue Differentiation Theorem*.

Theorem 4.29 (Lebesgue Differentiation Theorem). *If* $f \in L^1(\mathbb{T})$, *then* $\lim_{h\to 0} \frac{1}{2h}\int_{\theta-h}^{\theta+h} f(y)\,dy = f(\theta)$ *for a.e.* $\theta \in \mathbb{T}$.

Proofs of this result require a thorough understanding of measure theory. For example, see [**SS05**, Chapter 3, Section 1.2].

4.4. Fejér kernels and Cesàro means

Returning to an earlier question, can we recover the values of an integrable function f from knowledge of its Fourier coefficients? Perhaps just at points where f is continuous?

Du Bois-Reymond's example shows that even if f is continuous on the whole circle $\mathbb{T}$, we cannot hope to recover f by taking the limit as $n \to \infty$ of the traditional partial sums $S_N f$. However, Fejér discovered that we *can* always recover a *continuous* function f from its Fourier coefficients if we use a different method of summation, which had been developed by Cesàro. If f is just integrable, the method recovers f at all points where f is continuous.

Fejér's method boils down to using *Cesàro partial sums* $\sigma_N f$ corresponding to convolution with a particular family of good kernels, now called the Fejér kernels. The Cesàro partial sums, also known as *Cesàro means*, are defined by

$$\begin{aligned}\sigma_N f(\theta) &:= \big[S_0 f(\theta) + S_1 f(\theta) + \cdots + S_{N-1} f(\theta)\big]/N \\ &= \frac{1}{N}\sum_{0\le n\le N-1} S_n f(\theta) \;=\; \frac{1}{N}\sum_{0\le n\le N-1}\;\sum_{|k|\le n} \widehat{f}(k)\, e^{ik\theta}.\end{aligned}$$

Thus $\sigma_N f$ is the average of the first N partial Fourier sums $S_n f$, and it is a 2π-periodic trigonometric polynomial (see also Exercise 4.17). Interchanging the order of summation and counting the number of appearances of each Fourier summand $\widehat{f}(k)e^{ik\theta}$ gives a representation of $\sigma_N f$ as a weighted average of the Fourier coefficients of f corresponding to frequencies $|k| \le N$, where the coefficients corresponding to smaller frequencies are weighted most heavily. More precisely,

$$\sigma_N f(\theta) = \sum_{|k|\leq N-1} \big[(N-|k|)/N\big]\widehat{f}(k)e^{ik\theta}.$$

The *Fejér kernel* $F_N(\theta)$ is defined to be the arithmetic average of the first N Dirichlet kernels:

$$(4.9) \qquad F_N(\theta) := \big[D_0(\theta) + D_1(\theta) + \cdots + D_{N-1}(\theta)\big]/N.$$

By the same calculations performed above we conclude that the Fejér kernel can be written as a weighted average of the trigonometric functions corresponding to frequencies $|k| \leq N$, where the coefficients corresponding to smaller frequencies are weighted most heavily. Specifically, $F_N(\theta) = \sum_{|n|\leq N} \big[(N-|n|)/N\big]\, e^{in\theta}$. What is not so obvious is that there is a closed formula for the Fejér kernel,

$$(4.10) \qquad F_N(\theta) = (1/N)\big[\sin(N\theta/2)/\sin(\theta/2)\big]^2.$$

Exercise 4.30. Verify formula (4.10). ♢

Compare with formula (4.3) for the Dirichlet kernel in terms of sines. Notice that, unlike the Dirichlet kernel, the Fejér kernel is nonnegative. See Figure 4.5.

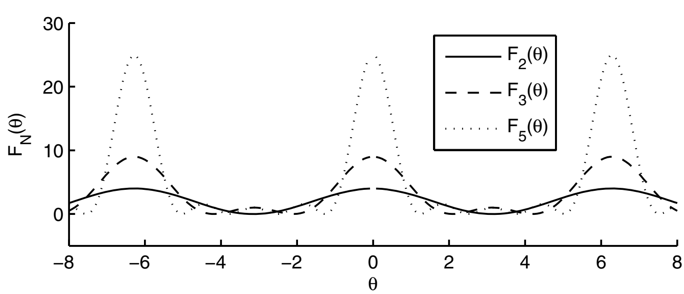


Figure 4.5. Graphs of Fejér kernels $F_N(\theta)$ for $N = 1$, 3, and 5, extended periodically from $[-\pi, \pi)$ to $\mathbb{R}$.

Exercise 4.31 (*The Fejér Kernel Is a Good Kernel*)**.** Show that the Fejér kernel F_N is a good kernel and that $\sigma_N f = F_N * f$. ♢

By Theorem 4.23 on good kernels, it follows that the Cesàro sums of f converge to f at points of continuity.

Theorem 4.32 (Fejér, 1900)**.** *Assume $f \in L^1(\mathbb{R})$ and that it is bounded. Then the following statements hold.*

(i) *At all points of continuity of f there is pointwise convergence of the Cesàro sums. In other words, if f is continuous at θ, then*

$$f(\theta) = \lim_{N\to\infty} (f * F_N)(\theta) = \lim_{N\to\infty} \sigma_N f(\theta).$$

(ii) *If f is continuous on $\mathbb{T}$, then $\sigma_N f \to f$ uniformly on $\mathbb{T}$.*

Thus we can indeed recover a continuous function f from knowledge of its Fourier coefficients.

Let us reiterate that the Cesàro sums give a way to approximate continuous functions on $\mathbb{T}$ uniformly by trigonometric polynomials. That is Weierstrass's Theorem 3.4, which we stated in Section 3.1 as an appetizer.

Fejér's Theorem gives a proof of the uniqueness of Fourier coefficients for continuous and periodic functions (the uniqueness principle): *If $f \in C(\mathbb{T})$ and all its Fourier coefficients vanish, then $f = 0$.*

Exercise 4.33 (*The Fourier Coefficients Are Unique*)**.** Prove the uniqueness principle for continuous functions on the circle. Show that the uniqueness principle implies that if $f, g \in C(\mathbb{T})$ and $\widehat{f}(n) = \widehat{g}(n)$ for all $n \in \mathbb{Z}$, then $f = g$. ♢

A similar argument shows that if f is Riemann integrable and all its Fourier coefficients are zero, then at all points of continuity of f, $f(\theta) = \lim_{N\to\infty} \sigma_N f(\theta) = 0$. Lebesgue's Theorem (Theorem 2.33) says that f is Riemann integrable if and only if f is continuous almost everywhere, so the conclusion can be strengthened to $f = 0$ a.e.

With the machinery of the Lebesgue integral one can prove the same result for Lebesgue-integrable functions.

Theorem 4.34 (Uniqueness Principle)**.** *If $f \in L^1(\mathbb{T})$ and $\widehat{f}(n) = 0$ for all n, then $f = 0$ a.e.*

4.5. Poisson kernels and Abel means

The Poisson kernel is another good kernel. Convolution of a function f with the Poisson kernel yields a new quantity known as the *Abel mean*[10], analogous to the way in which convolution with the Fejér kernel yields the Cesàro mean. Some differences are that the Poisson kernel is indexed by a continuous rather than a discrete parameter and that each Abel mean involves all the Fourier coefficients rather than just those for $|n| \leq N$. See [**SS03**, Sections 5.3 and 5.4] for a fuller discussion, including the notion of Abel summability of the Fourier series to f at points of continuity of the integrable function f, and an application to solving the heat equation on the unit disk.

Definition 4.35. The *Poisson kernel* $P_r(\theta)$ is defined by

$$P_r(\theta) := \sum_{n\in\mathbb{Z}} r^{|n|}e^{in\theta} = \frac{1-r^2}{1-2r\cos\theta+r^2} \quad \text{for } r \in [0,1). \qquad \Diamond$$

Notice that the Poisson kernel is indexed by all real numbers r between 0 and 1, not by the discrete positive integers N as in the Dirichlet and Fejér kernels. (See Figure 4.6.) We are now interested in the behavior as r increases towards 1, instead of $N \to \infty$. The Poisson kernel is positive for $r \in [0,1)$. The series for $P_r(\theta)$ is absolutely convergent and uniformly convergent.

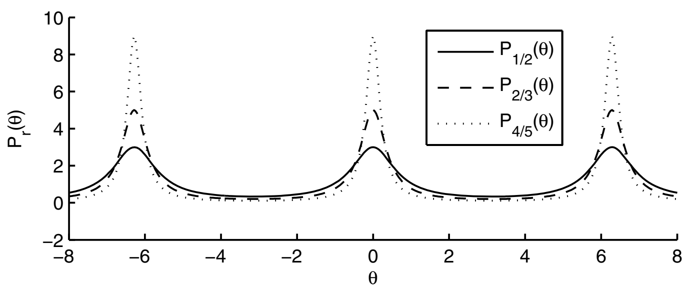


Figure 4.6. Graphs of Poisson kernels $P_r(\theta)$ for $r = 1/2$, $2/3$, and $4/5$, extended periodically from $[-\pi, \pi)$ to $\mathbb{R}$.

---

[10]Named after the Norwegian mathematician Niels Henrik Abel (1802–1829).

Exercise 4.36. Show that the two formulas in our definition of the Poisson kernel are actually equal. ◊

Exercise 4.37 (*The Poisson Kernel Is a Good Kernel*)**.** Show that the Poisson kernel $P_r(\theta)$ is a good kernel. (Modify the definition of *good kernel* to take into account the change in index from $n \in \mathbb{N}$ to $r \in [0,1)$.) ◊

Definition 4.38. Suppose f has Fourier series $\sum_{n\in\mathbb{Z}} \widehat{f}(n)\, e^{in\theta}$. For $r \in [0,1)$, the r^{th} *Abel mean* of f is defined by

$$A_r f(\theta) := \sum_{n\in\mathbb{Z}} r^{|n|} \widehat{f}(n)\, e^{in\theta}. \qquad \lozenge$$

Thus the Abel mean is formed from f by multiplying the Fourier coefficients $\widehat{f}(n)$ by the corresponding factor $r^{|n|}$.

The Abel mean arises from the Poisson kernel by convolution, just as the Cesàro mean arises from the Fejér kernel by convolution.

Exercise 4.39. Show that for integrable functions $f : \mathbb{T} \to \mathbb{C}$,

$$A_r f(\theta) = (P_r * f)(\theta). \tag{4.11}$$ ◊

It follows from Theorem 4.23, modified for the continuous index $r \in [0,1)$, that if $f : \mathbb{T} \to \mathbb{C}$ is Riemann integrable and f is continuous at $\theta \in \mathbb{T}$, then

$$f(\theta) \;=\; \lim_{r\to 1^-} (f * P_r)(\theta) \;=\; \lim_{r\to 1^-} A_r f(\theta).$$

Thus we can recover the values of f (at points of continuity of f) from the Abel means of f, just as we can from the Cesàro means. If f is continuous, then the convergence is uniform.

Exercise 4.40. Write a MATLAB script to plot the Dirichlet kernels D_N, the rectangular kernels K_n, the Féjer kernels F_N, or the Poisson kernels P_r. **Hint for the Poisson kernels:** Replace the continuous parameter r by the discrete parameter $r_n = 1 - 1/n$. ◊

See the project in Section 4.8 for more on summability methods.

4.6. Excursion into $L^p(\mathbb{T})$

We will be concerned with other modes of convergence for the convolution of a family of good kernels and a function. Examples are convergence in $L^2(\mathbb{T})$ and also in $L^p(\mathbb{T})$.

Theorem 4.41. *If $f \in C(\mathbb{T})$ and $\{K_n\}_{n\geq 1}$ is a family of good kernels, then* $\lim_{n\to\infty} \|f * K_n - f\|_{L^p(\mathbb{T})} = 0$ *for* $1 \leq p < \infty$.

Exercise 4.42. Prove Theorem 4.41. **Hint:** You can control the L^p norm with the L^∞ norm on $\mathbb{T}$. Recall the relations between different modes of convergence illustrated in Figure 2.4. ♢

In particular, this theorem implies the convergence in $L^p(\mathbb{T})$ of the Cesàro means $\sigma_N f$ and the Abel means $A_r f$ for continuous functions $f \in C(\mathbb{T})$. We state explicitly the mean-square convergence for continuous functions and the Cesàro means.

Theorem 4.43. *If $f \in C(\mathbb{T})$, then* $\lim_{N\to\infty} \|\sigma_N f - f\|_{L^2(\mathbb{T})} = 0$. *In other words, the Cesàro means $\sigma_N f$ converge to f in the L^2 sense.*

In the next chapter we are interested in proving the mean-square convergence of the partial Fourier sums of f. We will prove that $\|S_N f - f\|_{L^2(\mathbb{T})} \to 0$, and Theorem 4.43 will play a rôle. We will first better understand the geometry of $L^2(\mathbb{T})$ and of $S_N f$. In particular we will see that $S_N f$ is the *best approximation to f in the L^2 norm in the space $\mathcal{P}_N(\mathbb{T})$ of trigonometric polynomials of degree N.*

In Section 3.2 we proved the Riemann–Lebesgue Lemma for continuous functions; see Lemma 3.20. We can use an approximation argument to prove the lemma for integrable functions.

Lemma 4.44 (Riemann–Lebesgue Lemma for Integrable Functions)**.** *If $f \in L^1(\mathbb{T})$, then $\widehat{f}(n) \to 0$ as $|n| \to \infty$.*

Proof. The result holds for $f \in C(\mathbb{T})$ by Lemma 3.20. Given $f \in L^1(\mathbb{T})$, there is a sequence of continuous functions $\{f_k\}_{k\in\mathbb{N}}$ such that $\|f_k - f\|_{L^1(\mathbb{T})} \to 0$ as $k \to \infty$; see Theorem 2.75. In particular their Fourier coefficients are close: for each fixed n,

$$|\widehat{f}(n) - \widehat{f_k}(n)| \leq \frac{1}{2\pi}\|f_k - f\|_{L^1(\mathbb{T})} \to 0 \qquad \text{as } k \to \infty. \tag{4.12}$$

Now we estimate the size of the Fourier coefficients of f, knowing that for the continuous functions f_k, $\widehat{f_k}(n) \to 0$ as $|n| \to \infty$. Fix $\varepsilon > 0$. There is some $K > 0$ such that $\|f_k - f\|_{L^1(\mathbb{T})} < \pi\varepsilon$ for all $k > K$. Hence, by inequality (4.12), $|\widehat{f}(n) - \widehat{f_k}(n)| < \varepsilon/2$, for all $n > 0$ and all $k > K$. Fix $k > K$. There is an $N > 0$ such that for all $|n| > N$ and for the particular $k > K$, $|\widehat{f_k}(n)| < \varepsilon/2$. Finally, by the Triangle Inequality for real numbers, we see that for all $|n| > N$,

$$|\widehat{f}(n)| \leq |\widehat{f}(n) - \widehat{f_k}(n)| + |\widehat{f_k}(n)| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Since our argument is valid for all positive ε, we have shown that $\lim_{|n|\to\infty} |\widehat{f}(n)| = 0$ for all $f \in L^1(\mathbb{T})$, proving the lemma. □

4.7. Project: Weyl's Equidistribution Theorem

This project is based on Körner's account [**Kör**, Chapter 3]. For other accounts see [**Pin**, Section 1.4.4], [**SS03**, Section 4.2], and [**DM**, Section 1.7.6]. Let $\langle\gamma\rangle$ denote the *fractional part* of a real number γ, so that $0 \leq \langle\gamma\rangle < 1$ and $\gamma - \langle\gamma\rangle$ is an integer.

Given a real number γ, is there any pattern in the fractional parts $\langle\gamma\rangle$, $\langle 2\gamma\rangle$, $\langle 3\gamma\rangle$, …, $\langle n\gamma\rangle$ of the positive integer multiples of γ?

Weyl's Equidistribution Theorem[11] states that if γ is irrational, then for large n the fractional parts of its first n integer multiples are more or less uniformly scattered over the interval $[0,1]$. Specifically, these integer multiples are *equidistributed* in $[0,1]$, in the sense that the proportion of the first n fractional parts $\{\langle\gamma\rangle, \langle 2\gamma\rangle, \dots, \langle n\gamma\rangle\}$ that fall within any given subinterval $[a,b]$ of $[0,1]$ approaches, as $n \to \infty$, the proportion $b-a = (b-a)/(1-0)$ of $[0,1]$ that is occupied by $[a,b]$. Mathematically, Weyl's Equidistribution Theorem says that if γ is irrational, then for every a, b such that $0 \leq a \leq b \leq 1$, the proportion $\text{card}(\{r : 1 \leq r \leq n, a \leq \langle r\gamma\rangle \leq b\})/n \to b-a$ as $n \to \infty$. By $\text{card}(A)$ we mean the cardinality of the set A. For a finite set A, $\text{card}(A)$ is just the number of elements in A. We have also used the notation $\text{card}(A) = \#A$.

[11] Named after the German mathematician Hermann Klaus Hugo Weyl (1885–1955).

Perhaps surprisingly, Weyl's result can be proved using one of the fundamental ideas of Fourier analysis that we have learned in this chapter: every periodic continuous function can be uniformly approximated by trigonometric polynomials.

(a) Work through the proof of Weyl's Equidistribution Theorem given in [**Kör**, Chapter 3]. Focus on the use of continuous functions close to the characteristic function of $[2\pi a, 2\pi b]$ to approximate the cardinality of the set mentioned above, and the use of the density of the trigonometric polynomials in the continuous functions on $\mathbb{T}$.

(b) Choose several irrational and rational numbers, such as $\sqrt{2}$, π, $22/7$, the golden ratio $\phi = (1 + \sqrt{5})/2$, and so on. For each of your numbers, use MATLAB to create a labeled histogram showing the fractional parts of the first 2,000 integer multiples, divided into ten or more bins within $[0, 1]$.

(c) Explain why the fractional parts of the integer multiples of a rational number are not equidistributed in $[0, 1]$. It follows that the real numbers γ such that $\{\langle r\gamma \rangle\}$ is equidistributed in $[0, 1]$ are exactly the irrational numbers.

(d) The situation for integer *powers* is less clear-cut than that for integer *multiples* and indeed is not yet fully understood. Work through the short proof in [**Kör**, Chapter 3] showing that $\langle \phi^r \rangle$, the fractional parts of the integer powers of the golden ratio $\phi = (1+\sqrt{5})/2$, are *not* equidistributed in $[0, 1]$. (The proof makes use of the connection between the golden ratio and the Fibonacci numbers.) Thus for at least one irrational number, the sequence $\{\langle \gamma^r \rangle\}$ is not equidistributed. It is known that for almost all real numbers $\gamma > 1$, the sequence $\{\langle \gamma^r \rangle\}$ of fractional parts of the integer powers of γ is equidistributed. Use MATLAB to create histograms, like those in part (b), for integer powers of some rational and irrational numbers, looking for both equidistributed and nonequidistributed sequences. Can you formulate and prove any results suggested by your experiments?

(e) The concepts of being *dense* in $[0, 1]$ and being *equidistributed* in $[0, 1]$ are different. Does either imply the other? Can you find a sequence (not necessarily of fractional parts of integer multiples of a given number) that has one property but not the other?

4.8. Project: Averaging and summability methods

In the first part of this project we explore how the two averaging methods known as *Cesàro means* and *Abel means* do not damage, but can in fact improve, the convergence properties of sequences. In Sections 4.4 and 4.5 we used the Cesàro and Abel means in the special case where the sequence $\{b_n\}_{n=0}^{\infty}$ consists of the partial Fourier sums of a given function $f : \mathbb{T} \to \mathbb{C}$. For a captivating account of the history surrounding the use of summability methods for Fourier series, see [**KL**].

Let $\{b_n\}_{n=1}^{\infty}$ be a sequence of real numbers. The *Cesàro means* σ_n *of the sequence* $\{b_n\}_{n=1}^{\infty}$, also known as *arithmetic means* or just *averages*, are defined by $\sigma_n := (b_1 + b_2 + \cdots + b_n)/n$.

For $0 \leq r < 1$, the *Abel means* A_r *of the sequence* $\{b_n\}_{n=0}^{\infty}$ are defined by $A_r := (1-r)\sum_{n=0}^{\infty} r^n b_n$.

(a) Suppose that the sequence $\{b_n\}_{n=1}^{\infty}$ converges to the number b, in other words, $\lim_{n\to\infty} b_n = b$. Show that the sequence $\{\sigma_n\}_{n=1}^{\infty}$ of Cesàro means converges and that its limit is also b. First experiment with some examples, such as $b_n = 1$, $b_n = n$, and $b_n = 1/n$. Can you find an example of a sequence $\{b_n\}_{n=1}^{\infty}$ such that the sequence $\{\sigma_n\}_{n=1}^{\infty}$ of averages converges but the original sequence diverges?

(b) Given an arbitrary fraction p/q in the unit interval, can you find a sequence of ones and zeros such that the Cesàro means converge to p/q? Notice that such a sequence must have infinitely many ones and infinitely many zeros. What about any real number α in the unit interval, not necessarily a fraction?

(c) Suppose that the sequence $\{b_n\}_{n=0}^{\infty}$ converges to the number b. Show that as $r \to 1^-$ the Abel means A_r also approach b. Can you find an example of a sequence $\{b_n\}_{n=0}^{\infty}$ so that the limit exists but the sequence does not converge? Is there a sequence whose Abel means do not have a limit? Notice that we can think of the Abel mean as the double limit $\lim_{r\to1^-}\lim_{N\to\infty}(1-r)\sum_{n=0}^{N} r^n b_n$. What happens if you interchange the limits?

(d) In order to work with Abel means of sequences, it is useful to note the identity $(1-r)\sum_{n=0}^{\infty} r^n s_n = \sum_{n=0}^{\infty} r^n b_n$, valid for $0 \leq r < 1$

and for $s_n = \sum_{k=0}^{n} b_k$, $n \in \mathbb{N}$. This identity allows us to go back and forth between a sequence $\{b_n\}_{n=0}^{\infty}$ and its partial sums $\{s_n\}_{n=0}^{\infty}$. Prove the identity.

(e) What happens to the sequences of ones and zeros constructed in part (b) when you compute the limit in (c)? Is this a coincidence?

(f) Can you find a sequence $\{b_n\}_{n=0}^{\infty}$ such that the sequence of averages diverges, but the limit in (c) exists? Comparing our summability methods using the information gathered so far, which one is stronger, Cesàro means or Abel means? **Hint:** Consider the sequence $\{(-1)^n(n+1)\}_{n=0}^{\infty}$. See [**SS03**, Section 5.3]. For more on summability methods and matrices, see [**Pin**, Section 1.4.2].

Chapter 5

Mean-square convergence of Fourier series

This chapter has two main themes: the convergence in the L^2 sense of the partial Fourier sums $S_N f$ to f (Section 5.1) and the completeness of the trigonometric functions in $L^2(\mathbb{T})$ (Section 5.3). They are intimately related, as a review of the geometry of $L^2(\mathbb{T})$ shows (Section 5.2). We begin by setting the scene.

In the previous chapter, we discussed pointwise and uniform convergence of the partial Fourier sums $S_N f$ to the function f. We found that continuity of f is not sufficient to ensure pointwise convergence of $S_N f$. We need something more, such as f having two continuous derivatives, or the use of Cesàro sums instead of $S_N f$; either of these conditions guarantees uniform convergence.

In this chapter we discuss *mean-square convergence*, also known as *convergence in* $L^2(\mathbb{T})$. Suppose f is continuous, or just in $L^2(\mathbb{T})$. We will see that the N^{th} partial Fourier sum $S_N f$ is the trigonometric polynomial of degree at most N which is *closest* to f in the L^2 norm, in the sense that $S_N f$ is the *orthogonal projection* of f onto the subspace of $L^2(\mathbb{T})$ consisting of trigonometric polynomials of degree N. The *energy* or L^2 *norm* of f can be recovered from the Fourier coefficients $\{\widehat{f}(n)\}_{n\in\mathbb{Z}}$ alone, via the invaluable *Parseval's Identity*[1], which is an infinite-dimensional analogue of the Pythagorean Theorem[2]. What is behind all these results can be summarized in one very important sentence: *The trigonometric functions* $\{e^{in\theta}\}_{n\in\mathbb{Z}}$ *form an orthonormal basis for* $L^2(\mathbb{T})$.

It is immediate that the trigonometric functions are orthonormal on $\mathbb{T}$ (see formula (1.6)). What is really noteworthy here is that the family of trigonometric functions is *complete*, in the sense that they *span* $L^2(\mathbb{T})$: every square-integrable function on $\mathbb{T}$ can be written as a (possibly infinite) linear combination of the functions $e^{in\theta}$. We discuss equivalent conditions for the completeness of an orthonormal system of functions in $L^2(\mathbb{T})$ (Section 5.4).

5.1. Basic Fourier theorems in $L^2(\mathbb{T})$

In Section 2.1 we introduced the space $L^2(\mathbb{T})$ of square-integrable functions on $\mathbb{T}$ as the space of functions $f : \mathbb{T} \to \mathbb{C}$ such that

$$\int_{-\pi}^{\pi} |f(\theta)|^2 \, d\theta < \infty.$$

Recall that the L^2 *norm*, or *energy*, of a function $f \in L^2(\mathbb{T})$, denoted by $\|f\|_{L^2(\mathbb{T})}$, is the nonnegative, finite quantity defined by

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

Before we describe in more detail the geometric structure of the space $L^2(\mathbb{T})$, let us look ahead to the main results concerning Fourier series and $L^2(\mathbb{T})$.

The Fourier coefficients for $f \in L^2(\mathbb{T})$ are well-defined (since square-integrable functions on $\mathbb{T}$ are integrable), as are its partial Fourier sums

$$S_N f(\theta) = \sum_{|n| \le N} \widehat{f}(n) e^{2\pi i n\theta}.$$

[1] Named after the French mathematician Marc-Antoine Parseval de Chênes (1755–1836).

[2] Named after the Greek philosopher Pythagoras of Samos (c. 570–c. 495 BC).

Theorem 5.1 (Mean-square Convergence Theorem). *If $f \in L^2(\mathbb{T})$, then its partial Fourier sums $S_N f$ converge to f in the L^2 sense:*

$$\int_{-\pi}^{\pi} |S_N f(\theta) - f(\theta)|^2 \, d\theta \to 0 \qquad \text{as } N \to \infty.$$

Restating Theorem 5.1 in terms of norms, if $f \in L^2(\mathbb{T})$, then

$$\|S_N f - f\|_{L^2(\mathbb{T})} \to 0 \qquad \text{as } N \to \infty.$$

We can recover the square of the *energy*, or norm, of an $L^2(\mathbb{T})$ function f by adding the squares of the absolute values of its Fourier coefficients, as Parseval's Identity shows.

Theorem 5.2 (Parseval's Identity). *If $f \in L^2(\mathbb{T})$, then*

$$\|f\|^2_{L^2(\mathbb{T})} = \sum_{n \in \mathbb{Z}} |\widehat{f}(n)|^2.$$

We prove Theorems 5.1 and 5.2 later in this chapter. Parseval's Identity immediately gives us an insight into the behavior of Fourier coefficients of square-integrable functions. Since the series on the right-hand side of Parseval's Identity converges, the coefficients must vanish as $|n| \to \infty$, as described in the following corollary.

Corollary 5.3 (Riemann–Lebesgue Lemma). *If $f \in L^2(\mathbb{T})$, then*

$$\widehat{f}(n) \to 0 \qquad \text{as } |n| \to \infty.$$

The space of *square-summable sequences*, denoted by $\ell^2(\mathbb{Z})$ and pronounced "little ell-two of the integers", is defined by

$$\ell^2(\mathbb{Z}) := \Big\{ \{a_n\}_{n \in \mathbb{Z}} : a_n \in \mathbb{C} \text{ and } \sum_{n \in \mathbb{Z}} |a_n|^2 < \infty \Big\}. \tag{5.1}$$

The space $\ell^2(\mathbb{Z})$ is a normed space, with norm given by

$$\|\{a_n\}_{n \in \mathbb{Z}}\|_{\ell^2(\mathbb{Z})} := \Big(\sum_{n \in \mathbb{Z}} |a_n|^2 \Big)^{1/2}.$$

In the language of norms, we can restate Parseval's Identity as

$$\|f\|_{L^2(\mathbb{T})} = \|\{\widehat{f}(n)\}_{n \in \mathbb{Z}}\|_{\ell^2(\mathbb{Z})}. \tag{5.2}$$

In other words, the Fourier mapping $\mathcal{F} : L^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$, which associates to each function $f \in L^2(\mathbb{T})$ the sequence of its Fourier coefficients $\mathcal{F}(f) = \{\widehat{f}(n)\}_{n \in \mathbb{Z}}$, preserves norms: $\mathcal{F}$ is an isometry.

In the Appendix you will find the definitions of *normed spaces* and *complete inner-product vector spaces*, also known as *Hilbert spaces*. It turns out that $L^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$ are Hilbert spaces[3]. The Fourier mapping also preserves the inner products.

5.2. Geometry of the Hilbert space $L^2(\mathbb{T})$

The space $L^2(\mathbb{T})$ of square-integrable functions on $\mathbb{T}$ is a *vector space* with the usual addition and scalar multiplication for functions. The L^2 *norm* is induced by an *inner product* defined for f, $g \in L^2(\mathbb{T})$ by

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

in the sense that $\|f\|_{L^2(\mathbb{T})} = \langle f, f\rangle^{1/2}$.

Some inner-product vector spaces satisfy another property that has to do with convergence of Cauchy sequences of vectors. Namely, they are *complete inner-product vector spaces*, also known as *Hilbert spaces*. In particular, $L^2(\mathbb{T})$ is a Hilbert space. However, the space $\mathcal{R}(\mathbb{T})$ of Riemann integrable functions on $\mathbb{T}$, with the L^2 inner product, is not a Hilbert space.

Definition 5.4. A sequence of functions $f_n \in L^2(\mathbb{T})$ is a *Cauchy sequence* if for every $\varepsilon > 0$ there is an $N > 0$ such that $\|f_n - f_m\|_{L^2(\mathbb{T})} < \varepsilon$ for all n, $m > N$. $\diamondsuit$

We say that a sequence of functions $\{f_n\}_{n\geq 0}$ converges to f in the *mean-square sense*, or the L^2 *sense*, if

$$\|f_n - f\|_{L^2(\mathbb{T})} \to 0 \qquad \text{as } n \to \infty.$$

It is well known that if a sequence converges to a limit, then as $n \to \infty$ the functions in the sequence form a Cauchy sequence: they get closer to each other. It is a remarkable fact that for $L^2(\mathbb{T})$, the converse is also true.

[3]Here is one of the few places where we really need the Lebesgue integral: for $L^2(\mathbb{T})$ to be a Hilbert space. If we use the Riemann integral instead, then we still get an inner-product vector space, but it is not complete. There are Cauchy sequences of Riemann square-integrable functions that converge to functions that are not Riemann square integrable. Luckily, they are always Lebesgue square integrable. See Section 5.2.

Theorem 5.5. *The space $L^2(\mathbb{T})$ is a* Hilbert space*, that is, a* complete inner-product vector space*. In other words, every Cauchy sequence in $L^2(\mathbb{T})$, with respect to the norm induced by the inner product, converges to a function in $L^2(\mathbb{T})$.*

For a proof, see for example [**SS05**]. Here we use the word *complete* with a different meaning than in Definition 5.19, where we consider *complete systems of orthonormal functions*. It will be clear from the context whether we mean a *complete space* or a *complete system* $\{f_n\}_{n\geq 0}$ *of functions*.

Theorem 5.6. *The space $L^2(\mathbb{T})$ of Lebesgue square-integrable functions on $\mathbb{T}$ consists of the collection $\mathcal{R}(\mathbb{T})$ of all Riemann integrable functions on $\mathbb{T}$, together with all the functions that arise as limits of Cauchy sequences in $\mathcal{R}(\mathbb{T})$ with respect to the L^2 norm. In other words, $L^2(\mathbb{T})$ is the* completion *of $\mathcal{R}(\mathbb{T})$ with respect to the L^2 metric.*

In particular the Riemann integrable functions on $\mathbb{T}$ are dense in $L^2(\mathbb{T})$, as is every dense subset of $\mathcal{R}(\mathbb{T})$ with respect to the L^2 metric. For example step functions, polynomials, and continuous functions on $\mathbb{T}$ are all dense in $L^2(\mathbb{T})$. (Now is a good time to revisit Section 2.4.)

Example 5.7. The space $\ell^2(\mathbb{Z})$ of square-summable sequences with inner product $\langle\{a_n\}_{n\in\mathbb{Z}}, \{b_n\}_{n\in\mathbb{Z}}\rangle_{\ell^2(\mathbb{Z})} := \sum_{n\in\mathbb{Z}} a_n\overline{b_n}$ is a Hilbert space. ◊

Exercise 5.8. Prove that $\ell^2(\mathbb{Z})$ is complete. **Hint:** First generate a candidate for the limit, by observing that if a sequence $\{A^N\}_{N\in\mathbb{Z}}$ of sequences $A^N = \{a_n^N\}_{n\in\mathbb{Z}}$ is a Cauchy sequence in $\ell^2(\mathbb{Z})$, then for each fixed $n \in \mathbb{Z}$, the numerical sequence $\{a_n^N\}_{N\geq 1}$ is Cauchy and hence convergent. Let $A = \{a_n\}_{n\in\mathbb{Z}}$, where $a_n = \lim_{N\to\infty} a_n^N$. Show that $A \in \ell^2(\mathbb{Z})$. Finally show that $\lim_{N\to\infty} \|A^N - A\|_{\ell^2(\mathbb{Z})} = 0$. ◊

5.2.1. Three key results involving orthogonality. The important concept of *orthogonality* lets us generalize the familiar idea of two perpendicular vectors in the plane to infinite-dimensional inner-product spaces.

We define orthogonality here in the spaces $L^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$. See the Appendix for the definitions in a general inner-product vector space. The statements and proofs we give here do generalize to that context without much extra effort. It may take a little while to get used to thinking of functions as vectors, however.

Definition 5.9. Two functions f, $g \in L^2(\mathbb{T})$ are *orthogonal*, written $f \perp g$, if their inner product is zero: $\langle f, g\rangle = 0$. A collection of functions $A \subset L^2(\mathbb{T})$ is *orthogonal* if $f \perp g$ for all f, $g \in A$ with $f \neq g$. Two subsets A, B of $L^2(\mathbb{T})$ are *orthogonal*, written $A \perp B$, if $f \perp g$ for all $f \in A$ and all $g \in B$. ♢

The trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are *orthonormal* in $L^2(\mathbb{T})$. Orthonormal means that they are an orthogonal set and each element in the set has L^2 norm equal to one (the vectors have been *normalized*). Checking orthonormality of the trigonometric system is equivalent to verifying that

$$\langle e_k, e_m\rangle = \delta_{k,m} \qquad \text{for } k,\ m \in \mathbb{Z}, \tag{5.3}$$

where $\delta_{k,m}$ is the Kronecker delta (equation (1.7)).

The Pythagorean Theorem, the Cauchy–Schwarz Inequality, and the Triangle Inequality hold in every inner-product vector space over $\mathbb{C}$. We state them explicitly for the case of $L^2(\mathbb{T})$.

Theorem 5.10 (Pythagorean Theorem)**.** *If f, $g \in L^2(\mathbb{T})$ are orthogonal, then*

$$\|f+g\|^2_{L^2(\mathbb{T})} = \|f\|^2_{L^2(\mathbb{T})} + \|g\|^2_{L^2(\mathbb{T})}.$$

Theorem 5.11 (Cauchy–Schwarz Inequality)**.** *For all f, $g \in L^2(\mathbb{T})$,*

$$|\langle f, g\rangle| \leq \|f\|_{L^2(\mathbb{T})}\|g\|_{L^2(\mathbb{T})}.$$

Theorem 5.12 (Triangle Inequality)**.** *For all f, $g \in L^2(\mathbb{T})$,*

$$\|f+g\|_{L^2(\mathbb{T})} \leq \|f\|_{L^2(\mathbb{T})} + \|g\|_{L^2(\mathbb{T})}.$$

Exercise 5.13. Show that $\|f\|_{L^2(\mathbb{T})} := \sqrt{\langle f, f\rangle}$ is a norm on $L^2(\mathbb{T})$ (see the definition of norm in the Appendix). Prove Theorems 5.10, 5.11, and 5.12. ♢

Exercise 5.14. Deduce from the Cauchy–Schwarz Inequality that if f and g are in $L^2(\mathbb{T})$, then $|\langle f, g\rangle| < \infty$. ♢

An important consequence of the Cauchy–Schwarz Inequality is that we can interchange limits in the L^2 norm and inner products (or in language that we will meet later in the book, that the inner product is *continuous* in $L^2(\mathbb{T})$).

Proposition 5.15. *Fix $g \in L^2(\mathbb{T})$. If a sequence $\{f_n\}_{n\in\mathbb{N}}$ of L^2 functions converges to f in the L^2 sense, then*

$$\lim_{n\to\infty} \langle g, f_n \rangle = \langle g, \lim_{n\to\infty} f_n \rangle = \langle g, f \rangle.$$

Proof. Suppose g, $f_n \in L^2(\mathbb{T})$ and $\|f_n - f\|_{L^2(\mathbb{T})} \to 0$ as $n \to \infty$. Since $L^2(\mathbb{T})$ is complete, it follows that $f \in L^2(\mathbb{T})$. By the Cauchy–Schwarz Inequality,

$$|\langle g, f_n - f \rangle| \leq \|g\|_{L^2(\mathbb{T})} \|f_n - f\|_{L^2(\mathbb{T})}.$$

We conclude that $\lim_{n\to\infty} \langle g, f_n - f \rangle = 0$, and since the inner product is linear in the second variable, the result follows. □

It follows that we can also interchange the inner product with a convergent infinite sum in $L^2(\mathbb{T})$.

Exercise 5.16. Fix $g \in L^2(\mathbb{T})$. Suppose that $f, h_k \in L^2(\mathbb{T})$ for $k \in \mathbb{Z}$, and that $f = \sum_{k=1}^{\infty} h_k$ in the L^2 sense. Show that $\sum_{k=1}^{\infty} \langle g, h_k \rangle = \langle g, \sum_{k=1}^{\infty} h_k \rangle$. **Hint:** Set $f_n = \sum_{k=1}^{n} h_k$ in Proposition 5.15. ♢

Exercise 5.17. Let $\{e_n\}_{n\in\mathbb{Z}}$ be the set of trigonometric functions. Suppose that $\{a_n\}_{n\in\mathbb{Z}}$, $\{b_n\}_{n\in\mathbb{Z}}$ are sequences of complex numbers, $f = \sum_{n\in\mathbb{Z}} a_n e_n$, and $g = \sum_{n\in\mathbb{Z}} b_n e_n$, where the equalities are in the L^2 sense. Show that $\langle f, g \rangle = \sum_{n\in\mathbb{Z}} a_n \overline{b_n}$. In particular, show that $\|f\|^2_{L^2(\mathbb{T})} = \sum_{n\in\mathbb{Z}} |a_n|^2$. ♢

Exercise 5.18. Suppose $\{f_\lambda\}_{\lambda\in\Lambda}$ is an orthogonal family of nonzero functions in $L^2(\mathbb{T})$, with an arbitrary index set Λ. Show that the f_λ are *linearly independent*; in other words, for each finite subset of indices $\{\lambda_1, \ldots, \lambda_n\} \subset \Lambda$, $a_1 f_{\lambda_1} + \cdots + a_n f_{\lambda_n} = 0$ if and only if $a_1 = \cdots = a_n = 0$. ♢

5.2.2. Orthonormal bases. Orthogonality implies *linear independence* (see Exercise 5.18). If our geometric intuition is right, an orthogonal set is in some sense the most linearly independent set possible. In a finite-dimensional vector space of dimension N, if we find N linearly independent vectors, then we have found a basis of the space. Similarly, in an N-dimensional inner-product vector space, if we find N orthonormal vectors, then we have found an orthonormal basis.

The trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are an orthonormal set for $L^2(\mathbb{T})$. Since they are an infinite linearly independent set, the space $L^2(\mathbb{T})$ is infinite-dimensional. This fact alone is not yet enough to guarantee that there are no other functions in the space $L^2(\mathbb{T})$ orthogonal to the functions $\{e_n(\theta)\}_{n\in\mathbb{Z}}$, or in other words that the system of trigonometric functions is *complete*. It turns out that the completeness of the set of trigonometric functions in $L^2(\mathbb{T})$ is equivalent to the mean-square convergence of partial Fourier sums (Theorem 5.1) and also to Parseval's Identity (Theorem 5.2).

Definition 5.19. Let $\{f_n\}_{n\in\mathbb{N}}$ be an orthonormal family in $L^2(\mathbb{T})$. We say that the family $\{f_n\}_{n\in\mathbb{N}}$ is *complete*, or that $\{f_n\}_{n\in\mathbb{N}}$ is a *complete orthonormal system in* $L^2(\mathbb{T})$, or that the functions $\{f_n\}_{n\in\mathbb{N}}$ form an *orthonormal basis*, if each function $f \in L^2(\mathbb{T})$ can be expanded into a series of the basis elements that is convergent in the L^2 norm. That is, there is a sequence $\{a_n\}_{n\in\mathbb{N}}$ of complex numbers such that

$$\lim_{N\to\infty} \Big\| f - \sum_{n=1}^{N} a_n\, f_n \Big\|_{L^2(\mathbb{T})} = 0.$$

Equivalently, $f = \sum_{n=1}^{\infty} a_n f_n$ in the L^2 sense. ◊

The coefficients are uniquely determined by pairing the function f with the basis functions.

Lemma 5.20 (Uniqueness of the Coefficients)**.** *If $\{f_n\}_{n\in\mathbb{N}}$ is an orthonormal basis in $L^2(\mathbb{T})$ and if $f \in L^2(\mathbb{T})$, then there are complex numbers $\{a_n\}_{n\in\mathbb{N}}$ such that $f = \sum_{n=1}^{\infty} a_n f_n$, and the coefficients must be $a_n = \langle f, f_n\rangle$ for all $n \in \mathbb{N}$.*

Proof. Take the expansion of f in the basis elements. Pair it with f_k, use Exercise 5.16 to interchange the inner product and the sum, and use the orthonormality of the system to get

$$\langle f, f_k\rangle = \langle\sum_{n=1}^{\infty} a_n f_n, f_k\rangle = \sum_{n=1}^{\infty} a_n\langle f_n, f_k\rangle = a_k.$$

This proves the lemma. □

For the trigonometric system, more is true: not only is the system complete, but also the coefficients must be the Fourier coefficients. We state this result as a theorem, which will take us some time to prove.

Theorem 5.21. *The trigonometric system is a complete orthonormal system in $L^2(\mathbb{T})$. Hence if $f \in L^2(\mathbb{T})$, then*

$$f(\theta) = \sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{in\theta},$$

where the equality is in the L^2 sense.

It follows from Theorem 5.21 and Exercise 5.17 that the Fourier map $\mathcal{F} : L^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$, given by $\mathcal{F}(f) = \{\widehat{f}(n)\}_{n\in\mathbb{Z}}$, not only preserves norms (Parseval's Identity) but also preserves inner products.

5.3. Completeness of the trigonometric system

The trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are an orthonormal system in $L^2(\mathbb{T})$. Our goal is to prove Theorem 5.21, which says that the system is complete. We will prove that it is equivalent to showing that we have mean-square convergence of the partial Fourier sums $S_N f$ to f, that is, Theorem 5.1:

$$\|f - S_N f\|_{L^2(\mathbb{T})} \to 0 \qquad \text{as } N \to \infty, \text{ for } f \in L^2(\mathbb{T}).$$

More precisely, the Fourier coefficients of $f \in L^2(\mathbb{T})$ are given by the inner product of f against the corresponding trigonometric function,

$$\widehat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\overline{e^{in\theta}}\, d\theta = \langle f, e_n\rangle.$$

Hence the partial Fourier sums are given by

$$S_N f = \sum_{|n|\le N} \widehat{f}(n)\, e_n = \sum_{|n|\le N} \langle f, e_n\rangle\, e_n.$$

Now by Definition 5.19 and Lemma 5.20, to verify completeness of the trigonometric system, it suffices to check the mean-square convergence of the partial Fourier sums:

$$\lim_{N\to\infty} \Big\| f - \sum_{|n|\leq N} \langle f, e_n\rangle\, e_n \Big\|_{L^2(\mathbb{T})} = \lim_{N\to\infty} \|f - S_N f\|_{L^2(\mathbb{T})} = 0.$$

To prove Theorem 5.21 and hence Theorem 5.1, we first show that $S_N f$ is the trigonometric polynomial of degree N that best approximates f in the L^2 norm. In other words, $S_N f$ is the *orthogonal projection of f onto the subspace of trigonometric polynomials of degree less than or equal to N.* Next we show that the partial Fourier sums converge to f in $L^2(\mathbb{T})$ for continuous functions f. Finally, an approximation argument allows us to conclude the same for square-integrable functions.

5.3.1. $S_N f$ is the closest N^{th}-trigonometric polynomial to f. The first observation is that $(f - S_N f)$ is orthogonal to the subspace $\mathcal{P}_N(\mathbb{T})$ generated by $\{e_n\}_{|n|\leq N}$. That subspace is exactly the subspace of trigonometric polynomials of degree less than or equal to N, in other words, functions of the form $\sum_{|n|\leq N} c_n e_n$, for complex numbers c_n.

Lemma 5.22. *Given $f \in L^2(\mathbb{T})$, $(f - S_N f)$ is orthogonal to all trigonometric polynomials of degree less than or equal to N.*

Proof. It suffices to show that $(f - S_N f)$ is orthogonal to $\sum_{|n|\leq N} c_n e_n$ for all choices of complex numbers $\{c_n\}_{|n|\leq N}$.

First note that e_j is orthogonal to $(f - S_N f)$ for each j such that $|j| \leq N$. We deduce this by checking that the inner product vanishes:

$$\begin{aligned}\langle f - S_N f, e_j\rangle &= \langle f - \sum_{|n|\leq N} \widehat{f}(n) e_n, e_j\rangle \\ &= \langle f, e_j\rangle - \sum_{|n|\leq N} \widehat{f}(n)\langle e_n, e_j\rangle = \langle f, e_j\rangle - \widehat{f}(j) = 0,\end{aligned}$$

using the linearity of the inner product and the orthonormality of the trigonometric system.

Now plug in any linear combination $\sum_{|j|\leq N} c_j e_j$ instead of e_j, and use the fact that the inner product is conjugate linear in the second variable, and what we just showed, to get

$$\langle f - S_N f, \sum_{|j|\leq N} c_j e_j\rangle = \sum_{|j|\leq N} \overline{c_j}\, \langle f - S_N f, e_j\rangle = 0.$$

Thus $(f - S_N f) \perp \sum_{|j|\leq N} c_j e_j$ for every such linear combination. □

Since the partial sum $S_N f = \sum_{|n|\leq N} \widehat{f}\,(n) e_n$ is a trigonometric polynomial of degree N, by Lemma 5.22 it is orthogonal to $(f - S_N f)$. By the Pythagorean Theorem 5.10,

$$\|f\|^2_{L^2(\mathbb{T})} = \|f - S_N f\|^2_{L^2(\mathbb{T})} + \|S_N f\|^2_{L^2(\mathbb{T})}.$$

Lemma 5.23. *If $f \in L^2(\mathbb{T})$, then*

$$\|f\|^2_{L^2(\mathbb{T})} = \|f - S_N f\|^2_{L^2(\mathbb{T})} + \sum_{|n|\leq N} |\widehat{f}\,(n)|^2.$$

Proof. All we need to check is that $\|S_N f\|^2_{L^2(\mathbb{T})} = \sum_{|n|\leq N} |\widehat{f}\,(n)|^2$.

First, by a direct calculation,

$$\|S_N f\|^2_{L^2(\mathbb{T})} = \Big\| \sum_{|n|\leq N} \widehat{f}(n) e_n \Big\|^2_{L^2(\mathbb{T})} = \Big\langle \sum_{|n|\leq N} \widehat{f}(n) e_n, \sum_{|j|\leq N} \widehat{f}(j) e_j \Big\rangle.$$

Second, by linearity in the first variable and conjugate linearity in the second variable of the inner product and by the orthonormality of the trigonometric functions,

$$\begin{aligned} \|S_N f\|^2_{L^2(\mathbb{T})} &= \sum_{|n|\leq N} \sum_{|j|\leq N} \widehat{f}\,(n) \overline{\widehat{f}\,(j)} \langle e_n, e_j\rangle \\ &= \sum_{|n|\leq N} \widehat{f}\,(n) \overline{\widehat{f}\,(n)} = \sum_{|n|\leq N} |\widehat{f}\,(n)|^2, \end{aligned}$$

as required. Notice that this is exactly the Pythagorean theorem for $2N + 1$ orthogonal summands. □

Bessel's inequality[4] for trigonometric functions follows immediately.

Lemma 5.24 (Bessel's Inequality)**.** *For all $f \in L^2(\mathbb{T})$,*

$$\sum_{n\in\mathbb{Z}} |\widehat{f}\,(n)|^2 \leq \|f\|^2_{L^2(\mathbb{T})}. \tag{5.4}$$

[4]Named after the German mathematician Friedrich Wilhelm Bessel (1784–1846).

Equality (Parseval's Identity!) holds exactly when the orthonormal system is complete, which we have yet to prove.

The partial Fourier sum $S_N f$ is the *orthogonal projection* of f onto the subspace $\mathcal{P}_N(\mathbb{T})$ of trigonometric polynomials of degree less than or equal to N. In particular, among all trigonometric polynomials of degree less than or equal to N, the N^{th} partial Fourier sum $S_N f$ is the one that best approximates f in the L^2 norm. Here are the precise statement and proof of this assertion.

Lemma 5.25 (Best Approximation Lemma)**.** *Take $f \in L^2(\mathbb{T})$. Then for each $N \geq 0$ and for all trigonometric polynomials $P \in \mathcal{P}_N(\mathbb{T})$,*

$$\|f - S_N f\|_{L^2(\mathbb{T})} \leq \|f - P\|_{L^2(\mathbb{T})}.$$

Equality holds if and only if $P = S_N f$.

Proof. Let $P \in \mathcal{P}_N(\mathbb{T})$. Then there exist complex numbers c_n, $|n| \leq N$, such that $P = \sum_{|n|\leq N} c_n e_n$. Then

$$\begin{aligned} f - P &= f - \sum_{|n|\leq N} c_n e_n - S_N f + S_N f \\ &= \big(f - S_N f\big) + \Big(\sum_{|n|\leq N} \big(\widehat{f}(n) - c_n\big) e_n\Big). \end{aligned}$$

The terms in parentheses on the right-hand side are orthogonal to each other by Lemma 5.22. Hence, by the Pythagorean Theorem,

$$\|f - P\|^2_{L^2(\mathbb{T})} = \|f - S_N f\|^2_{L^2(\mathbb{T})} + \Big\| \sum_{|n|\leq N} \big(\widehat{f}(n) - c_n\big) e_n \Big\|^2_{L^2(\mathbb{T})}.$$

Therefore $\|f - P\|^2_{L^2(\mathbb{T})} \geq \|f - S_N f\|_{L^2(\mathbb{T})}$. Equality holds if and only if the second summand on the right-hand side is equal to zero, but by the argument in the proof of Lemma 5.23,

$$\Big\| \sum_{|n|\leq N} \big(\widehat{f}(n) - c_n\big) e_n \Big\|^2_{L^2(\mathbb{T})} = \sum_{|n|\leq N} \big|\widehat{f}(n) - c_n\big|^2.$$

The right-hand side is equal to zero if and only if $\widehat{f}(n) = c_n$ for all $|n| \leq N$. That is, $P = S_N f$ as required for the equality to hold. The lemma is proved. □

Aside 5.26. Every lemma in this subsection has a counterpart for any orthonormal system $X = \{\psi_n\}_{n\in\mathbb{N}}$ in $L^2(\mathbb{T})$, where the subspace $\mathcal{P}_N(\mathbb{T})$ of trigonometric polynomials of degree less than or equal to N is replaced by the subspace generated by the first N functions in the system, $\mathcal{W}_N = \{f \in L^2(\mathbb{T}) : f = \sum_{n=1}^{N} a_n\psi_n, a_n \in \mathbb{C}\}$.

In fact all the lemmas in Subsection 5.3.1 are valid for any complete inner-product space, any orthonormal system in the space, and any closed subspace spanned by a finite or countable subset of the given orthonormal system. In the case of a subspace spanned by a countable but not finite subset, one has to prove and use the analogues in this setting of Proposition 5.15 and its consequences (Exercises 5.16 and 5.17) to justify the interchange of infinite sums and the inner product. ♢

Exercise 5.27. Let X be an orthonormal system in $L^2(\mathbb{T})$, not necessarily the trigonometric system. State and prove the analogues for X of Lemma 5.22, Lemma 5.23, Bessel's Inequality (Lemma 5.24), and the Best Approximation Lemma (Lemma 5.25). ♢

5.3.2. Mean-square convergence for continuous functions.

It is natural to prove Theorem 5.1 first for nice functions, in our case continuous functions, known to be dense in $L^2(\mathbb{T})$, and then for general square-integrable functions via an approximation argument.

Proof of Theorem 5.1 for continuous functions. We are ready to prove that for *continuous* functions g,

$$\|g - S_N g\|_{L^2(\mathbb{T})} \to 0 \qquad \text{as } N \to \infty.$$

Fix $\varepsilon > 0$. By Fejér's Theorem, the Cesàro sums $\sigma_N g$ converge *uniformly* to g on $\mathbb{T}$. So there is a trigonometric polynomial of degree M, say, namely $P(\theta) = \sigma_M g(\theta)$, such that $|g(\theta) - P(\theta)| < \varepsilon$ for all $\theta \in \mathbb{T}$. Therefore the L^2 norm of the difference is bounded by ε as well:

$$\|g - P\|^2_{L^2(\mathbb{T})} = \frac{1}{2\pi}\int_{-\pi}^{\pi} |g(\theta) - P(\theta)|^2\, d\theta < \varepsilon^2.$$

Notice that if $M \le N$, then $P \in \mathcal{P}_N(\mathbb{T})$. We can now use the Best Approximation Lemma (Lemma 5.25) to conclude that

$$\|g - S_N g\|_{L^2(\mathbb{T})} \le \|g - P\|_{L^2(\mathbb{T})} < \varepsilon \qquad \text{for all } N \ge M.$$

Thus, $\lim_{N\to\infty} \|g - S_N g\|_{L^2(\mathbb{T})} = 0$, for continuous functions g. □

5.3.3. Mean-square convergence for L^2 functions. We can now prove mean-square convergence for arbitrary functions $f \in L^2(\mathbb{T})$ and verify the completeness of the trigonometric system in $L^2(\mathbb{T})$.

Proof of Theorem 5.1. We now know that for $g \in C(\mathbb{T})$

$$\lim_{N\to\infty} \|g - S_N g\|_{L^2(\mathbb{T})} = 0.$$

The continuous functions on $\mathbb{T}$ are *dense* in $L^2(\mathbb{T})$; see Theorem 2.75. This means that given a function $f \in L^2(\mathbb{T})$ and given $\varepsilon > 0$, there exists a continuous function g on $\mathbb{T}$ such that

$$\|f - g\|_{L^2(\mathbb{T})} \leq \varepsilon.$$

Since g is continuous, we can find a trigonometric polynomial P (for example $P = S_M g$ for M large enough) such that

$$\|g - P\|_{L^2(\mathbb{T})} \leq \varepsilon.$$

Therefore, by the Triangle Inequality,

$$\|f - P\|_{L^2(\mathbb{T})} \leq \|f - g\|_{L^2(\mathbb{T})} + \|g - P\|_{L^2(\mathbb{T})} \leq 2\varepsilon.$$

Let M be the degree of P. By Lemma 5.25, for all $N \geq M$ we have

$$\|f - S_N f\|_{L^2(\mathbb{T})} \leq \|f - P\|_{L^2(\mathbb{T})} \leq 2\varepsilon.$$

Lo and behold, $\lim_{N\to\infty} \|f - S_N f\|_{L^2(\mathbb{T})} = 0$, for all $f \in L^2(\mathbb{T})$. □

Notice that it suffices to know that there is a trigonometric polynomial ε-close in the L^2 norm to g. For example $P = \sigma_N g$ for N large enough would also work. For this argument we do not really need to know that $S_N g \to g$ in $L^2(\mathbb{T})$ for continuous functions g, because we have the Best Approximation Lemma at our disposal.

There are settings other than $L^2(\mathbb{T})$, for example $L^p(\mathbb{T})$ for $1 \leq p < 2$ or $2 < p < \infty$, where there is no inner-product structure and no Best Approximation Lemma. It is still true that

$$\lim_{N\to\infty} \|S_n f - f\|_{L^p(\mathbb{T})} = 0. \tag{5.5}$$

However a different proof is required. A similar argument to the one presented in Section 5.3.2 shows that equation (5.5) holds for continuous functions g. Given $f \in L^p(\mathbb{T})$, take $g \in C(\mathbb{T})$ that is ε-close to f in the L^p norm. Then estimate the L^p norm of $(f - S_N f)$ by adding and subtracting both g and $S_N g$ and using the Triangle Inequality for the L^p norm:

$$\|f - S_N f\|_{L^p(\mathbb{T})} \leq \|f - g\|_{L^p(\mathbb{T})} + \|g - S_N g\|_{L^p(\mathbb{T})} + \|S_N g - S_N f\|_{L^p(\mathbb{T})}.$$

The first term on the right-hand side is less than ε, while the second can be made less than ε for N large enough by Exercise 5.28. The only concern is the third term. However, the mapping that takes f to its N^{th} partial Fourier sum $S_N f$ is linear, meaning that $S_N(af + bh) = aS_N f + bS_N h$. Moreover, these mappings are *uniformly bounded* or *continuous* on $L^p(\mathbb{T})$, meaning that there is a constant $C > 0$ such that for all $f \in L^p(\mathbb{T})$ and all $N > 0$,

$$\|S_N f\|_{L^p(\mathbb{T})} \leq C\|f\|_{L^p(\mathbb{T})}. \tag{5.6}$$

We can now control the third term:

$$\|S_N g - S_N f\|_{L^p(\mathbb{T})} = \|S_N(g - f)\|_{L^p(\mathbb{T})} \leq C\|f - g\|_{L^p(\mathbb{T})} \leq C\varepsilon.$$

Altogether, we have shown that for N large enough we can make $\|f - S_N f\|_{L^p(\mathbb{T})}$ arbitrarily small, so we are done.

This argument is valid as long as the mappings S_N are uniformly bounded on $L^p(\mathbb{T})$. For $p = 2$, this boundedness follows from Bessel's Inequality; see Exercise 5.30. For $p \neq 2$, we will prove in Chapter 12 that it follows from the boundedness on $L^p(\mathbb{T})$ of the *periodic Hilbert transform*.

Exercise 5.28. Show that $\lim_{N\to\infty} \|S_n g - g\|_{L^p(\mathbb{T})} = 0$ for continuous functions $g \in C(\mathbb{T})$. ♢

Exercise 5.29. Show that $S_N(f + h) = S_N f + S_N h$ for integrable functions f, $h \in L^1(\mathbb{T})$. ♢

Exercise 5.30. Show that $\|S_N f\|_{L^2(\mathbb{T})} \leq \|f\|_{L^2(\mathbb{T})}$ for all $f \in L^2(\mathbb{T})$ and for all $N > 0$ and that there is some $f \in L^2(\mathbb{T})$ for which equality holds. We say that S_N is a bounded *operator* with constant 1. ♢

5.3.4. Parseval's Identity. Given the completeness of the trigonometric system, we can now prove *Parseval's Identity* (Theorem 5.2): for $f \in L^2(\mathbb{T})$, the square of the L^2 norm of f is the sum of the squares of the absolute values of its Fourier coefficients.

Proof of Parseval's Identity. By Lemma 5.23, for each $N \in \mathbb{N}$,

$$\|f\|^2_{L^2(\mathbb{T})} = \|f - S_N f\|^2_{L^2(\mathbb{T})} + \sum_{|n| \leq N} |\widehat{f}(n)|^2. \tag{5.7}$$

Thus the series on the right-hand side converges because all terms are nonnegative and the sum is bounded above by $\|f\|^2_{L^2(\mathbb{T})}$. Furthermore,

$$\sum_{n \in \mathbb{Z}} |\widehat{f}(n)|^2 \leq \|f\|^2_{L^2(\mathbb{T})}.$$

We have just shown that if $f \in L^2(\mathbb{T})$, then $\|f - S_N f\|_{L^2(\mathbb{T})} \to 0$ as $N \to \infty$. In fact,

$$\|f\|^2_{L^2(\mathbb{T})} = \lim_{N \to \infty} \sum_{|n| \leq N} |\widehat{f}(n)|^2 =: \sum_{n \in \mathbb{Z}} |\widehat{f}(n)|^2. \qquad \square$$

5.4. Equivalent conditions for completeness

How can we tell whether a given orthonormal family X in $L^2(\mathbb{T})$ is actually an orthonormal basis? One criterion for completeness is that the only function in $L^2(\mathbb{T})$ that is orthogonal to all the functions in X is the zero function. Another is that Plancherel's Identity[5], which is an infinite-dimensional Pythagorean Theorem, must hold. These ideas are summarized in Theorem 5.31, which holds for all complete inner-product vector spaces V over $\mathbb{C}$, not just for $L^2(\mathbb{T})$; see the Appendix. We will prove Theorem 5.31 for the particular case when X is the trigonometric system. Showing that the same proof extends to any orthonormal system in $L^2(\mathbb{T})$ is left as an exercise.

Theorem 5.31. *Let $X = \{\psi_n\}_{n \in \mathbb{N}}$ be an orthonormal family in $L^2(\mathbb{T})$. Then the following are equivalent:*

(i) (Completeness) *X is a complete system, thus an orthonormal basis of $L^2(\mathbb{T})$.*

(ii) ($X^\perp = 0$) *If $f \in L^2(\mathbb{T})$ and $f \perp X$, then $f = 0$ in $L^2(\mathbb{T})$.*

[5] Named after the Swiss mathematician Michel Plancherel (1885–1967).

(iii) (Plancherel's Identity) *For all* $f \in L^2(\mathbb{T})$,

$$\|f\|^2_{L^2(\mathbb{T})} = \sum_{n=1}^{\infty} |\langle f, \psi_n \rangle|^2.$$

Note that Parseval's Identity, namely $\|f\|^2_{L^2(\mathbb{T})} = \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2$ for all $f \in L^2(\mathbb{T})$, is the special case of Plancherel's Identity when X is the trigonometric system $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$.

Proof of Theorem 5.31 for the trigonometric system. For the case when X is the trigonometric system, we showed in Section 5.3.4 that completeness implies Parseval's Identity, in other words that (i) $\Rightarrow$ (iii). Here, however, we establish the circle of implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

We show that (i) $\Rightarrow$ (ii) for the particular case of X being the trigonometric system. If $f \in L^2(\mathbb{T})$ were orthogonal to all trigonometric functions, then necessarily $\widehat{f}(n) = \langle f, e_n \rangle = 0$ for all $n \in \mathbb{Z}$. In that case the partial Fourier sums $S_N f$ would be identically equal to zero for all $N \geq 0$. The completeness of the trigonometric system now implies that

$$\|f\|_{L^2(\mathbb{T})} = \lim_{N\to\infty} \|f - S_N f\|_{L^2(\mathbb{T})} = 0.$$

Therefore $f = 0$. In other words, the completeness of the trigonometric system implies that the only L^2 function orthogonal to all trigonometric functions is the zero function.

We show that in the case of the trigonometric system, (ii) $\Rightarrow$ (iii). We proceed by the contrapositive. (Notice that we are not assuming the trigonometric system to be complete, only orthonormal.) Assume Parseval's Identity does not hold for some function $g \in L^2(\mathbb{T})$. Then the following inequality (the strict case of Bessel's Inequality) holds:

$$\sum_{n\in\mathbb{Z}} |\langle g, e_n \rangle|^2 < \|g\|^2_{L^2(\mathbb{T})} < \infty.$$

Define a function $h := \sum_{n\in\mathbb{Z}} \langle g, e_n \rangle e_n$. Then $h \in L^2(\mathbb{T})$, $\langle h, e_k \rangle = \langle g, e_k \rangle$ by the continuity of the inner product (see Exercise 5.16), and $\|h\|^2_{L^2(\mathbb{T})} = \sum_{n\in\mathbb{Z}} |\langle g, e_n \rangle|^2$, by Exercise 5.17. Let $f := g - h$. First, $f \neq 0$ in the L^2 sense; otherwise Parseval's Identity would hold for g. Second, f is orthogonal to e_n for all $n \in \mathbb{Z}$, as we see by a calculation:

$\langle f, e_n\rangle = \langle g - h, e_n\rangle = \langle g, e_n\rangle - \langle h, e_n\rangle = 0$. If Parseval's Identity does not hold, then there is a nonzero function f orthogonal to all trigonometric functions.

To close the loop, we show that (iii) $\Rightarrow$ (i) for the particular case of X being the trigonometric system. We have shown that equation (5.7) holds for all $N > 0$, and so we can take the limit as $N \to \infty$ on both sides to get

$$\|f\|^2_{L^2(\mathbb{T})} = \lim_{N\to\infty} \|f - S_N f\|^2_{L^2(\mathbb{T})} + \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2.$$

Now use Parseval's Identity to cancel the left-hand side and the right-most summand. We conclude that for all $f \in L^2(\mathbb{T})$,

$$\lim_{N\to\infty} \|f - S_N f\|_{L^2(\mathbb{T})} = 0.$$

Thus X is a complete system, as required. □

Exercise 5.32. Prove Theorem 5.31 for an arbitrary orthonormal system X in $L^2(\mathbb{T})$. ♢

5.4.1. Orthogonal projection onto a closed subspace. We know from Lemma 5.25 that $S_N f$ is the best approximation to f in the subspace of trigonometric polynomials of degree N in the L^2 norm.

More is true. We first recall that a subspace $\mathcal{W}$ of $L^2(\mathbb{T})$ is called *closed* if every convergent sequence in $\mathcal{W}$ converges to a point in $\mathcal{W}$.

Theorem 5.33 (Orthogonal Projection). *Given any closed subspace $\mathcal{W}$ of the complete inner-product vector space $L^2(\mathbb{T})$ and given $f \in L^2(\mathbb{T})$, there exists a unique function $P_{\mathcal{W}} f \in \mathcal{W}$ that minimizes the distance in $L^2(\mathbb{T})$ to $\mathcal{W}$. That is,*

$$\|f - g\|_{L^2(\mathbb{T})} \geq \|f - P_{\mathcal{W}} f\|_{L^2(\mathbb{T})} \qquad \text{for all } g \in \mathcal{W}.$$

Furthermore, the vector $f - P_{\mathcal{W}} f$ is orthogonal to $P_{\mathcal{W}} f$.

If $\mathcal{B} = \{g_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{W}$, then

$$P_{\mathcal{W}} f = \sum_{n\in\mathbb{N}} \langle f, g_n\rangle g_n.$$

We proved Theorem 5.33 for the special case of subspaces $\mathcal{P}_N(\mathbb{T})$ of trigonometric polynomials of degree at most N. These subspaces

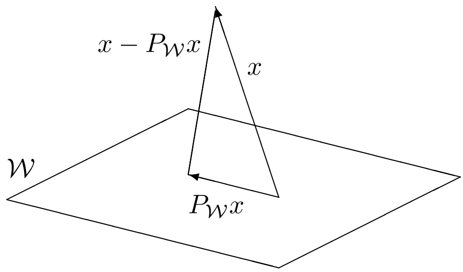


Figure 5.1. Orthogonal projection of a vector x onto a closed subspace $\mathcal{W}$ of a complete inner-product vector space V. Here V is represented by three-dimensional space, and $\mathcal{W}$ is represented by a plane. The orthogonal projection of x onto $\mathcal{W}$ is represented by the vector $P_{\mathcal{W}}x$. The distance from x to $\mathcal{W}$ is the length $\|x - P_{\mathcal{W}}x\|$ of the difference vector $x - P_{\mathcal{W}}x$.

are finite dimensional (of dimension $2N + 1$) and therefore automatically closed.

Theorem 5.33 says that we can draw pictures in $L^2(\mathbb{T})$ as we do in Euclidean space of a vector and its orthogonal projection onto a closed subspace (represented by a plane), and that the difference of these two vectors minimizes the distance to the subspace. See Figure 5.1.

We will use this orthogonal projection theorem for the Hilbert space $L^2(\mathbb{R})$ instead of $L^2(\mathbb{T})$, both in Section 9.4 when we discuss the Haar basis on $\mathbb{R}$ and in Chapters 10 and 11 when we talk about wavelets and multiresolution analyses.

Exercise 5.34. Prove Theorem 5.33 for a closed subspace $\mathcal{W}$ of $L^2(\mathbb{T})$, assuming that $\mathcal{W}$ has an orthonormal basis $\mathcal{B}_{\mathcal{W}} = \{g_n\}_{n\in\mathbb{N}}$. Then the elements of $\mathcal{W}$ can be characterized as (infinite) linear combinations of the basis elements g_n, with coefficients in $\ell^2(\mathbb{N})$, so that $\mathcal{W} = \{f \in L^2(\mathbb{T}) : f = \sum_{n\in\mathbb{N}} a_n g_n, \sum_{n\in\mathbb{N}} |a_n|^2 < \infty\}$. $\Diamond$

5.5. Project: The isoperimetric problem

A classical problem known as the *isoperimetric problem*, dating back to the ancient Greeks, is to find among all planar figures with fixed perimeter the one with the largest area. It is not hard to believe that the figure should be a circle. Steiner[6] found a geometric proof in 1841. In 1902, Hurwitz[7] published a short proof using Fourier series. An elegant direct proof that compares a smooth simple closed curve with an appropriate circle was given by Schmidt[8], in 1938. It uses only the arclength formula, the expression for the area of a plane region from Green's Theorem[9] and the Cauchy–Schwarz Inequality.

(a) Work through Steiner's geometric argument; see for example [**Kör**, Chapter 35].

(b) Work through Hurwitz's Fourier series argument. A delightful account can be found in [**Kör**, Chapter 36]. See also [**KL**, p. 151].

(c) Find a readable source for Schmidt's argument, and work through the proof.

(d) To further your knowledge on the isoperimetric problem and its applications to hydrodynamics, "in particular to problems concerning shapes of electrified droplets of perfectly conducting fluid", see [**BK**].

There is plenty of material online about the isoperimetric problem. See for instance the article [**Treib**] by Andrejs Treibergs, inspired by Treibergs' talk *My Favorite Proofs of the Isoperimetric Inequality* given in the Undergraduate Colloquium at the University of Utah on November 20, 2001. Jennifer Wiegert's article [**Wieg**] on the history of the isoperimetric problem was one of the winning articles in the 2006 competition for the best history-of-mathematics article by a student, sponsored by the History of Mathematics SIGMAA of the Mathematical Association of America (MAA). Viktor Blåsjö's article [**Blå**] on the same topic won the MAA's Lester R. Ford award for expository excellence. See also [**SS03**, Section 4.1].

[6]Swiss mathematician Jacob Steiner (1796–1863).

[7]German mathematician Adolf Hurwitz (1859–1919).

[8]German mathematician Erhard Schmidt (1876–1959).

[9]Named after British mathematical physicist George Green (1793–1841).

Chapter 6

A tour of discrete Fourier and Haar analysis

In this chapter we capitalize on the knowledge acquired while studying Fourier series to develop the simpler Fourier theory in a finite-dimensional setting.

In this context, the *discrete trigonometric basis* for $\mathbb{C}^N$ is a certain set of N orthonormal vectors, which necessarily form an orthonormal basis (Section 6.1). The *Discrete Fourier Transform* is the linear transformation that, for each vector v in $\mathbb{C}^N$, gives us the coefficients of v in the trigonometric basis (Section 6.3). Many of the features of the Fourier series theory are still present, but without the nuisances and challenges of being in an infinite-dimensional setting ($L^2(\mathbb{T})$), where one has to deal with integration and infinite sums. The tools required in the finite-dimensional setting are the tools of linear algebra which we review (Section 6.2). For practical purposes, this finite theory is what is needed, since computers deal with finite vectors. The Discrete Fourier Transform (DFT) of a given vector v in $\mathbb{C}^N$ is calculated by applying a certain invertible $N \times N$ matrix to v, and the Discrete Inverse Fourier Transform by applying the inverse matrix to the transformed vector (Section 6.3).

For an applied mathematician, an algorithm is only as good as how efficiently it can be executed. Reducing the number of operations required to perform an algorithm can drastically reduce the length of time required for execution, thus increasing efficiency. This is especially important for numerical applications where an increase in efficiency can make an algorithm practical. Surprisingly, one can perform each of the matrix multiplications in the DFT and its inverse in order $N \log_2 N$ operations, as opposed to the expected N^2 operations, by using the *Fast Fourier Transform*, or FFT (Section 6.4).

In this chapter we introduce another orthonormal basis, the *discrete Haar basis* for $\mathbb{C}^n$, in preparation for later chapters where the Haar bases for $L^2(\mathbb{R})$ and for $L^2([0,1])$ appear (Section 6.5). We highlight the similarities of the discrete trigonometric basis and the discrete Haar basis, and their differences, and discuss the Discrete Haar Transform (Section 6.6). There is also a *Fast Haar Transform* (FHT), of order N operations (Section 6.7). The Fast Haar Transform is a precursor of the *Fast Wavelet Transform* (FWT), studied in more depth in later chapters.

6.1. Fourier series vs. discrete Fourier basis

We recapitulate what we have learned about Fourier series and present the discrete Fourier basis in $\mathbb{C}^N$.

6.1.1. Fourier series: A summary. We begin by summarizing the ingredients of the theory of Fourier series, as we have developed it in Chapters 1–5.

(a) To each integrable function $f : \mathbb{T} \to \mathbb{C}$ we associate a sequence $\{\widehat{f}(n)\}_{n\in\mathbb{Z}}$ of complex numbers. These numbers are known as the *Fourier coefficients* of f and are given by the formula

$$\widehat{f}(n) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta \qquad \text{for } n \in \mathbb{Z}.$$

(b) The *building blocks* of the Fourier series are the *trigonometric functions* $e_n : \mathbb{T} \to \mathbb{C}$ given by $e_n(\theta) := e^{in\theta}$, where $n \in \mathbb{Z}$ and $\theta \in \mathbb{T}$.

They are called *trigonometric* rather than *exponential* to emphasize the connection with cosines and sines given by Euler's famous formula $e^{in\theta} = \cos n\theta + i \sin n\theta$.

Thus the n^{th} Fourier coefficient of f can be written in terms of the usual complex L^2 inner product as

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

(c) The Fourier series associated to f is the doubly infinite series (a series summed from $-\infty$ to ∞) formed by summing the products of the n^{th} Fourier coefficients $\widehat{f}(n)$ with the n^{th} trigonometric function $e^{in\theta}$:

$$f(\theta) \sim \sum_{n=-\infty}^{\infty} \widehat{f}(n) e^{in\theta} = \sum_{n=-\infty}^{\infty} \langle f, e_n \rangle\, e_n(\theta).$$

We use $\sim$ rather than $=$ in this formula in order to emphasize that for a given point θ, the Fourier series may not sum to the value $f(\theta)$, and indeed it may not converge at all.

(d) The set $\{e^{in\theta}\}_{n\in\mathbb{Z}}$ of trigonometric functions forms an *orthonormal basis* for the vector space $L^2(\mathbb{T})$ of square-integrable functions on $\mathbb{T}$. Here *orthonormal* means with respect to the $L^2(\mathbb{T})$ inner product. In particular this means that f equals its Fourier series in $L^2(\mathbb{T})$, or equivalently that the partial Fourier sums $S_N f$ converge in $L^2(\mathbb{T})$ to f, where $S_N f(\theta) = \sum_{|n|\leq N} \widehat{f}(n) e^{in\theta}$,

$$\lim_{N\to\infty} \|f - S_N f\|_{L^2(\mathbb{T})} = 0.$$

Also, the *energy*, or norm, of an $L^2(\mathbb{T})$ function coincides with the $\ell^2(\mathbb{Z})$ norm of its Fourier coefficients; that is, Parseval's Identity holds:

$$\|f\|^2_{L^2(\mathbb{T})} = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2.$$

Furthermore, the partial Fourier sum $S_N f$ is the orthogonal projection onto the closed subspace $\mathcal{P}_N(\mathbb{T})$ of trigonometric polynomials of degree less than or equal to N. In particular $S_N f$ is the best approximation in $L^2(\mathbb{T})$ to f by a polynomial in $\mathcal{P}_N(\mathbb{T})$.

6.1.2. The discrete Fourier basis. We present the analogues in the discrete setting of the four ingredients of Fourier analysis listed above. Fix a positive integer N, and let $\omega := e^{2\pi i/N}$. Recall that $\mathbb{Z}_N := \{0, 1, \ldots, N-1\}$. As usual, the superscript t indicates the transpose, so for example $v = [z_0, \ldots, z_{N-1}]^t$ is a column vector.

(a) To each *vector* $v = [z_0, \ldots, z_{N-1}]^t \in \mathbb{C}^N$ we associate a second *vector* $\widehat{v} = [a_0, \ldots, a_{N-1}]^t \in \mathbb{C}^N$. The vector $\widehat{v}$ is known as the *Discrete Fourier Transform* of v. The entries of the Discrete Fourier Transform are given by the formula

$$a_n = \widehat{v}(n) := \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} z_k \omega^{-kn} \qquad \text{for } n \in \mathbb{Z}_N.$$

(b) The *building blocks* of the Discrete Fourier Transform are the N discrete trigonometric functions $e_n : \mathbb{Z}_N \to \mathbb{C}$ given by

$$e_n(k) := \frac{1}{\sqrt{N}} \omega^{kn}, \qquad k \in \mathbb{Z}_N,$$

for $n \in \{0, 1, \ldots, N-1\}$. Thus the n^{th} Fourier coefficient of v can be written in terms of the $\mathbb{C}^N$ inner product as

$$a_n = \widehat{v}\,(n) = \sum_{k=0}^{N-1} z_k \overline{e_n(k)} =: \langle v, e_n \rangle.$$

(c) The original vector v can be exactly determined from its Discrete Fourier Transform. The k^{th} entry of v is given, for $k \in \mathbb{Z}_N$, by

$$v(k) := z_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \widehat{v}\,(n) \omega^{kn} = \sum_{n=0}^{N-1} \langle v, e_n \rangle e_n(k).$$

(d) The set $\{e_0, \ldots, e_{N-1}\}$ forms an *orthonormal basis* for the vector space $\mathbb{C}^N$ with respect to the standard inner product, or *dot product*, in $\mathbb{C}^N$, and the Pythagorean theorem holds:

$$\|v\|^2_{\ell^2(\mathbb{Z}_N)} = \|v\|^2_{\mathbb{C}^n} = \sum_{n=0}^{N-1} |\widehat{v}\,(n)|^2.$$

In this sense the norm of a vector v in $\mathbb{C}^N$ is the same as the norm of its Discrete Fourier Transform $\widehat{v}$.

Here we have referred to the Hilbert space $\mathbb{C}^N$ with the standard inner product as the space $\ell^2(\mathbb{Z}_N)$, to emphasize the similarities with the Hilbert space $L^2(\mathbb{T})$. The notation $\ell^2(\mathbb{Z}_N)$ means that we are viewing $\mathbb{C}^N$ as an inner-product vector space whose elements are complex-valued functions defined on the discrete set $\mathbb{Z}_N = \{0, 1, \ldots, N-1\}$, with the standard Euclidean inner product.

Note that as the dimension changes, the number ω and the vectors e_n also change. We should really label them with the dimension N, for example $\omega = \omega_N = e^{2\pi i/N}$, $e_n^N(k) = (1/\sqrt{N})\,\omega_N^{nk}$. However, for brevity we will omit the dependence on N and simply write ω, e_n and expect the reader to be aware of the underlying dimension.

We give two examples showing the Fourier orthonormal bases obtained for $\mathbb{C}^5$ and $\mathbb{C}^8$ using the scheme described above.

Example 6.1 (*Fourier Orthonormal Basis for* $\mathbb{C}^5$)**.** Let $N = 5$, so that $\omega = e^{2\pi i/5}$. We have the pairwise orthogonal vectors given by

$$f_0 = \begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}, \quad f_1 = \begin{bmatrix}1\\\omega\\\omega^2\\\omega^3\\\omega^4\end{bmatrix}, \quad f_2 = \begin{bmatrix}1\\\omega^2\\\omega^4\\\omega\\\omega^3\end{bmatrix}, \quad f_3 = \begin{bmatrix}1\\\omega^3\\\omega\\\omega^4\\\omega^2\end{bmatrix}, \quad f_4 = \begin{bmatrix}1\\\omega^4\\\omega^3\\\omega^2\\\omega\end{bmatrix}.$$

Notice the interesting arrangement of the powers of ω in each vector f_l. Dividing by the length $\sqrt{5}$ of each f_l, we obtain an orthonormal basis for $\mathbb{C}^5$, namely $e_0 = (1/\sqrt{5})f_0$, $e_1 = (1/\sqrt{5})f_1$, $e_2 = (1/\sqrt{5})f_2$, $e_3 = (1/\sqrt{5})f_3$, $e_4 = (1/\sqrt{5})f_4$. $\Diamond$

Example 6.2 (*Fourier Orthonormal Basis for* $\mathbb{C}^8$)**.** Let $N = 8$, so that $\omega = e^{2\pi i/8} = e^{\pi i/4} = (1+i)/\sqrt{2}$. We have pairwise orthogonal vectors $f_0, f_1, \ldots, f_7$ given by

$$\begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\\omega\\\omega^2\\\omega^3\\\omega^4\\\omega^5\\\omega^6\\\omega^7\end{bmatrix}, \begin{bmatrix}1\\\omega^2\\\omega^4\\\omega^6\\1\\\omega^2\\\omega^4\\\omega^6\end{bmatrix}, \begin{bmatrix}1\\\omega^3\\\omega^6\\\omega\\\omega^4\\\omega^7\\\omega^2\\\omega^5\end{bmatrix}, \begin{bmatrix}1\\\omega^4\\1\\\omega^4\\1\\\omega^4\\1\\\omega^4\end{bmatrix}, \begin{bmatrix}1\\\omega^5\\\omega^2\\\omega^7\\\omega^4\\\omega\\\omega^6\\\omega^3\end{bmatrix}, \begin{bmatrix}1\\\omega^6\\\omega^4\\\omega^2\\1\\\omega^6\\\omega^4\\\omega^2\end{bmatrix}, \begin{bmatrix}1\\\omega^7\\\omega^6\\\omega^5\\\omega^4\\\omega^3\\\omega^2\\\omega\end{bmatrix}.$$

Notice the patterns that arise because N is not prime. Dividing by the length $\sqrt{8}$ of each f_l, we obtain an orthonormal basis for $\mathbb{C}^8$, namely $e_0 = (1/\sqrt{8})f_0$, $e_1 = (1/\sqrt{8})f_1$, $\ldots$, $e_7 = (1/\sqrt{8})f_7$. $\Diamond$

Exercise 6.3. Verify that the set $\{e_0, e_1, \ldots, e_{N-1}\}$ is an orthonormal set in $\mathbb{C}^N$. **Hint:** The numbers $w_k = e^{2\pi ik/N}$, for $k = 0, 1, \ldots, N-1$, are the N^{th}-*roots of unity*, in other words, the solutions of the equation $z^N = 1$. Simple algebra shows that $z^N - 1 = (z-1)(z^{N-1} + \cdots + z + 1)$. Hence if $k \neq 0$, then $w_k^{N-1} + w_k^{N-2} + \cdots + w_k + 1 = 0$. $\Diamond$

We make four remarks. First, the discrete setting has a certain symmetry in that both the original signal and the transformed signal are of the same form: vectors in $\mathbb{C}^N$. In the Fourier series setting, the original signal is a square-integrable function $f : \mathbb{T} \to \mathbb{C}$, while the transformed signal is a doubly infinite sequence $\{\widehat{f}(n)\}_{n\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$. In Chapters 7 and 8 we will develop Fourier theory in yet another setting, that of the *Fourier transform*. There the original signal and the transformed signal can again be of the same form. For instance, they can both be functions in the Schwartz class $\mathcal{S}(\mathbb{R})$, or they can both be functions in $L^2(\mathbb{R})$. We will also discuss generalizations of the Fourier transform to much larger classes of input signals, such as functions in $L^p(\mathbb{R})$, and, more generally, tempered distributions.

Second, our signal vectors $v = [z_0, \ldots, z_{N-1}]^t \in \mathbb{C}^N$ in the discrete setting can be thought of as functions $v : \mathbb{Z}_N \to \mathbb{C}$, in the sense that for each number $n \in \mathbb{Z}_N := \{0, 1, \ldots, N-1\}$, the function v picks out a complex number $v(n) = z_n$. So

$$v = [v(0), v(1), \ldots, v(N-1)]^t.$$

This formulation of the input signal is closer to the way we presented the Fourier series setting. With this notation the Discrete Fourier Transform reads

$$\widehat{v}(m) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} v(n) e^{-2\pi inm/N}. \tag{6.1}$$

Third, in the discrete setting there are no difficulties with convergence of series, since all the sums are finite. Thus the subtleties discussed in Chapters 3, 4, and 5 are absent from the setting of the finite Fourier transform.

Fourth, in the discrete setting the identity known as *Parseval's Identity* or *Plancherel's Identity* holds:

$$\sum_{n=0}^{N-1} |v(n)|^2 = \sum_{n=0}^{N-1} |\widehat{v}(n)|^2.$$

This identity is exactly the Pythagorean Theorem for vectors with N entries. Rewritten in a form closer to the way we saw it in the Fourier series setting, Parseval's Identity reads

$$\|v\|_{\ell^2(\mathbb{Z}_N)} = \|\widehat{v}\|_{\ell^2(\mathbb{Z}_N)}.$$

Here the norm, known as the ℓ^2 norm, is the Euclidean norm induced by the complex inner product. Explicitly, it is given by the square root of the sum of the squares of the absolute values of the entries of the vector v, and thus by the square root of the inner product of v with its complex conjugate.

Exercise 6.4. Prove Parseval's Identity in the discrete setting. $\Diamond$

As in the Fourier series setting, Parseval's Identity says that the energy of the transformed signal is the same as that of the original signal. In the Fourier series setting, the *energy* means the L^2 norm of a function or the ℓ^2 norm of a doubly infinite series. In the discrete setting, the energy means the ℓ^2 norm of a vector. Parseval's Identity also holds in the Fourier transform setting. For example when the original and transformed signals, f and $\widehat{f}$, are both functions in $L^2(\mathbb{R})$, the energy is represented by the L^2 norm, and

$$\|f\|_{L^2(\mathbb{R})} = \|\widehat{f}\|_{L^2(\mathbb{R})}.$$

Aside 6.5. Some authors use the orthogonal basis $\{f_0, \dots, f_{N-1}\}$ of $\mathbb{C}^N$ defined by $f_l = \sqrt{N}\, e_l$, instead of the orthonormal basis $\{e_l\}_{l\in\mathbb{Z}_N}$. Then a factor of $1/N$ appears in the definition of the coefficients and also in Parseval's Identity (but not in the reconstruction formula). To mirror the Fourier series theory, one could work with the vectors $\{f_0, f_1, \dots, f_{N-1}\}$ and consider them to be normalized, by defining the following inner product for vectors v, $w \in \mathbb{C}^N$:

$$\langle v, w\rangle_{\text{nor}} := \frac{1}{N}\sum_{n=0}^{N-1} v(n)\overline{w(n)}. \tag{6.2}$$

We made an analogous convention in $L^2(\mathbb{T})$ when we included the factor $1/(2\pi)$ in the definition of the inner product. In our formulation of the discrete theory we prefer to use the standard inner product, or dot product, on $\mathbb{C}^N$, not the normalized one. Reader beware: it is wise to check which normalization a given author is using. ◊

Exercise 6.6. Show that formula (6.2) defines an inner product in $\mathbb{C}^N$. Consider the vectors $\{f_0, f_1, \dots, f_{N-1}\}$. Show that they are orthonormal with respect to the normalized inner product in $\mathbb{C}^N$. ◊

Exercise 6.7. Define the normalized discrete Fourier coefficients of $v \in \mathbb{C}^N$ to be $\widehat{v}^{\,\mathrm{nor}}(n) = \langle v, f_n\rangle_{\mathrm{nor}}$. Use the theory of orthonormal bases to justify the reconstruction formula $v = \sum_{n=0}^{N-1} \widehat{v}^{\,\mathrm{nor}}(n) f_n$. Verify that Parseval's Identity holds in this setting:

$$\sum_{n=0}^{N-1} |\widehat{v}^{\,\mathrm{nor}}(n)|^2 = \|\widehat{v}^{\,\mathrm{nor}}\|^2_{\ell^2(\mathbb{Z}_N)} = \|v\|^2_{\mathrm{nor}} = \frac{1}{N}\sum_{n=0}^{N-1} |v(n)|^2. \qquad ◊$$

6.2. Short digression on dual bases in $\mathbb{C}^N$

Before discussing the Discrete Fourier Transform in more detail, let us review some fundamental linear algebra facts about bases.

Definition 6.8. A *basis* for $\mathbb{C}^N$ is a set of N vectors $\{v_1, v_2, \dots, v_N\}$ such that each vector $v \in \mathbb{C}^N$ can be written as a unique linear combination of the elements of the basis: there are complex numbers $a_1, a_2, \dots, a_N$, uniquely determined by the vector v, such that

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_N v_N. \tag{6.3}$$ ◊

Aside 6.9. In this section we are counting N vectors, columns, or entries starting from 1 up to N. In other sections we are counting starting from 0 up to $N-1$, because it offers a notational advantage when dealing with trigonometric and Haar vectors. ◊

Let B denote the $N \times N$ matrix whose j^{th} column is the vector v_j. We write

$$B = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_N \\ | & & | \end{bmatrix},$$

where the vertical line segments are a reminder that the v_j are vectors, not scalars. With this notation, we have the following matrix representation of equation (6.3):

$$v = Ba, \quad \text{where } a = [a_1, a_2, \ldots, a_N]^t.$$

As usual, the superscript t indicates the transpose, so here a is a column vector.

Theorem 6.10. *The following statements are equivalent.*

(1) *A set of N vectors $\{v_1, v_2, \ldots, v_N\}$ is a basis for $\mathbb{C}^N$.*

(2) *The vectors $\{v_1, v_2, \ldots, v_N\}$ are linearly independent. That is, if $0 = a_1v_1 + a_2v_2 + \cdots + a_Nv_N$, then $a_1 = \cdots = a_N = 0$.*

(3) *The $N \times N$ matrix B whose columns are the vectors v_j is invertible (also called nonsingular). Therefore we can find the coefficients a_1, a_2, $\ldots$, a_N of a vector v with respect to the vectors $\{v_1, v_2, \ldots, v_N\}$ by applying the inverse matrix B^{-1} to v:*

$$a = B^{-1}v.$$

Exercise 6.11. Prove the equivalences in Theorem 6.10. ♢

Given N linearly independent vectors $\{v_1, v_2, \ldots, v_N\}$ in $\mathbb{C}^N$, in other words, a basis for $\mathbb{C}^N$, let B be the invertible matrix with columns v_j, and denote by w_i the complex conjugate of the i^{th} row vector of its inverse B^{-1}. Thus

$$B^{-1} = \begin{bmatrix} — & \overline{w_1} & — \\ & \vdots & \\ — & \overline{w_N} & — \end{bmatrix}.$$

The horizontal line segments emphasize that the $\overline{w_j}$ are vectors, not scalars. With this notation, we see by carrying out the matrix multiplication that

$$B^{-1}v = \begin{bmatrix} — & \overline{w_1} & — \\ & \vdots & \\ — & \overline{w_N} & — \end{bmatrix} \begin{bmatrix} v(1) \\ \vdots \\ v(N) \end{bmatrix} = \begin{bmatrix} \langle v, w_1 \rangle \\ \vdots \\ \langle v, w_N \rangle \end{bmatrix},$$

where $v = [v(1), \ldots, v(N)]^t$. We choose to put the complex conjugate on the rows of B^{-1}, so that when applied to a vector v, we get the inner products of v with the vectors w_i with no complex conjugate.

Exercise 6.12. Verify that the i^{th} entry of the vector $B^{-1}v$ is the (complex!) inner product between the vector v and the vector w_i. ♢

The coefficients of v in the basis $\{v_1, v_2, \dots, v_N\}$ are exactly the entries of the vector $B^{-1}v$. Therefore, not only are they uniquely determined, but we have an algorithm to find them:

$$a_j = \langle v, w_j \rangle,$$

where the conjugates of the vectors w_j are the rows of B^{-1}. Thus there is a set of uniquely determined *dual vectors* $\{w_1, w_2, \dots, w_N\}$ forming a *dual basis* to the basis $\{v_1, v_2, \dots, v_N\}$, so that *to compute the coefficients of a given vector v in the original basis, all we need to do is compute the inner products of v with the dual vectors.*

Exercise 6.13 (*Orthonormality Property of the Dual Basis*)**.** Show that every basis $\{v_1, v_2, \dots, v_N\}$ of $\mathbb{C}^N$, together with its dual basis $\{w_1, w_2, \dots, w_N\}$, satisfies the following *orthonormality* condition:

$$\langle v_k, w_j \rangle = \delta_{k,j}. \tag{6.4}$$

Furthermore show that $v = \sum_{j=1}^{N} \langle v, w_j \rangle v_j = \sum_{j=1}^{N} \langle v, v_j \rangle w_j$. ♢

Exercise 6.14 (*Uniqueness of the Dual Basis*)**.** Given a basis of $\mathbb{C}^N$, $\{v_1, v_2, \dots, v_N\}$, suppose there are vectors $\{w_1, w_2, \dots, w_N\}$ with the orthonormality property (6.4). Show that the vectors $\{w_1, w_2, \dots, w_N\}$ are the dual basis for $\{v_1, v_2, \dots, v_N\}$. ♢

We will encounter dual bases in an infinite-dimensional context when we discuss biorthogonal wavelets (Chapter 10).

In particular, if the basis $\{v_1, v_2, \dots, v_N\}$ is orthonormal, then the dual vectors coincide with the original basis vectors. In other words, an orthonormal basis is its own dual basis. Equivalently, on the matrix side, if the basis is orthonormal, then $B^{-1} = \overline{B^t}$.

Definition 6.15. An $N \times N$ matrix whose columns are orthonormal vectors in $\mathbb{C}^N$ is called a *unitary matrix.* ♢

Unitary matrices are always invertible. The inverse of a unitary matrix U is equal to the conjugate of the transpose of U; that is,

$$U^{-1} = \overline{U^t}.$$

Exercise 6.16. Let U be a unitary matrix. Show that U preserves the inner product in $\mathbb{C}^N$ and the $\ell^2(\mathbb{Z}_N)$ norm. In particular U is an *isometry*: (i) $\langle Uv, Uw\rangle = \langle v, w\rangle$ and (ii) $\|Uv\|_{\ell^2(\mathbb{Z}_N)} = \|v\|_{\ell^2(\mathbb{Z}_N)}$. ♢

6.3. The Discrete Fourier Transform and its inverse

It is not hard to see that the transformation that maps the vector $v \in \mathbb{C}^N$ to the vector $\widehat{v} \in \mathbb{C}^N$ is linear. We call this transformation the *Discrete Fourier Transform*, or the *Fourier transform* on $\ell^2\big(\mathbb{Z}_N\big)$.

Definition 6.17. The *Fourier matrix* F_N is the matrix with entries

$$F_N(m,n) = e^{2\pi i mn/N} \qquad \text{for } m,\, n \in \{0, 1, \dots, N-1\}.$$

We emphasize that the rows and columns of F_N are labeled from 0 to $N-1$ and not from 1 to N. We call the columns of F_N the *orthogonal Fourier vectors*, and we denote them by f_0, f_1, $\dots$, f_{N-1}. ♢

The Fourier matrix is *symmetric*, meaning that it coincides with its transpose: $F_N^t = F_N$.

Example 6.18 (*Fourier Matrix for* $\mathbb{C}^8$). Here is the Fourier matrix F_N in the case when $N = 8$, where $\omega = e^{2\pi i/8}$ and $\omega^8 = 1$:

$$F_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \omega^4 & \omega^5 & \omega^6 & \omega^7 \\ 1 & \omega^2 & \omega^4 & \omega^6 & 1 & \omega^2 & \omega^4 & \omega^6 \\ 1 & \omega^3 & \omega^6 & \omega & \omega^4 & \omega^7 & \omega^2 & \omega^5 \\ 1 & \omega^4 & 1 & \omega^4 & 1 & \omega^4 & 1 & \omega^4 \\ 1 & \omega^5 & \omega^2 & \omega^7 & \omega^4 & \omega & \omega^6 & \omega^3 \\ 1 & \omega^6 & \omega^4 & \omega^2 & 1 & \omega^6 & \omega^4 & \omega^2 \\ 1 & \omega^7 & \omega^6 & \omega^5 & \omega^4 & \omega^3 & \omega^2 & \omega \end{bmatrix}. \qquad ♢$$

The finite Fourier transform is completely determined by the $N \times N$ complex-valued matrix $(1/\sqrt{N})\overline{F_N}$. Specifically, the *Discrete Fourier Transform* $\widehat{v}$ of a vector v is given by

$$\widehat{v} = (1/\sqrt{N})\,\overline{F_N}\,v. \tag{6.5}$$

The matrix $(1/\sqrt{N})\,\overline{F_N}$ is *unitary*, because its columns are orthonormal vectors. Therefore its inverse is its complex conjugate,

$\left((1/\sqrt{N})\,\overline{F_N}\right)^{-1} = (1/\sqrt{N})\,F_N$. Thus the *Discrete Inverse Fourier Transform* can be written as

$$v = \frac{1}{\sqrt{N}}\,F_N\widehat{v}. \tag{6.6}$$

In this finite-dimensional context, we are making an orthogonal change of basis in $\mathbb{C}^N$, from the standard basis to the Fourier basis.

Exercise 6.19. Show that $(1/N)F_N\,\overline{F_N} = (1/N)\overline{F_N}\,F_N = I_N$, where I_N denotes the $N \times N$ identity matrix. ♢

Exercise 6.20. Show that if $v, w \in \mathbb{C}^N$, then $\langle \widehat{v}, \widehat{w}\rangle = \langle v, w\rangle$. ♢

Aside 6.21. When describing the efficiency of an algorithm (matrix multiplication in this chapter) by counting the number of significant operations in terms of some parameter (in this case the dimensions of the matrix), some operations count more than others. Because they take up an insignificant amount of computer time, the following operations are not counted: multiplication by 1 or 0, additions, and swapping positions (permutations). Only multiplication by a complex number that is neither 0 nor 1 counts as an operation. In particular multiplication by -1 counts as one operation. ♢

Both F_N and $\overline{F_N}$ are full matrices, in the sense that all their entries are nonzero. So applying one of them to a vector to compute the Discrete Fourier Transform, or the Discrete Inverse Fourier Transform, requires N^2 multiplications if the vector has no zero entries. They have an operation count of the order of N^2 operations. However, one can dramatically improve on this operation count by exploiting the hidden structure of the Fourier matrix, as we now show.

6.4. The Fast Fourier Transform (FFT)

In the 1960s, Cooley and Tukey[1] discovered such an algorithm. Their famous *Fast Fourier Transform* (FFT) algorithm, explained in this section, makes use of an ingenious recursive factorization of matrices to reduce the number of multiplications required from order N^2 to order $N\log_2 N$, where N is the length of the signal vector. This

[1] American applied mathematician James Cooley (born in 1926) and statistician John Wilder Tukey (1915–2000).

improvement revolutionized the field of digital signal processing. One measure of the influence of the FFT algorithm is that over the period 1945–1988, Cooley and Tukey's paper [**CT**] was the sixth most-cited paper in the Science Citation Index in the areas of mathematics, statistics, and computer science [**Hig**, p. 217].

It is interesting to note that in the early 1800s, Gauss[2] had found an algorithm similar to the FFT for computing the coefficients in finite Fourier series. At the time, both matrices and digital signal processing were completely unforeseen.

Mathematically, the Fast Fourier Transform is based on a factorization of the Fourier matrix into a product of *sparse* matrices, meaning matrices with many zero entries. We illustrate this principle in dimension $N = 4$ first.

Exercise 6.22. The conjugate $\overline{F_4}$ of the 4×4 Fourier matrix F_4 can be written as the product of three sparse matrices, as follows:

$$(6.7) \qquad \overline{F_4} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -i \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & i \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Use matrix multiplication to check formula (6.7). Can you spot two copies of $\overline{F_2}$ on the right-hand side? ♢

The process of the Fast Fourier Transform is easiest to explain, and to implement, when $N = 2^j$. (If N is not a power of two, one can simply add an appropriate number of zeros to the end of the signal vector to make N into a power of two, a step known as *zero-padding*.) It begins with a factorization that reduces $\overline{F_N}$ to two copies of $\overline{F_{N/2}}$. This reduction is applied to the smaller and smaller matrices $\overline{F_{N/2^\ell}}$ for a total of $j = \log_2 N$ steps, until the original matrix is written as a product of $2\log_2 N$ sparse matrices, half of which can be collapsed into a single permutation matrix. The total number of multiplications required to apply the Fourier transform to a vector is reduced from order N^2, for the original form of the matrix $\overline{F_N}$, to the much smaller order $N \log_2 N$, for the factorized version of $\overline{F_N}$.

[2]German mathematician Johann Carl Friedrich Gauss (1777–1855).

To understand how this reduction works at the level of individual entries of the transformed vector, we rearrange formula (6.1) for the Discrete Fourier Transform to express $\widehat{v}(m)$ as a linear combination of two sums A_m and B_m that have the same structure as the sum in $\widehat{v}(m)$, so we can do this process recursively, first by definition, $\widehat{v}(m) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} v(n) e^{\frac{-2\pi imn}{N}}$. Now separate the sum into the even and odd terms,

$$\begin{aligned} \widehat{v}(m) &= \sum_{n=0}^{\frac{N}{2}-1} \frac{v(2n)}{\sqrt{N}} e^{\frac{-2\pi imn}{N/2}} + e^{\frac{-2\pi im}{N}} \sum_{n=0}^{\frac{N}{2}-1} \frac{v(2n+1)}{\sqrt{N}} e^{\frac{-2\pi imn}{N/2}} \\ &=: A_m + e^{\frac{-\pi im}{N/2}} B_m . \end{aligned} \tag{6.8}$$

The key is to note that $\widehat{v}(m+N/2)$ can be written as another linear combination of the same two sums, namely

$$\begin{aligned} \widehat{v}\left(m+\frac{N}{2}\right) &= \sum_{n=0}^{\frac{N}{2}-1} \frac{v(2n)}{\sqrt{N}} e^{\frac{-2\pi imn}{N/2}} - e^{\frac{-2\pi im}{N}} \sum_{n=0}^{\frac{N}{2}-1} \frac{v(2n+1)}{\sqrt{N}} e^{\frac{-2\pi imn}{N/2}} \\ &= A_m - e^{\frac{-\pi im}{N/2}} B_m , \end{aligned} \tag{6.9}$$

where we have used the fact that $e^{\frac{-2\pi i N/2}{N}} = e^{-\pi i} = -1$.

By clever ordering of operations we can reuse A_m and B_m, reducing the total number of operations required. We use the language of matrices to describe more precisely what we mean by "clever ordering". We have factored the conjugate Fourier matrix $\overline{F_N}$, for N even, into a product of three $N \times N$ block matrices,

$$\overline{F_N} = D_1^N B_1^N S_1^N, \tag{6.10}$$

where the matrices on the right side of equation (6.10) are

$$D_1^N = \begin{bmatrix} I_{N/2} & D_{N/2} \\ I_{N/2} & -D_{N/2} \end{bmatrix}, \quad B_1^N = \begin{bmatrix} \overline{F_{N/2}} & 0_{N/2} \\ 0_{N/2} & \overline{F_{N/2}} \end{bmatrix}, \quad S_1^N = \begin{bmatrix} \text{Even}_{N/2} \\ \text{Odd}_{N/2} \end{bmatrix}.$$

Here I_M denotes the $M \times M$ identity matrix, F_M is the $M \times M$ Fourier matrix, 0_M denotes the $M \times M$ zero matrix, Even_M and Odd_M are the $M \times 2M$ matrices that select in order the even and odd entries, respectively, of the vector $(v(0), v(1), \ldots, v(2M-1))^t$, and D_M is a diagonal $M \times M$ matrix to be described below.

Exercise 6.23. Compare equation (6.10) with the factorization of $\overline{F_4}$ given in Exercise 6.22. In particular, in equation (6.7) identify the matrices I_2, D_2, $\overline{F_2}$, Even_2, and Odd_2. ◊

We now describe the diagonal $M \times M$ matrix D_M for general M. The matrix D_M has diagonal entries $e^{-m\pi i/M}$ for $m = 0, 1, \ldots, M-1$. For example, here are D_2 and D_4:

$$D_2 = \begin{bmatrix} e^0 & 0 \\ 0 & e^{-\pi i/2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix},$$

$$D_4 = \begin{bmatrix} e^0 & 0 & 0 & 0 \\ 0 & e^{-\pi i/4} & 0 & 0 \\ 0 & 0 & e^{-2\pi i/4} & 0 \\ 0 & 0 & 0 & e^{-3\pi i/4} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1-i}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & \frac{-1-i}{\sqrt{2}} \end{bmatrix}.$$

The appearance in D_2 of the fourth root of unity, in D_4 of the eighth root of unity, and so on, illuminates how the Fourier matrix F_N can contain the N^{th} root of unity, while $F_{N/2}$ contains only the $(N/2)^{\text{th}}$ root, and the other matrices in the recursion equation contain only real numbers.

The $2M \times 2M$ matrix S_1^{2M} selects the even-numbered entries (starting with the zeroth entry) of a given $2M$-vector and moves them to the top half of the vector, preserving their order, and it moves the odd-numbered entries of the $2M$-vector to the bottom half of the vector, preserving their order. This is a *permutation matrix.* Here is the 8×8 matrix S_1^8 acting on a vector $v = [v(0), v(1), \ldots, v(7)]$ in $\mathbb{C}^8$:

$$S_1^8 v = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v(0) \\ v(1) \\ v(2) \\ v(3) \\ v(4) \\ v(5) \\ v(6) \\ v(7) \end{bmatrix} = \begin{bmatrix} v(0) \\ v(2) \\ v(4) \\ v(6) \\ v(1) \\ v(3) \\ v(5) \\ v(7) \end{bmatrix}.$$

Exercise 6.24. Convince yourself that equations (6.8) and (6.9) translate into the matrix decomposition (6.10). ◊

Turning to the number of operations required in the Fast Fourier Transform, we see that the matrix D_1^N in the decomposition (6.10) requires only N multiplications. The matrix B_1^N ostensibly requires $2(N/2)^2 = N^2/2$ multiplications, but it has the conjugate Fourier matrix $\overline{F_{N/2}}$ in the diagonal, and if 4 divides N, we can use the recursion equation (6.10) again to replace each of the two $(N/2) \times (N/2)$ matrix $\overline{F_{N/2}}$ that are sitting in the diagonal of B_1^N by products of three sparser matrices, $\overline{F_{N/2}} = D_1^{N/2} B_1^{N/2} S_1^{N/2}$. Notice that the matrix $B_1^{N/2}$ is a block diagonal matrix with two copies of matrix $\overline{F_{N/4}}$ on its diagonal. If $N = 2^j$, we can apply the recursion equation to the $(N/4) \times (N/4)$ matrix $\overline{F_{N/4}}$ and keep on reducing in the same way until we cannot subdivide any further, when we reach the 2×2 matrix $\overline{F_2}$, after $j - 1 \sim \log_2 N$ steps.

Exercise 6.25. What three matrices result from applying the recursion equation (6.10) to $\overline{F_2}$? ♢

Exercise 6.26. Verify that for $\overline{F_N}$, after two recursive steps the matrix B_1^N is decomposed as the product of three $N \times N$ block matrices: $B_1^N = D_2^N B_2^N S_2^N$, defined as follows:

$$D_2^N = \begin{bmatrix} D_1^{N/2} & \\ & D_1^{N/2} \end{bmatrix} = \begin{bmatrix} I_{N/4} & D_{N/4} & & \\ I_{N/4} & -D_{N/4} & & \\ & & I_{N/4} & D_{N/4} \\ & & I_{N/4} & -D_{N/4} \end{bmatrix},$$

$$B_2^N = \begin{bmatrix} B_1^{N/2} & \\ & B_1^{N/2} \end{bmatrix} = \begin{bmatrix} \overline{F_{N/4}} & & & \\ & \overline{F_{N/4}} & & \\ & & \overline{F_{N/4}} & \\ & & & \overline{F_{N/4}} \end{bmatrix},$$

$$S_2^N = \begin{bmatrix} S_1^{N/2} & \\ & S_1^{N/2} \end{bmatrix}.$$

Here, blank spaces represent zero submatrices of appropriate sizes. ♢

Assuming that $N = 2^\ell K$, $K > 1$, so that we can iterate ℓ times, at stage ℓ of this recurrence we will obtain a decomposition of $\overline{F_N}$ as the product of $2\ell + 1$ matrices:

$$\overline{F_N} = D_1^N D_2^N \cdots D_\ell^N B_\ell^N S_\ell^N \cdots S_2^N S_1^N,$$

where the $N \times N$ matrices are defined recursively by

$$D_\ell^N = \begin{bmatrix} D_1^{N/2^{\ell-1}} & & \\ & \ddots & \\ & & D_1^{N/2^{\ell-1}} \end{bmatrix},$$

and similarly for B_ℓ^N and S_ℓ^N, for which we will place $2^{\ell-1}$ copies of the matrices $B_1^{N/2^{\ell-1}}$ or $S_1^{N/2^{\ell-1}}$ on the diagonal.

Exercise 6.27. Write down the decomposition of the conjugate Fourier matrix $\overline{F_8}$ into a product of five matrices, by using equation (6.10) twice. Check your answer by matrix multiplication. How many multiplication operations are required to multiply a vector by the matrix $\overline{F_8}$? How many are required to multiply a vector by the five matrices in your factorization? (Count only multiplications by matrix entries that are neither one nor zero.) ♢

Note that the $N/2^{\ell-1} \times N/2^{\ell-1}$ matrix $D_1^{N/2^{\ell-1}}$ is the 2×2 block matrix with $N/2^\ell \times N/2^\ell$ diagonal blocks described earlier. More precisely,

$$D_1^{N/2^{\ell-1}} = \begin{bmatrix} I_{N/2^\ell} & D_{N/2^\ell} \\ I_{N/2^\ell} & -D_{N/2^\ell} \end{bmatrix}.$$

Multiplying by the matrix $D_1^{N/2^{\ell-1}}$ requires only $N/2^{\ell-1}$ multiplications, and thus, for each $\ell \geq 1$ multiplying by the matrix D_ℓ^N requires only $N = 2^{\ell-1}N/2^{\ell-1}$ significant multiplications.

After two iterations, the first two matrices in the decomposition of F_N require only N multiplications each. If $N = 2^j$, then one can do this $j - 1 = \log_2 N - 1$ times, until the block size is 2×2. After $j - 1$ steps, the matrix $\overline{F_N}$ will be decomposed into the product of $j - 1$ matrices, each requiring only N multiplications, for a total of $N(\log_2 N - 1)$ multiplications, followed by the product of B_{j-1}^N, and the $j - 1$ *scrambling* or permutation matrices, namely the Odd/Even matrices. The matrix B_{j-1}^N has 2^{j-1} copies of $\overline{F_2}$ on its diagonal and requires N multiplications, and so do each of the $j - 1$ permutation matrices, bringing the total number of multiplications to no more than $2N \log N$ operations.

How can one guess the decomposition (6.7)? The following exercise attempts to give you an explanation.

Exercise 6.28. Write explicitly the output of $\overline{F_4}$ times the column vector $v = [v(0), v(1), v(2), v(3)]^t$. In the output you will see linear combinations (complex) of the sums and differences of the numbers $v(0)$, $v(1)$ and of the numbers $v(2)$, $v(3)$. Observe that those sums and differences are the outputs of the 2×2 identity matrix I_2 and the 2×2 matrix $\overline{F_2}$ times the column vectors $[v(0), v(1)]^t$ and $[v(2), v(3)]^t$, respectively. You should now be in a position to discover the decomposition (6.7) on your own. See Section 6.7 where a similar analysis is carried out to discover the Fast Haar Transform. ♢

6.4.1. Scrambling matrices: A numerical miracle. When the $j-1$ scrambling matrices, S_ℓ^N, for $\ell = 1, 2, \ldots, j-1$, are applied in the right order, they become a very simple operation which costs essentially nothing in terms of operation counts. A miracle is observed when describing the numbers $a \in \mathbb{N}$, $0 \le a \le N-1$, in binary notation: "the successive application of the Odd/Even matrices will simply put in place n the m^{th} entry, where m is found by reversing the digits in the binary decomposition of n". See [**Stra88**, pp. 192–193]. In computer science this step is called *bit reversal*.

Let us see the miracle in action when $N = 8 = 2^3$. Denote the successive scrambling matrices by

$$S_1^8 := \begin{bmatrix} \mathrm{Even}_4 \\ \mathrm{Odd}_4 \end{bmatrix}, \quad S_2^8 = \begin{bmatrix} S_1^4 & 0_4 \\ 0_4 & S_1^4 \end{bmatrix}.$$

We start with the numbers 0, 1, ..., 7 in binary notation and apply the matrix S_1^8, followed by the matrix S_2^8:

$$\begin{bmatrix} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{bmatrix} \overset{S_1^8}{\to} \begin{bmatrix} 000 \\ 010 \\ 100 \\ 110 \\ 001 \\ 011 \\ 101 \\ 111 \end{bmatrix} \overset{S_2^8}{\to} \begin{bmatrix} 000 \\ 100 \\ 010 \\ 110 \\ 001 \\ 101 \\ 011 \\ 111 \end{bmatrix}.$$

We see that the binary numbers in the right-hand column are obtained from the binary numbers in the left-hand column by reading the binary digits in reverse order.

Exercise 6.29. Can you justify the miracle for all $N = 2^j$? Try an argument by induction.The j^{th} miracle matrix M_j is the product of $j-1$ scrambling $N \times N$ matrices, $M_j = S^N_{j-1} S^N_{j-2} \dots S^N_1$. The scrambling matrix S^N_ℓ for $\ell < j$ is a block diagonal matrix with $2^{\ell-1}$ copies of the $N/2^{\ell-1} \times N/2^{\ell-1}$ matrix $S_1^{N/2^{\ell-1}}$ on the diagonal. $\Diamond$

Even without a miracle, the scrambling $N \times N$ matrices S^N_ℓ are very sparse. Only N entries of S^N_ℓ are nonzero, and in fact the nonzero entries are all ones. Applying each of the scrambling matrices requires only N operations, and applying $j-1$ such matrices ($N = 2^j$) requires at most $N(\log_2 N - 1)$ operations. This means that the operation count for the Fourier matrix using this decomposition is no more than $2N(\log_2 N)$ operations. So even without the miracle, the algorithm is still of order $N \log_2 N$.

6.4.2. Fast convolution. Given two vectors v, $w \in \mathbb{C}^N$, define their *discrete circular convolution* by

$$v * w(n) := \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} v(k) w(n-k), \quad n = 0, 1, \dots, N-1.$$

If $n < k$, then $-N \le n - k < 0$. Define $w(n-k) := w(N - (n-k))$, or, equivalently, think of w as being extended periodically over $\mathbb{Z}$ so that $w(n) = w(N+n)$ for all $n \in \mathbb{Z}$. In matrix notation, the linear transformation T_w that maps the vector v into the vector $v * w$ is given by the *circulant matrix*

$$\begin{bmatrix} w(0) & w(N-1) & w(N-2) & \dots & w(1) \\ w(1) & w(0) & w(N-1) & \dots & w(2) \\ w(2) & w(1) & w(0) & \dots & w(3) \\ \vdots & \vdots & \vdots & & \vdots \\ w(N-1) & w(N-2) & w(N-3) & \dots & w(0) \end{bmatrix}.$$

(The term circulant indicates that down each diagonal, the entries are constant.) A priori, multiplying a vector by this full matrix requires N^2 operations.

Exercise 6.30 (*Fast Convolution of Vectors*)**.** (a) Show that

$$\widehat{v * w}(m) = \widehat{v}(m)\,\widehat{w}(m).$$

(b) Describe a fast algorithm to compute the convolution of two vectors in order $N \log_2 N$ operations. ♢

Convolution with a fixed vector w is a linear transformation given by a circulant matrix, as described in the paragraph preceding Exercise 6.30. These are also the only *translation-invariant* or *shift-invariant* linear transformations, meaning that if we first shift a vector and then convolve, the output is the same as first convolving and then shifting. More precisely, denote by S_k the linear transformation that shifts a vector $v \in \mathbb{C}^N$ by k units, that is, $S_k v(n) := v(n-k)$. Here we are viewing the vector as extended periodically so that $v(n-k)$ is defined for all integers. Then $S_k T_w = T_w S_k$, where, as before, T_w denotes the linear transformation given by convolution with the vector w, $T_w v = v * w$. As the underlying dimension N changes, so does the shift transformation S_k, but for brevity we write S_k and not S_k^N.

Exercise 6.31. Write down the 8×8 shift matrix S_3 in $\mathbb{C}^8$. Verify that $S_3 = (S_1)^3$. Now describe the $N \times N$ shift matrix S_k in $\mathbb{C}^N$ for any $k \in \mathbb{Z}$. Show that $S_k = (S_1)^k$. ♢

Exercise 6.32. Show that $S_k T_w = T_w S_k$. Conversely, show that if a linear transformation T is shift invariant, meaning that $S_k T = T S_k$ for all $k \in \mathbb{Z}$, then there is a vector $w \in \mathbb{C}^N$ such that $Tv = v * w$. ♢

The book by Frazier [**Fra**] takes a linear algebra approach and discusses these ideas and much more in the first three chapters, before moving on to infinite-dimensional space. Strang and Nguyen's book [**SN**] also has a linear algebra perspective.

6.5. The discrete Haar basis

In this section we discuss the *discrete Haar basis*[3]. Unlike the discrete Fourier basis, the Haar basis is localized, a concept that we now describe.

[3] Named after the German mathematician Alfréd Haar (1885–1933).

By *localized* we mean that the vector is zero except for a few entries. All entries of the discrete trigonometric vectors are nonzero, and in fact each entry has the same absolute value $1/\sqrt{N}$. Thus the discrete trigonometric vectors are certainly not localized. In particular, since the Fourier coefficients of a vector v are given by the inner product v against a trigonometric vector e_j, all entries of v are involved in the computation.

An example of a basis which is as localized as possible is the *standard basis* in $\mathbb{C}^N$. The standard basis vectors have all but one entry equal to zero; the nonzero entry is 1. Denote by s_j the vector in the standard basis whose j^{th} entry is 1, for $j = 0, 1, \ldots, N-1$:

$$s_0 = [1, 0, \ldots, 0]^t, s_1 = [0, 1, 0, \ldots, 0]^t, \ldots, s_{N-1} = [0, \ldots, 0, 1]^t.$$

The standard basis changes with the dimension (we should tag the vectors $s_j = s_j^N$), but as with the Fourier basis, we will omit the reference to the underlying dimension and hope the reader does not get confused.

Exercise 6.33. Show that the Discrete Fourier Transform in $\mathbb{C}^N$ of the Fourier basis vector e_j is given by the standard basis vector s_j, that is, $\widehat{e_j} = s_j$, for $0 \leq j \leq N-1$. Start with the case $N = 4$. ♢

Although the Fourier basis is not localized at all, its Fourier transform is as localized as possible. We say the Fourier basis is *localized in frequency*, but not in space or time.

Exercise 6.34. Verify that the Discrete Fourier Transform of the standard basis vector s_j is the complex conjugate of the Fourier basis vector e_j, that is, $\widehat{s_j} = \overline{e_j}$. ♢

Although the standard basis is as localized as possible, its Fourier transform is not localized at all. We say the standard basis is *localized in space* or *time* but not in frequency.

Exercise 6.35. Let $\{v_0, v_1, \ldots, v_{N-1}\}$ be an orthonormal basis for $\mathbb{C}^N$. Show that the Discrete Fourier Transforms $\{\widehat{v_0}, \widehat{v_1}, \ldots, \widehat{v_{N-1}}\}$ form an orthonormal basis for $\mathbb{C}^N$. Exercise 6.20 may be helpful. ♢

These examples are incarnations of a general principle, the *Uncertainty Principle*, that we discuss in more depth in subsequent chapters. The Uncertainty Principle says that it is impossible for a vector to be simultaneously localized in space and frequency.

The *discrete Haar basis* is an intermediate basis between the discrete Fourier basis and the standard basis, in terms of its space and frequency localization. We begin with the example of the discrete Haar basis for $\mathbb{C}^8$, where $N = 2^3 = 8$.

Exercise 6.36. Define the vectors $\tilde{h}_n$ for $n = 0, 1, \ldots, 7$ by

$$\begin{aligned}
\tilde{h}_0 &:= [1,1,1,1,1,1,1,1]^t,\\
\tilde{h}_1 &:= [-1,-1,-1,-1,1,1,1,1]^t,\\
\tilde{h}_2 &:= [-1,-1,1,1,0,0,0,0]^t,\\
\tilde{h}_3 &:= [0,0,0,0,-1,-1,1,1]^t,\\
\tilde{h}_4 &:= [-1,1,0,0,0,0,0,0]^t,\\
\tilde{h}_5 &:= [0,0,-1,1,0,0,0,0]^t,\\
\tilde{h}_6 &:= [0,0,0,0,-1,1,0,0]^t,\\
\tilde{h}_7 &:= [0,0,0,0,0,0,-1,1]^t.
\end{aligned}$$

Notice the $\tilde{h}_2, \ldots, \tilde{h}_7$ can be seen as compressed and translated versions of $\tilde{h}_1$. (See also Exercise 6.44.) Show that the vectors $\{\tilde{h}_0, \tilde{h}_1, \ldots, \tilde{h}_7\}$ are orthogonal vectors in $\mathbb{C}^8$. Show that the (normalized) vectors

$$h_0 = \frac{\tilde{h}_0}{\sqrt{8}}, \quad h_1 = \frac{\tilde{h}_1}{\sqrt{8}}, \quad h_2 = \frac{\tilde{h}_2}{\sqrt{4}}, \quad h_3 = \frac{\tilde{h}_3}{\sqrt{4}},$$
$$h_4 = \frac{\tilde{h}_4}{\sqrt{2}}, \quad h_5 = \frac{\tilde{h}_5}{\sqrt{2}}, \quad h_6 = \frac{\tilde{h}_6}{\sqrt{2}}, \quad h_7 = \frac{\tilde{h}_7}{\sqrt{2}} \tag{6.11}$$

are orthonormal. They form the discrete Haar basis for $\mathbb{C}^8$. ◊

Exercise 6.37. Compute the Discrete Fourier Transform $\widehat{h_n}$ of each of the vectors in the Haar basis for $\mathbb{C}^8$, displayed in the formulas (6.11). Either analytically or using MATLAB, sketch the graphs of each h_n and $\widehat{h_n}$. Notice that the more localized h_n is, the less localized $\widehat{h_n}$ is. In other words, the more localized in space h_n is, the less localized it is in frequency. Compare the shapes of the vectors h_n for $n \geq 1$. ◊

When $N = 2^j$, $j \in \mathbb{N}$, the N Haar vectors are defined according to the analogous pattern. The first nonnormalized vector $\tilde{h}_0$ is a string of ones; the second vector $\tilde{h}_1$ has the first half of its entries equal to -1 and the second half equal to 1; the third vector $\tilde{h}_2$ has the first quarter of its entries equal to -1, the second quarter 1, and the rest of the entries zero; and so on. The last nonnormalized vector $\tilde{h}_{N-1}$ starts with $N-2$ zeros, and its last two entries are -1 and 1.

Exercise 6.38. Write down the discrete Haar basis for $\mathbb{C}^N = \mathbb{C}^{2^j}$ for some $j \geq 4$. ♢

Notice that half of the vectors are very localized, having all entries zero except for two consecutive nonzero entries. These are the 2^{j-1} vectors $\{\tilde{h}_{2^{j-1}}, \tilde{h}_{2^{j-1}+1}, \ldots, \tilde{h}_{2^j-1}\}$. The 2^{j-2} vectors $\{\tilde{h}_{2^{j-2}}, \tilde{h}_{2^{j-2}+1}, \ldots, \tilde{h}_{2^{j-1}-1}\}$ have all entries zero except for four consecutive nonzero entries. The 2^{j-k} vectors $\{\tilde{h}_{2^{j-k}}, \tilde{h}_{2^{j-k}+1}, \ldots, \tilde{h}_{2^{j-k+1}-1}\}$ have all entries zero except for 2^k consecutive nonzero entries. The vectors $\{\tilde{h}_2, \tilde{h}_3\}$ have half of the entries zero. Finally $\tilde{h}_0$ and $\tilde{h}_1$ have all entries nonzero. Furthermore the nonzero entries for $\tilde{h}_1$ consist of a string of consecutive -1's followed by the same number of 1's and for $\tilde{h}_0$ is just a string of 1's.

We now give a precise definition of the nonnormalized N^{th} Haar basis. We repeat: we should really label each of the N vectors with an N to indicate the dimension of $\mathbb{C}^N$. However, to make the notation less cluttered, we will not bother to do so.

Given $N = 2^j$ and n with $1 \leq n < N$, there are unique $k, m \in \mathbb{N}$ with $0 \leq k < j$ and $0 \leq m < 2^k$ such that $n = 2^k + m$. Note that for each k this describes $2^k \leq n < 2^{k+1}$.

Definition 6.39. For $1 \leq n \leq N-1$, define the *nonnormalized Haar vectors* $\tilde{h}_n$ in $\mathbb{C}^N$, with $N = 2^j$, in terms of the unique numbers k and m, $0 \leq k < j$, $0 \leq m < 2^k$, determined by n so that $n = 2^k + m$:

$$\tilde{h}_n(l) = \begin{cases} -1, & \text{if} \quad mN2^{-k} \leq l < (m + \frac{1}{2})N2^{-k}; \\ 1, & \text{if} \quad (m+\frac{1}{2})N2^{-k} \leq l < (m+1)N2^{-k}; \\ 0, & \text{otherwise.} \end{cases}$$

For $n = 0$, we define $\tilde{h}_0(l) = 1$ for all $0 \leq l \leq N-1$. ♢

Exercise 6.40. Verify that when $N = 8$ the formula above for $\tilde{h}_n(l)$ gives the same vectors as in Exercise 6.36. $\Diamond$

Notice that $\tilde{h}_n(l)$ is nonzero if and only if $m2^{j-k} \leq l < (m+1)2^{j-k}$. In fact those $2^{j-k} = N/2^k$ entries are ± 1.

Definition 6.41. The (*normalized*) *Haar vectors* are defined by

$$h_n = 2^{k/2} N^{-1/2} \tilde{h}_n, \quad 0 \leq n = 2^k + m \leq N-1,\ 0 \leq m < 2^k. \quad \Diamond$$

Exercise 6.42. Show that the vectors $\{h_n\}_{n=0}^{N-1}$ are orthonormal. It follows that N Haar vectors are linearly independent. Hence they form an orthonormal basis for $\mathbb{C}^N$. $\Diamond$

The parameters k, m, uniquely determined by n so that $n = 2^k + m$ are *scaling* and *translation* parameters, respectively. The scaling parameter k, $0 \leq k < j$, tells us that we will decompose the set $\{0, 1, 2, \ldots, 2^j - 1\}$ into 2^k disjoint subsets of $2^{j-k} = N2^{-k}$ consecutive numbers. The translation parameter m, $0 \leq m < 2^k$, indicates which subinterval of length $N2^{-k}$ we are considering, namely the interval $[mN2^{-k}, (m+1)N2^{-k})$.

Remark 6.43. It would make sense to index the n^{th} Haar vector with the pair (k, m), and we will do so when we study the Haar basis and some wavelet bases on $L^2(\mathbb{R})$. These are orthonormal bases, $\{\psi_{k,m}\}_{k,m\in\mathbb{Z}}$, indexed by two integer parameters (k, m). The Haar and wavelet bases have the unusual feature that they are found by dilating and translating one function ψ, known as the *wavelet* (or sometimes the *mother wavelet*). More precisely,

$$\psi_{k,m}(x) := 2^{k/2}\psi(2^k x - m).$$

The functions $\psi_{k,m}$ retain the shape of the wavelet ψ. If ψ is concentrated around zero on the interval $[-1, 1]$, then the function $\psi_{k,m}$ is now localized around the point $m2^{-k}$ and concentrated on the interval $[(m-1)2^{-k}, (m+1)2^{-k}]$ of length 2×2^{-k}. As $k \to \infty$, the *resolution* increases (that is, the localization improves), while as $k \to -\infty$, the localization gets worse. $\Diamond$

Exercise 6.44. Show that

$$h_{k,m}(l) = 2^{k/2} h_{0,0}(2^k l - mN).$$

Here $N = 2^j$ and $n = 2^k + m$, for $0 \leq k < j$, $0 \leq m < 2^k$. We write $h_{k,m} := h_n$. It is understood that $h_{0,0}(z) = h_1(z) = 0$ whenever z is not in $\{0, 1, \ldots, N-1\}$. We are padding the signal with zeros to the right and left of $h_{0,0}$, not extending it periodically. ♢

6.6. The Discrete Haar Transform

The *Discrete Haar Transform* is the linear transformation that applied to a vector $v \in \mathbb{C}^N$, $N = 2^j$, gives the coefficients of v in the orthonormal Haar basis. The *Inverse Discrete Haar Transform* is the linear transformation that reconstructs the vector v from its Haar coefficients $\{c_n\}_{n=1}^{N-1}$. Notice that

$$v = c_0 h_0 + c_1 h_1 + \cdots + c_{N-1} h_{N-1}.$$

In matrix notation the Inverse Discrete Haar Transform is given by

$$v = H_N\, c, \quad \text{where } c = [c_0, c_1, \ldots, c_{N-1}]^t,$$

and the *Haar matrix* H_N is the $N \times N$ matrix whose columns are the Haar vectors h_j. Both the vectors and the columns are numbered starting at $j = 0$ and ending at $j = N - 1$. That is,

$$H_N = \begin{bmatrix} | & & | \\ h_0 & \ldots & h_{N-1} \\ | & & | \end{bmatrix}.$$

The Haar matrix H_N is a unitary matrix (it has orthonormal columns); hence its inverse is its conjugate transpose. The Haar matrix H_N is also a real matrix (the entries are the real numbers $0, \pm 2^{(k-j)/2}$); hence $H_N^{-1} = H_N^t$. We can recover the coefficients $c \in \mathbb{C}^N$ of the vector $v \in \mathbb{C}^N$ by applying the transpose of H_N to the vector v. Therefore the *Discrete Haar Transform* is given by

$$c = H_N^t\, v.$$

Notice that either from the above matrix formula or from the fact that the Haar basis is an orthonormal basis, the j^{th} coefficient is calculated by taking the inner product of v with the j^{th} Haar vector:

$$c_j = \langle v, h_j \rangle.$$

Example 6.45. Here is the matrix H_8^t:

$$H_8^t = \begin{bmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}. \qquad \Diamond$$

The Haar matrix H_8^t is sparse; it has many zero entries. Counting the nonzero entries, we note that the first and second rows are both full. The third and fourth rows are half-full, so they make one full row together. The fifth, sixth, seventh, and eighth rows together make one full row of nonzero entries. Adding, we get $2+1+1=4$ full rows in H_8^t where $N=8=2^3$. In H_{16}^t we get the equivalent of four full rows from the first 8 rows as in H_8^t and the nineth to sixteenth would make up one more, i.e., $2+1+1+1=5$ full rows in H_{16}^t where $N=16=2^4$. The general formula is: if $N=2^j$, we get $j+1$ full columns, each of length N. So the total number of nonzero entries is $N(j+1)$ where $j=\log_2 N$. Hence, multiplying a vector of length N by H_N takes $N(1+\log_2 N)$ multiplications. This implies that the Discrete Haar Transform can be performed in order $N\log_2 N$ operations.

Here is an argument for seeking localized basis vectors: *The brute force operation count for the Haar basis is of the same order as the FFT, because of the sparseness of the Haar matrix.* Can this count be improved by using some smart algorithm as we did for the Discrete Fourier Transform? Yes, the operation count can be brought down to order N operations. We call such an algorithm the *Fast Haar Transform*, and it is the first example of the *Fast Wavelet Transform.*

6.7. The Fast Haar Transform

We illustrate with some examples how one can apply the Discrete Haar Transform more efficiently. We choose here to argue, as we did for the FFT, in terms of matrix decomposition. We will revisit this algorithm when we discuss the *multiresolution analysis* for wavelets.

First consider $N = 2$. Let us explicitly compute the action of the matrix H_2^t on a vector $v = [v(0), v(1)]^t \in \mathbb{C}^2$:

$$H_2^t v = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} v(0) \\ v(1) \end{bmatrix} = \begin{bmatrix} \frac{v(1)+v(0)}{\sqrt{2}} \\ \frac{v(1)-v(0)}{\sqrt{2}} \end{bmatrix}.$$

Notice that the output consists of scaled averages and differences of $v(0)$ and $v(1)$. In terms of multiplications, we only need two multiplications, namely $\frac{1}{\sqrt{2}}v(0)$ and $\frac{1}{\sqrt{2}}v(1)$; then we add these two numbers and subtract them. For operation counts, what is costly is the multiplications. *Applying H_2^t requires only 3 multiplications.*

Now consider the case $N = 4$. Let us explicitly compute the action of the matrix H_4^t on a vector $v = [v(0), v(1), v(2), v(3)]^t \in \mathbb{C}^4$:

$$H_4^t v = \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} v(0) \\ v(1) \\ v(2) \\ v(3) \end{bmatrix} = \begin{bmatrix} \frac{v(3)+v(2)+v(1)+v(0)}{\sqrt{4}} \\ \frac{\left(v(3)+v(2)\right)-\left(v(1)+v(0)\right)}{\sqrt{4}} \\ \frac{v(1)-v(0)}{\sqrt{2}} \\ \frac{v(3)-v(2)}{\sqrt{2}} \end{bmatrix}.$$

Denote the scaled averages of pairs of consecutive entries by a_m^1 and their differences by d_m^1:

$$a_0^1 := \frac{v(1)+v(0)}{\sqrt{2}}, \quad d_0^1 := \frac{v(1)-v(0)}{\sqrt{2}}.$$

$$a_1^1 := \frac{v(3)+v(2)}{\sqrt{2}}, \quad d_1^1 := \frac{v(3)-v(2)}{\sqrt{2}}.$$

Notice that they are the outputs of the 2×2 matrix H_2^t applied to the vectors $[v(0), v(1)]^t, [v(2), v(3)]^t \in \mathbb{C}^2$. With this notation, the output of the matrix $H_4^t v$ is the vector

$$H_4^t v = \begin{bmatrix} \frac{a_1^1+a_0^1}{\sqrt{2}} & \frac{a_1^1-a_0^1}{\sqrt{2}} & d_0^1 & d_1^1 \end{bmatrix}^t.$$

The first two entries are the output of the matrix H_2^t applied to the average vector $a^1 = [a_0^1, a_1^1]$. The last two entries are the output of the 2×2 identity matrix applied to the difference vector $d^1 = [d_0^1, d_1^1]$.

Aside 6.46. The superscript 1 is at this stage unnecessary. However we are computing averages and differences of the averages. We could and will define a vector $a^2 \in \mathbb{C}^4$ by $a^2_m = v(m)$ for $0 \le m \le 3$ (here 2 corresponds to the power j where $N = 2^j$, in our case, $N = 4$). With this convention the vector $a^1 \in \mathbb{C}^2$ is found averaging consecutive entries of a^2, and the vector $d^1 \in \mathbb{C}^2$ is found taking differences. We can introduce $a^0 = a^0_0 := \frac{a^1_1 + a^1_0}{\sqrt{2}}$, $d^0 = d^0_0 := \frac{a^1_1 - a^1_0}{\sqrt{2}}$. With this notation we have

$$H_4^t v = \begin{bmatrix} a^0 & d^0 & d^1 \end{bmatrix}^t .$$

Notice that the vectors d^0 and a^0 have dimension $1 = 2^0$ and d^1 has dimension $2 = 2^1$, so the output has dimension $4 = 2^2$. See Aside 6.49. ♢

It is now clear that we can decompose H_4^t as the product of two 4×4 matrices:

$$H_4^t = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} H_2^t & 0_2 \\ 0_2 & H_2^t \end{bmatrix}. \tag{6.12}$$

Notice that the first matrix can be decomposed as the product of a block diagonal matrix (with H_2^t and the 2×2 identity matrix I_2 in the diagonal) with a permutation matrix. In other words,

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} H_2^t & 0_2 \\ 0_2 & I_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The second matrix in equation (6.12) requires 2×2 multiplications, since it involves applying H_2^t twice. The first matrix requires only 3 multiplications, since it involves applying H_2^t once and then the 2×2 identity matrix which involves no multiplications. The total multiplication count for H_4^t is thus $2 \times 3 + 3 = 9 = 3(4 - 1)$.

Exercise 6.47. Show that H_8^t can be decomposed as the product of three 8×8 matrices as follows:

$$H_8^t = \begin{bmatrix} H_2^t & \\ & I_6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} H_4^t & \\ & H_4^t \end{bmatrix}.$$

Here I_6 is the 6×6 identity. Blank spaces denote zero matrices of the appropriate dimensions (in this case 2×6, 6×2, and 4×4). Notice that the last matrix on the right-hand side requires $2 \times 9 = 18$ multiplications, since it involves applying H_4^t twice. The first matrix on the right-hand side requires only 3 multiplications, since it involves applying H_2^t once and then the 6×6 identity matrix I_6 which involves no multiplications. The middle matrix is a permutation matrix, denoted by P_8. It involves no multiplications. The total multiplication count for H_8^t is thus $2 \times 9 + 3 = 21 = 3(8-1)$. ♢

We can see a pattern emerging. The matrix H_N^t for $N = 2^j$ can be decomposed as the product of three $N \times N$ matrices. The first and the last are block diagonal matrices while the middle matrix is a permutation matrix:

$$H_N^t = \begin{bmatrix} H_2^t & \\ & I_{N-2} \end{bmatrix} P_N \begin{bmatrix} H_{N/2}^t & \\ & H_{N/2}^t \end{bmatrix}. \tag{6.13}$$

Here I_{N-2} is the $(N-2) \times (N-2)$ identity, blank spaces denote zero matrices of the appropriate dimensions (in this case $2 \times (N-2)$, $(N-2) \times 2$, and $N/2 \times N/2$), and P_N is an $N \times N$ permutation matrix given by a permutation of the columns of I_N.

Denote by k_N the number of products required to compute multiplication by the matrix H_N^t. The last matrix in the decomposition (6.13) requires $2 \times k_{N/2}$ multiplications, since it involves applying $H_{N/2}^t$ twice, and $k_{N/2}$ is the number of multiplications required to compute multiplication by $H_{N/2}^t$. The first matrix requires only 3 multiplications, since it involves applying H_2^t once and then the

$(N-2) \times (N-2)$ identity matrix which involves no multiplications. The permutation matrix costs nothing in terms of multiplications. The total multiplication count for H_N^t obeys the recursive equation

$$k_N = 2 \times k_{N/2} + 3,$$

with initial value for the recurrence given by $k_2 = 3$.

Exercise 6.48. Show by induction that if the above decomposition holds for each $N = 2^j$, then the number of multiplications required to apply H_N^t is $3(N-1)$. That is, $k_N = 3(N-1)$. $\Diamond$

Aside 6.49. We return to the averages and differences, but now in dimension $N = 2^j$. Starting with the vector $a^j := v \in \mathbb{C}^N$, we create two vectors a^{j-1} and d^{j-1}, each of dimension $N/2 = 2^{j-1}$, by taking averages and differences of pairs of consecutive entries. We now take the vector $a^{j-1} \in \mathbb{C}^{N/2}$ and create two vectors of dimension $N/4 = 2^{j-2}$ by the same procedure. We repeat the process until we reach dimension $1 = 2^0$. The process can be repeated j times and can be represented by the following tree, or *herringbone*, algorithm:

$$\begin{array}{ccccccccccc} v := a^j & \to & a^{j-1} & \to & a^{j-2} & \to & \cdots & \to & a^1 & \to & a^0 \\ & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & \\ & & d^{j-1} & & d^{j-2} & & \cdots & & d^1 & & d^0 \,. \end{array}$$

With this notation it can be seen that

$$H_N^t v = \begin{bmatrix} a^0 & d^0 & d^1 & \ldots & d^{j-1} \end{bmatrix}^t.$$

Notice that the vectors a^{j-k} and d^{j-k} have dimension $2^{j-k} = N2^{-k}$, so the vector $\begin{bmatrix} a^0 & d^0 & d^1 & \ldots & d^{j-1} \end{bmatrix}$ has the correct dimension

$$1 + 1 + 2 + 2^2 + \cdots + 2^{j-1} = 2^j = N.$$

We will encounter averages and differences again when discussing wavelets in later chapters. $\Diamond$

Exercise 6.50. Give a precise description of the permutation matrix P_N in the decomposition of H_N^t. $\Diamond$

6.8. Project: Two discrete Hilbert transforms

In this project we introduce the finite discrete Hilbert transform and the sequential discrete Hilbert transform. These are discrete analogues of the continuous and periodic Hilbert transforms that will be studied in detail in Chapter 12. We define them on the space side and ask the reader to discover the corresponding definitions on the Fourier side. (We will do the opposite when we define the (continuous) Hilbert transform and its periodic analogue in Chapter 12.) These operators are convolution operators. In the finite case we mean the circular convolution introduced in Section 6.4.2, and in the sequential case we mean convolution of two sequences as defined below (see also Definition 11.13).

For vectors $x = (x_{-N}, \dots, x_{-1}, x_0, x_1, \dots, x_N)$ in $\mathbb{R}^{2N+1}$, the *finite Hilbert transform* H_N is defined by

$$(H_N x)(i) := \sum_{|j| \leq N, j \neq i} \frac{x_j}{i-j} \quad \text{for } |i| \leq N.$$

The infinite-dimensional analogue, known as the *sequential (or discrete) Hilbert transform* H^d, can be identified with the doubly infinite matrix defined by

$$H^d = \{h_{m,n}\}_{m,n \in \mathbb{Z}}, \quad h_{mn} := \begin{cases} \dfrac{1}{m-n}, & \text{if } m \neq n; \\ 0, & \text{if } m = n. \end{cases}$$

Its action on a doubly infinite sequence $x = \{x_n\}_{n \in \mathbb{Z}}$ is given by

$$(H^d x)(i) := \sum_{j \in \mathbb{Z}, j \neq i} \frac{x_j}{i-j} \quad \text{for } i \in \mathbb{Z}.$$

(a) Find some applications of the finite and sequential Hilbert transforms in the literature.

(b) Show that the finite Hilbert transform H_N is bounded on $\ell^2(\mathbb{Z}_{2N+1})$, with bounds independent of the dimension N, meaning that there is a constant $C > 0$ independent of N such that $\|H_N x\|_{\ell^2(\mathbb{Z}_{2N+1})} \leq C\|x\|_{\ell^2(\mathbb{Z}_{2N+1})}$ for all vectors $x \in \ell^2(\mathbb{Z}_{2N+1})$.

On the other hand, show that the $\ell^1(\mathbb{Z}_{2N+1})$ norms of H_N grow with the dimension, and find the rate of growth. Do some numerical experiments to validate your findings. Repeat the experiments for the ℓ^2 norm. What can you conclude?

(c) Show that $H_N x$ is given by circular convolution on $\mathbb{C}^{2N+1}$ with some vector $K_N \in \mathbb{C}^{2N+1}$. See Section 6.4.2. Calculate the Discrete Fourier Transform of the vector K_N in $\mathbb{C}^{2N+1}$. By hand or using MATLAB, plot the max-norm of the vector K_N as a function of N. What do you observe?

(d) We now turn to the sequential Hilbert transform. We want to show that $H^d : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$. Item (b) suggests that if we can view H^d as the limit of H_N as $N \to \infty$, then H^d must be a bounded operator in $\ell^2(\mathbb{Z})$, meaning that there is a constant $C > 0$ such that $\|H^d x\|_{\ell^2(\mathbb{Z})} \leq C\|x\|_{\ell^2(\mathbb{Z})}$ for all vectors $x \in \ell^2(\mathbb{Z})$. Try to make this idea more precise. See **[Graf08]**.

(e) We can view $L^2(\mathbb{T})$ as the frequency domain and $\ell^2(\mathbb{Z})$ as time or space. Taking advantage of the isometry we already know exists between these two spaces via the Fourier transform, we can define the *Fourier transform of a square-summable sequence* $\{x_j\}_{j\in\mathbb{Z}}$ to be the periodic function in $L^2(\mathbb{T})$ given by

$$\widehat{x}\,(\theta) := \sum_{j\in\mathbb{Z}} x_j e^{-ij\theta}.$$

Show that Plancherel's Identity holds: $\|x\|_{\ell^2(\mathbb{Z})} = \|\widehat{x}\|_{L^2(\mathbb{T})}$. Check that the convolution of two sequences in $\ell^2(\mathbb{Z})$ is well-defined by

$$x * y(i) = \sum_{j\in\mathbb{Z}} x(j)y(i-j).$$

Show also that $x * y \in \ell^2(\mathbb{Z})$ and that the Fourier transform of the convolution is the product of the Fourier transforms of the sequences x and y. That is, $(x * y)^\wedge(\theta) = \widehat{x}(\theta)\,\widehat{y}(\theta)$.

(f) Show that the sequential Hilbert transform H^d is given by convolution with the ℓ^2 sequence $k(m) = h_{m0}$ defined at the start of this project. Find the Fourier transform $\widehat{k}\,(\theta)$ of the sequence $\{k(m)\}_{m\in\mathbb{Z}}$, and check that it is a bounded periodic function. Use Plancherel's Identity to show that H^d is bounded on $\ell^2(\mathbb{Z})$. Compare the doubly infinite sequence $H^d x$ with the periodic function $-i\,\mathrm{sgn}(\theta)\widehat{\chi}(\theta)$, for $\theta \in \mathbb{T}$.

6.9. Project: Fourier analysis on finite groups

The goal of this project is to develop Fourier analysis on finite abelian groups G. We assume the reader has some familiarity with algebra concepts such as *finite abelian group*. If not, gaining that familiarity becomes part of the project. To achieve our goal, we must construct an appropriate Hilbert space of square-integrable functions defined on the group G. We need to develop the ideas of integration, inner product, and orthonormal bases. In particular we need some functions, the *characters* of G, to play the rôle of the trigonometric functions.

(a) Find a definition of a *finite abelian group* G and some examples of such groups. Make sure you are comfortable with the notion of the *order* of the group G, denoted by $\#G$. Consider multiplicative groups. See [**DM**, Chapter 4].

(b) Let $L^2(G) = \{f : G \to \mathbb{C}\}$. This is a finite-dimensional vector space over $\mathbb{C}$. For each $a \in G$ define a function $\delta_a : G \to \mathbb{C}$ by $\delta_a(x) = 0$ for all $x \neq a$ and $\delta_a(a) = 1$. The finite collection of functions $\{\delta_a\}_{a\in G}$ is a basis for $L^2(G)$. Show that if $f \in L^2(G)$, then $f(x) = \sum_{a\in G} f(a)\delta_a(x)$, for all $x \in G$.

Check that the functions $\{\delta_a\}_{a\in G}$ are linearly independent. Conclude that the dimension of $L^2(G)$ is equal to the order of the group G.

(c) For $f \in L^2(G)$ and $U \subset G$, define the integral over U of f to be $\int_U f := \sum_{a\in U} f(a)$. Calculate $\int_U \delta_a$. Verify that the integral so defined is linear and that if U_1 and U_2 are disjoint subsets of G, then $\int_{U_1\cup U_2} f = \int_{U_1} f + \int_{U_2} f$.

(d) Define a mapping $\langle\cdot,\cdot\rangle : L^2(G) \times L^2(G) \to \mathbb{C}$ with the "integral formula", $\langle f, g\rangle := \int_G f\,\overline{g} = \sum_{a\in G} f(a)\,\overline{g(a)}$. Verify that this mapping defines an inner product in $L^2(G)$. The induced norm is then given by $\|f\|_{L^2(G)} = \sqrt{\sum_{a\in G} f(a)\,\overline{f(a)}}$. Since $L^2(G)$ is finite dimensional, it is a complete inner-product vector space, and hence a Hilbert space. Verify that the set $\{\delta_a\}_{a\in G}$ is an orthonormal basis for $L^2(G)$.

(e) We now develop the theory of characters. Characters play a rôle in representation theory analogous to the rôle of trigonometric functions in classical Fourier analysis. A *character of* G is a group homomorphism $\chi : G \to \mathbb{C}$; thus for all a, $b \in G$, $\chi(ab) = \chi(a)\chi(b)$.

The set of characters is denoted $\widehat{G}$, and one can introduce a group structure by defining $(\chi_1\chi_2)(a) = \chi_1(a)\chi_2(a)$ for χ_1, $\chi_2 \in \widehat{G}$. Verify that this group operation is associative, that the identity element is χ_0, where $\chi_0(a) = 1$ for all $a \in G$, and that the inverse of a character χ is given by $\chi^{-1}(a) = \chi(a^{-1})$. We also have conjugation on $\widehat{G}$, given by $\overline{\chi}(a) = \overline{\chi(a)}$. Check that $\overline{\chi} = \chi^{-1}$. Verify that if G is a finite abelian group of order n, if $\chi, \chi_1, \chi_2 \in \widehat{G}$ and if $a, b \in G$, then

$\sum_{a\in G} \chi_1(a)\overline{\chi_2}(a) = 0$ if $\chi_1 \neq \chi_2$ and is equal to n if $\chi_1 = \chi_2$,

$\sum_{\chi\in\widehat{G}} \chi(a)\overline{\chi}(b) = 0$ if $a \neq b$ and is equal to n if $a = b$.

(f) Define the dual space of $L^2(G)$ to be $L^2(\widehat{G})$. The Fourier transform of $f \in L^2(G)$ is the function $\widehat{f} \in L^2(\widehat{G})$ defined pointwise, via inner product with the character, $\widehat{f}(\chi) = \langle f, \chi\rangle = \sum_{a\in G} f(a)\overline{\chi(a)}$. Verify that the Fourier transform is a linear operator from $L^2(G)$ into $L^2(\widehat{G})$. Let G be a finite abelian group of order n, i.e., $\#G = n$. Verify the following Fourier Inversion Formula and Plancherel's Identity in this context: if $f \in L^2(G)$, then $f = \frac{1}{n}\sum_{\chi\in\widehat{G}} \widehat{f}(\chi)\,\chi$ and $\|f\|_{L^2(G)} = \frac{1}{n}\|\widehat{f}\|_{L^2(\widehat{G})}$.

(g) Define the translation by $a \in G$ of the function $f \in L^2(G)$ by $\tau_a f(x) = f(xa^{-1})$ (we are considering a multiplicative group). Show that for all $\chi \in \widehat{G}$, $\widehat{\tau_a f}(\chi) = \overline{\chi(a)}\widehat{f}(\chi)$.

Can you define an appropriate notion of the convolution $f * g \in L^2(G)$ of f, $g \in L^2(G)$? Verify that your convolution is commutative and that $\widehat{f * g} = \widehat{f}\,\widehat{g}$.

Chapter 7

The Fourier transform in paradise

The key idea of harmonic analysis is to express a function or signal as a superposition of simpler functions that are well understood. We briefly summarize the two versions of this idea that we have already seen: the Fourier series, which expresses a periodic function in terms of trigonometric functions (Chapters 1–5), and the Discrete Fourier Transform (DFT), which expresses a vector in terms of the trigonometric vectors (Chapter 6). Then we embark on our study of the third type of Fourier analysis, namely the Fourier transform, which expresses a nonperiodic function as an integral of trigonometric functions (Chapters 7 and 8).

We will see that the Fourier integral formulas work for smooth functions all of whose derivatives decay faster than any polynomial increases; these functions form the *Schwartz class*[1] (Sections 7.2 and 7.3). In the Schwartz class the Fourier theory is perfect: we are in paradise (Sections 7.4 and 7.6). We will encounter convolution and approximations of the identity in paradise (Section 7.5), and to close the chapter, we will see that convolution of a Schwartz function with a suitable approximation of the identity converges in L^p (Section 7.7). We will see in Chapter 8 that there is also life beyond paradise, in

[1]Named after the French mathematician Laurent Schwartz (1915–2002). He was awarded the Fields Medal in 1950 for his work in distribution theory.

the sense that we have a very nice Fourier theory for larger classes of functions, such as $L^2(\mathbb{R})$ and even the class of generalized functions, or *tempered distributions*.

7.1. From Fourier series to Fourier integrals

As we have seen, traditional Fourier series express a periodic function as a sum of pure harmonics (sines and cosines):

$$f(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n) e^{2\pi i n x}. \tag{7.1}$$

In this chapter we use functions of period 1 instead of 2π. For aesthetic reasons we prefer to place the factor 2π in the exponent, where it will also be in the case of the Fourier transform on $\mathbb{R}$.

In the early 1800s, mathematics was revolutionized by Fourier's assertion that "every periodic function" could be expanded in such a series, where the coefficients (amplitudes) for each frequency n are calculated from f via the formula

$$\widehat{f}(n) := \int_0^1 f(x) e^{-2\pi i n x}\, dx \qquad \text{for } n \in \mathbb{Z}. \tag{7.2}$$

It took almost 150 years to resolve exactly what this assertion meant. In a remarkable paper **[Car]** that appeared in 1966, Lennart Carleson showed that for square-integrable functions on $[0,1]$ the Fourier partial sums converge pointwise a.e. As a consequence the same holds for continuous functions (this fact was unknown until then).

We discussed these facts in Chapters 1–5. Among other things, in Chapter 5 we established the following fact.

> *The exponential functions* $\{e^{2\pi i n x}\}_{n\in\mathbb{Z}}$ *form an orthonormal basis for* $L^2([0,1])$.

It follows immediately that equation (7.1) for the inverse Fourier transform holds in the L^2 sense when $f \in L^2([0,1])$.

We also discussed, in Chapter 6, the Discrete Fourier Transform on the finite-dimensional vector space $\mathbb{C}^N$. There the trigonometric vectors $\{e_l\}_{l=0}^{N-1}$, with k^{th} entry given by $e_l(k) = (1/\sqrt{N})\, e^{2\pi i k l/N}$, form an orthonormal basis for $\mathbb{C}^N$. Therefore the Discrete Inverse

Fourier Transform is given by

$$v(k) = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} \widehat{v}(l)\, e^{2\pi ikl/N},$$

and the l^{th} Fourier coefficient is given by the inner product in $\mathbb{C}^N$ of the vector $v \in \mathbb{C}^N$ and the vector e_l:

$$\widehat{v}(l) = \langle v, e_l \rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} v(k)\, e^{-2\pi ikl/N}.$$

In the nonperiodic setting, the *Fourier transform* $\widehat{f}$ of an integrable function $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$\widehat{f}\,(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i\xi x}\, dx. \tag{7.3}$$

The *inverse Fourier transform* $(g)^\vee$ of an integrable function $g : \mathbb{R} \to \mathbb{R}$ is defined by

$$(g)^\vee(x) = \int_{\mathbb{R}} g(\xi) e^{2\pi i\xi x}\, d\xi. \tag{7.4}$$

Note that $\widehat{f}$ is pronounced "f hat" and $(g)^\vee$ is pronounced "g check" or "g inverse hat". These formulas make sense for integrable functions. There is much to be said about for which integrable functions, and in what sense, the following *Fourier Inversion Formula* holds:

$$f(x) = \int_{\mathbb{R}} \widehat{f}\,(\xi) e^{2\pi i\xi x}\, d\xi, \tag{7.5}$$

or more concisely $(\widehat{f}\,)^\vee = f$.

Heuristically, we could arrive at the integral formulas (7.3) and (7.5) by calculating the Fourier series on larger and larger intervals, until we cover the whole line, so to speak. See Section 1.3.2. For nice enough functions the L-Fourier series converges uniformly to the function f on $[-L/2, L/2)$:

$$f(x) = \sum_{n\in\mathbb{Z}} a_L(n) e^{2\pi inx/L} \qquad \text{for all } x \in [-L/2, L/2), \tag{7.6}$$

where the L-Fourier coefficients are given by the formula

$$a_L(n) := \frac{1}{L} \int_{-L/2}^{L/2} f(y) e^{-2\pi iny/L}\, dy.$$

Let $\xi_n = n/L$, so that $\Delta\xi := \xi_{n+1} - \xi_n = 1/L$. Then we can rewrite the Fourier series in equation (7.6) as $\sum_{n\in\mathbb{Z}} F_L(\xi_n)\Delta\xi$, where $F_L(\xi) := e^{2\pi i\xi x} \int_{-L/2}^{L/2} f(y)e^{-2\pi i\xi y}\, dy$.

This expression returns $f(x)$ for $x \in [-L/2, L/2)$ and resembles a Riemann sum for the improper integral $\int_{-\infty}^{\infty} F_L(\xi)\, d\xi$, except that the parameter L appears in the function to be integrated, as well as in the partition of the real line in steps of length $1/L$. Now pretend that the function f is zero outside the interval $[-M, M]$, for some $M > 0$ (that is, the function f has *compact support*). For L sufficiently large that $M < L/2$, the function F_L is now a function independent of L:

$$F_L(\xi) = e^{2\pi i\xi x} \int_{-\infty}^{\infty} f(y)e^{-2\pi i\xi y}\, dy = e^{2\pi i\xi x}\, \widehat{f}(\xi),$$

where $\widehat{f}(\xi)$ is the Fourier transform of the function f defined on $\mathbb{R}$. Heuristically we would expect as $L \to \infty$ that the sum $\sum_{n\in\mathbb{Z}} F_L(\xi_n)\Delta\xi$ should go to $\int_{-\infty}^{\infty} e^{2\pi i\xi x}\, \widehat{f}(\xi)\, d\xi$ and that this integral returns the value of $f(x)$ for all $x \in \mathbb{R}$.

It could be argued that to recover a nonperiodic function on $\mathbb{R}$ we need *all* frequencies and that the following integrals play the same rôle that equations (7.1) and (7.2) play in the Fourier series theory:

$$f(x) = \int_{\mathbb{R}} \widehat{f}(\xi)e^{2\pi i\xi x}\, d\xi \quad \text{and} \quad \widehat{f}(\xi) := \int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}\, dx. \tag{7.7}$$

To make this argument rigorous, we should specify what "nice" means and stablish some type of convergence that entitles us to perform all these manipulations. We will not pursue this line of argument here, other than to motivate the appearance of the integral formulas.

7.2. The Schwartz class

It turns out that formulas (7.7) hold when f belongs to the *Schwartz space* $\mathcal{S}(\mathbb{R})$, also known as the *Schwartz class*. First, $C^\infty(\mathbb{R})$ denotes the space of infinitely differentiable functions $f : \mathbb{R} \to \mathbb{C}$. Next, the Schwartz space $\mathcal{S}(\mathbb{R})$ consists of those functions $f \in C^\infty(\mathbb{R})$ such that f and all its derivatives f', f'', ..., $f^{(\ell)}$, ... decrease faster than any polynomial increases. Rephrasing, we obtain the following definition.

Definition 7.1. The *Schwartz space* $\mathcal{S}(\mathbb{R})$ is the collection of C^∞ functions f such that they and their derivatives are rapidly decreasing, meaning that

$$\lim_{|x|\to\infty} |x|^k |f^{(\ell)}(x)| = 0 \quad \text{for all integers } k,\ \ell \geq 0. \tag{7.8}$$ ◊

This limiting property is equivalent to the statement that the products $|x|^k|f^{(\ell)}(x)|$ are bounded functions for all k and ℓ.

Exercise 7.2. Let $f \in C^\infty(\mathbb{R})$. Show that (7.8) holds if and only if

$$\sup_{x\in\mathbb{R}} |x|^k |f^{(\ell)}(x)| < \infty \quad \text{for all integers } k,\ \ell \geq 0. \tag{7.9}$$ ◊

The Schwartz space is a vector space over the complex numbers. We invite you to check that it is closed under the operations of multiplication by Schwartz functions, multiplication by polynomials, multiplication by trigonometric functions, and differentiation. Note that neither polynomials nor trigonometric functions are Schwartz functions; they are infinitely differentiable but do not decay at infinity.

Exercise 7.3 ($\mathcal{S}(\mathbb{R})$ *Is Closed under Multiplication and Differentiation*)**.** Show that if f, $g \in \mathcal{S}(\mathbb{R})$, then the products fg, $x^k f(x)$ for all $k \geq 0$, and $e^{-2\pi i x\xi} f(x)$ for all $\xi \in \mathbb{R}$ belong to $\mathcal{S}(\mathbb{R})$. Show also that the derivatives $f^{(\ell)}$ belong to $\mathcal{S}(\mathbb{R})$, for all $\ell \geq 0$. ◊

Our first examples of Schwartz functions are the compactly supported functions in $C^\infty(\mathbb{R})$.

Definition 7.4. A function f defined on $\mathbb{R}$ is *compactly supported* if there is a closed interval $[a, b] \subset \mathbb{R}$ such that $f(x) = 0$ for all $x \notin [a, b]$. We say that such a function f *has compact support* and informally that f *lives on the interval* $[a, b]$. ◊

Example 7.5 (*Compactly Supported* $C^\infty(\mathbb{R})$ *Functions Are Schwartz*)**.** If $f \in C^\infty(\mathbb{R})$ is compactly supported, then a fortiori f and all its derivatives are rapidly decreasing, and so $f \in \mathcal{S}(\mathbb{R})$. ◊

Compactly supported C^∞ functions are examples of Schwartz functions, but can we find a compactly supported C^∞ function that is not identically equal to zero? Yes, we can.

Example 7.6 (*A Bump Function*)**.** Here is an example of a *compactly supported Schwartz function* that is supported on the interval $[a,b]$ and nonzero on the interval (a,b):

$$B(x) = \begin{cases} e^{-1/(x-a)}e^{-1/(b-x)}, & \text{if } a < x < b; \\ 0, & \text{otherwise.} \end{cases} \qquad \Diamond$$

Exercise 7.7. Show that the bump function in Example 7.6 is in the Schwartz class. Notice that it suffices to check that $B(x)$ is C^∞ at $x = a$ and at $x = b$. $\Diamond$

Here is an example of a function that is in the Schwartz class but is not compactly supported.

Example 7.8 (*The Gaussian Function*[2])**.** The canonical example of a function in $\mathcal{S}(\mathbb{R})$ is the *Gaussian function* $G(x)$, defined by

$$G(x) := e^{-\pi x^2}. \qquad \Diamond$$

We define the integral over $\mathbb{R}$ of a Schwartz function f to be the limit of the integrals of f over larger and larger intervals (notice that those integrals can be defined either in the Riemann or in the Lebesgue sense, with the same result):

$$\int_{\mathbb{R}} f(x)\,dx := \lim_{T\to\infty} \int_{-T}^{T} f(x)\,dx. \tag{7.10}$$

Exercise 7.9. Show that the limit of $\int_{-T}^{T} f(x)\,dx$ as $T \to \infty$ is finite, for functions $f \in \mathcal{S}(\mathbb{R})$. $\Diamond$

Definition 7.10. The *Fourier transform* $\widehat{f} : \mathbb{R} \to \mathbb{C}$ of a Schwartz function f for $\xi \in \mathbb{R}$ is defined as follows:

$$\widehat{f}(\xi) := \int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}\,dx = \lim_{T\to\infty} \int_{-T}^{T} f(x)e^{-2\pi i\xi x}\,dx. \qquad \Diamond \tag{7.11}$$

Notice that the integrand is in $\mathcal{S}(\mathbb{R})$ (see Exercise 7.3).

[2]Named after the German mathematician Johann Carl Friedrich Gauss (1777–1855).

The Gaussian function plays a very important rôle in Fourier analysis, probability theory, and physics. It has the unusual property of being equal to its own Fourier transform:

$$\widehat{G}(\xi) = e^{-\pi\xi^2} = G(\xi).$$

One can prove this fact using the observation that both the Gaussian and its Fourier transform satisfy the ordinary differential equation $f'(x) = -2\pi x f(x)$, with initial condition $f(0) = 1$. Therefore, by uniqueness of the solution, they must be the same function.

Exercise 7.11 (*The Gaussian Is Its Own Fourier Transform*)**.** Convince yourself that the Gaussian belongs to $\mathcal{S}(\mathbb{R})$. Find the Fourier transform of the Gaussian either by filling in the details in the previous paragraph or by directly evaluating the integral in formula (7.11), for instance using contour integration. ♢

Definition 7.12. The *convolution* $f * g$ of two functions f, $g \in \mathcal{S}(\mathbb{R})$ is the function defined by

$$f * g(x) := \int_{\mathbb{R}} f(x-y)g(y)\,dy. \tag{7.12}$$ ♢

One can verify that if $f \in \mathcal{S}(\mathbb{R})$, then for each $x \in \mathbb{R}$ the new function $(\widetilde{\tau_x f})(y) := f(x-y)$ is in $\mathcal{S}(\mathbb{R})$. By Exercise 7.3 the integral in the definition of convolution is well-defined for each $x \in \mathbb{R}$ and for f, $g \in \mathcal{S}(\mathbb{R})$.

Convolution is a very important operation in harmonic analysis. In the context of Fourier series, we saw how to convolve two periodic functions on $\mathbb{T}$. In Chapter 6 we touched on circular convolution for vectors. In Section 7.8 we say more about convolutions on $\mathbb{R}$. In Chapter 8 we discuss convolution of a Schwartz function with a distribution, and in Chapter 12 we discuss convolution of an L^p function with an L^q function.

7.3. The time–frequency dictionary for $\mathcal{S}(\mathbb{R})$

The Fourier transform interacts very nicely with a number of operations. In particular, *differentiation is transformed into polynomial* *multiplication* and vice versa, which sheds light on the immense success of Fourier transform techniques in the study of differential equations. Also, *convolutions are transformed into products* and vice versa, which underlies the success of Fourier transform techniques in signal processing, since convolution, or *filtering*, is one of the most important signal-processing tools. In this section we present a time–frequency dictionary that lists all these useful interactions.

The Fourier transform is a *linear transformation*, meaning that the Fourier transform of a linear combination of Schwartz functions f and g is equal to the same linear combination of the Fourier transforms of f and g: $(af + bg)^\wedge = a\widehat{f} + b\,\widehat{g}$ for all a, $b \in \mathbb{C}$.

The Fourier transform also interacts very nicely with *translations*, *modulations*, and *dilations*. In Table 7.1 we list ten of these extremely useful fundamental properties of the Fourier transform.

We refer to the list in Table 7.1 as the *time–frequency dictionary.* Here the word *time* is used because the variable x in $f(x)$ often stands for time, for instance when $f(x)$ represents a voice signal taking place over some interval of time, as in Example 1.1. Similarly, the word *frequency* refers to the variable ξ in $\widehat{f}(\xi)$. Operations on the function $f(x)$ are often said to be happening *in the time domain* or *on the time side*, while operations on the Fourier transform $\widehat{f}(\xi)$ are said to be *in the frequency domain* or *on the frequency side*, or simply *on the Fourier side.*

As in a dictionary used for converting between two languages, an entry in the right-hand column of the time–frequency dictionary gives the equivalent on the frequency side of the corresponding entry expressed on the time side in the left-hand column.

For example, the expressions $e^{2\pi ihx}f(x)$ and $\widehat{f}(\xi - h)$ convey in two different ways the same idea, that of shifting by the amount h all the frequencies present in the signal $f(x)$. To give some terminology, *modulation* of a function $f(x)$ means multiplying the function by a term of the form $e^{2\pi ihx}$ for some real number h, as shown in the term $e^{2\pi ihx}f(x)$ on the left-hand side of property (c) in Table 7.1. We denote such a multiplication by $M_h f(x)$. As usual, horizontal *translation* of a function means adding a constant, say $-h$, to its

Table 7.1. The time–frequency dictionary in $\mathcal{S}(\mathbb{R})$.

	Time	**Frequency**
(a)	linear properties $af + bg$	linear properties $a\widehat{f} + b\widehat{g}$
(b)	translation $\tau_h f(x) := f(x-h)$	modulation $\widehat{\tau_h f}\,(\xi) = M_{-h}\widehat{f}\,(\xi)$
(c)	modulation $M_h f(x) := e^{2\pi i h x} f(x)$	translation $\widehat{M_h f}(\xi) = \tau_h \widehat{f}\,(\xi)$
(d)	dilation $D_s f(x) := s f(sx)$	inverse dilation $\widehat{D_s f}(\xi) = s D_{s^{-1}}\widehat{f}(\xi)$
(e)	reflection $\widetilde{f}(x) := f(-x)$	reflection $\widehat{\widetilde{f}}\,(\xi) = \widetilde{\widehat{f}}\,(\xi)$
(f)	conjugation $\overline{f}(x) := \overline{f(x)}$	conjugate reflection $\widehat{\overline{f}}(\xi) = \overline{\widehat{f}\,(-\xi)} = \overline{(f^\vee)}(\xi)$
(g)	derivative $f'(x)$	multiply by polynomial $\widehat{f'}(\xi) = 2\pi i \xi \widehat{f}\,(\xi)$
(h)	multiply by polynomial $-2\pi i x f(x)$	derivative $[-2\pi i x f(x)]^\wedge(\xi) = \frac{d}{d\xi}\widehat{f}\,(\xi)$
(i)	convolution $f * g(x) := \int f(x-y) g(y)\, dy$	product $\widehat{f * g}(\xi) = \widehat{f}\,(\xi)\widehat{g}\,(\xi)$
(j)	product $f(x)g(x)$	convolution $\widehat{fg}(\xi) = \widehat{f} * \widehat{g}\,(\xi)$

argument, as shown in the term $f(x-h)$ on the left-hand side of property (b). We denote such a translation by $\tau_h f(x)$.

Thus, row (c) in the table should be read as saying that *the Fourier transform of* $e^{2\pi i h x} f(x)$ *is equal to* $\widehat{f}\,(\xi - h)$. In other words, *modulation by h on the time side is transformed into translation by h on the frequency side.*

Notice that if f, $g \in \mathcal{S}(\mathbb{R})$, then all the functions in the time column of the table belong to the Schwartz class as well. In other words, the Schwartz class is closed not only under products, multiplication by polynomials and trigonometric functions (modulations), differentiation, and convolution, but also under simpler operations such as linear operations, translations, dilations, and conjugations.

Property (a) says that the Fourier transform is linear. Properties (b)–(e), known as *symmetry* properties or *group invariance* properties, explain how the Fourier transform interacts with the symmetries (translation, dilation, reflection) of the domain of the function f and of $\widehat{f}$. For us this domain is either $\mathbb{T}$, $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$, or $\mathbb{R}$; all three are groups. Properties (f)–(j) explain how the Fourier transform interacts with conjugation and differentiation of functions f and with multiplication of functions f by polynomials, as well as with products and convolution of functions f with other Schwartz functions.

We leave it as an exercise to prove most of the properties listed in the time–frequency dictionary. We have already proved most of these properties for periodic functions and their Fourier coefficients, and the proofs in the present setting are almost identical.

Exercise 7.13. Verify properties (a)–(g) in Table 7.1. $\Diamond$

Convolution in $\mathcal{S}(\mathbb{R})$ as well as properties (i) and (j) will be discussed in Section 7.5.

Let us prove property (h), which gives the connection between taking a derivative on the Fourier side and multiplying f by an appropriate polynomial on the time side. Namely, multiplying a function f by the polynomial $-2\pi i x$ yields a function whose Fourier transform is the derivative with respect to ξ of the Fourier transform $\widehat{f}$ of f. Property (h) does not have an immediate analogue in the Fourier series setting, and for this reason we sketch its proof here.

Proof of property (h). Differentiating (formally) under the integral sign leads us immediately to the correct formula. It takes some effort to justify the interchange of the derivative and the integral. We present the details because they illustrate a standard way of thinking in analysis that we will use repeatedly. One could skip ahead and return to this proof later.

We will check that $\widehat{f}$ is differentiable (as a bonus we get continuity of $\widehat{f}$) and that its derivative coincides with the Fourier transform of $-2\pi i x f(x)$, both in one stroke. By the definition of the derivative, we must show that

$$\lim_{h\to 0}\big[\widehat{f}(\xi+h)-\widehat{f}(\xi)\big]/h = [-2\pi i x f(x)]^\wedge(\xi). \tag{7.13}$$

By definition of the Fourier transform,

$$\begin{aligned}&\big[\widehat{f}(\xi+h)-\widehat{f}(\xi)\big]/h - [-2\pi i x f(x)]^\wedge(\xi)\\ &\qquad = \int_{\mathbb{R}} f(x)e^{-2\pi i x\xi}\Big[\frac{e^{-2\pi i x h}-1}{h}+2\pi i x\Big]\,dx.\end{aligned} \tag{7.14}$$

It suffices to check that for each $\varepsilon > 0$ we can make the absolute value of the integral on the right-hand side of equation (7.14) smaller than a constant multiple of ε, provided h is small enough, say for $|h| < h_0$. We bound separately the contributions for large x and for small x. For each $N > 0$, the right-hand side of equation (7.14) is bounded above by the sum of the integrals (A) and (B) defined here:

$$\begin{aligned}(A) &:= \int_{|x|\le N} |f(x)|\Big|\frac{e^{-2\pi i x h}-1}{h}+2\pi i x\Big|\,dx,\\ (B) &:= 2\pi\int_{|x|>N} |x|\,|f(x)|\Big|\frac{e^{-2\pi i x h}-1}{2\pi x h}+i\Big|\,dx.\end{aligned}$$

Since $f\in\mathcal{S}(\mathbb{R})$, we have that f is bounded by some $M>0$. Moreover, for $|x|\le N$, there is some h_0 such that for all $|h|<h_0$,

$$\big|\big(e^{-2\pi i x h}-1\big)/h+2\pi i x\big| \le \varepsilon/2N. \tag{7.15}$$

Therefore $(A)\le M\varepsilon$. We leave the proof of (7.15) to Exercise 7.14.

Since $f\in\mathcal{S}(\mathbb{R})$, f is rapidly decreasing. In particular there is some $C>0$ such that $|x|^3|f(x)|\le C$ for all $|x|>1$, and so

$$|xf(x)|\le Cx^{-2} \quad \text{for all } |x|>1.$$

Since $\int_{|x|>1} x^{-2}\,dx < \infty$, the tails of this integral must be going to zero, meaning that $\lim_{N\to\infty}\int_{|x|>N} x^{-2}\,dx = 0$. Therefore, given $\varepsilon>0$, there is an $N>0$ such that

$$\int_{|x|>N}|xf(x)|\,dx \le \int_{|x|>N}\frac{C}{x^2}\,dx \le \varepsilon.$$

Also, for all real θ, $|e^{i\theta} - 1| \leq |\theta|$, and so $\left|(e^{-i\theta} - 1)/\theta + i\right| \leq 2$. All these facts together imply that $(B) \leq 4\pi\varepsilon$. Hence for each $\varepsilon > 0$ there exists $h_0 > 0$ such that for all $|h| < h_0$,

$$\left|[\widehat{f}(\xi + h) - \widehat{f}(\xi)]/h - [-2\pi i x f(x)]^\wedge(\xi)\right| \leq (M + 4\pi)\varepsilon.$$

Therefore equation (7.13) holds, by definition of the limit. Hence $(d/d\xi)\widehat{f}(\xi) = [-2\pi i x f(x)]^\wedge(\xi)$, which proves property (h). □

Exercise 7.14. Prove inequality (7.15). You could consider the continuous function $g(x, y) := (\partial/\partial y)e^{-2\pi i x y} = -2\pi i x e^{-2\pi i x y}$, which is necessarily uniformly continuous on any compact set in $\mathbb{R}^2$, in particular on a closed rectangle $R_N := [-N, N] \times [-1, 1]$. ♢

To prove property (h), we used only the decay properties of f. First, to estimate term (A), we used the boundedness of f to show that $\int_{|x| \leq N} |f(x)|\, dx < 2NM$. Second, to estimate term (B), we used the fact that $\int |x||f(x)|\, dx < \infty$ to ensure that the tail of the integral (for $|x| > N$) can be made small enough. Note that the integrability of $|x|\,|f(x)|$ is the minimum necessary to ensure that we can calculate the Fourier transform of $-2\pi i x f(x)$.

Exercise 7.15. Show that if $\int_{\mathbb{R}} (1 + |x|)^k |f(x)|\, dx < \infty$, then $\widehat{f}$ is k times differentiable. Moreover, in that case

$$(d^k/d\xi^k)\widehat{f}(\xi) = [(-2\pi i x)^k f(x)]^\wedge(\xi). \tag{7.16}$$

Hint: Do the case $k = 1$ first. Then use induction on k. For $k = 1$, assume that $\int_{\mathbb{R}} (1 + |x|)|f(x)|\, dx < \infty$. Hence both $\int_{\mathbb{R}} |f(x)|\, dx < \infty$ and $\int_{\mathbb{R}} |x||f(x)|\, dx < \infty$. Then follow the proof of property (h). ♢

7.4. The Schwartz class and the Fourier transform

In this section we show that the Fourier transform of a Schwartz function f is also a Schwartz function (Theorem 7.18). We must check that $\widehat{f}$ is infinitely differentiable and that $\widehat{f}$ and all its derivatives are rapidly decaying.

First, Exercise 7.15 above shows that $f \in \mathcal{S}(\mathbb{R})$ implies that $\widehat{f} \in C^\infty(\mathbb{R})$ and gives a formula for the k^{th} derivative of $\widehat{f}$. Second, we prove a version of the Riemann–Lebesgue Lemma in $\mathcal{S}(\mathbb{R})$ (Lemma 7.16), showing that the Fourier transform $\widehat{f}$ of a Schwartz function f tends to zero as $|\xi| \to \infty$. This result must hold if $\widehat{f}$ is to be in the Schwartz class $\mathcal{S}(\mathbb{R})$. Third, we show that the Fourier transform $\widehat{f}$ of a Schwartz function f is rapidly decreasing (Lemma 7.17). With these ingredients in hand we prove Theorem 7.18.

Lemma 7.16 (Riemann–Lebesgue Lemma). *If $f \in \mathcal{S}(\mathbb{R})$, then*

$$\lim_{|\xi|\to\infty} \widehat{f}(\xi) = 0.$$

Proof. Using a symmetrization trick that we saw when proving the Riemann–Lebesgue Lemma for periodic functions, we find that

$$\begin{aligned}
\widehat{f}(\xi) &= \int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}\,dx = -\int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}e^{\pi i}\,dx \\
&= -\int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi\left(x-1/(2\xi)\right)}\,dx \\
&= -\int_{-\infty}^{\infty} f\Big(y+\frac{1}{2\xi}\Big)e^{-2\pi i\xi y}\,dy.
\end{aligned}$$

The last equality uses the change of variable $y = x - 1/(2\xi)$. It follows, by averaging the two integral expressions for $\widehat{f}(\xi)$, that

$$\widehat{f}(\xi) = \frac{1}{2}\int_{-\infty}^{\infty}\Big[f(x) - f\Big(x+\frac{1}{2\xi}\Big)\Big]e^{-2\pi i\xi x}\,dx.$$

Hence using the Triangle Inequality for integrals and the fact that $|e^{-2\pi i\xi x}| = 1$,

$$|\widehat{f}(\xi)| \le \frac{1}{2}\int_{-\infty}^{\infty}\Big|f(x) - f\Big(x+\frac{1}{2\xi}\Big)\Big|\,dx.$$

Since $f \in \mathcal{S}(\mathbb{R})$, there exists $M > 0$ such that $|f(y)| \le M/y^2$ for all $y \in \mathbb{R}$; in particular it holds for $y = x$ and $y = x+1/(2\xi)$. Given $\varepsilon > 0$, there is some $N > 0$ so that $\int_{|x|>N}\big[1/x^2 + 1/\big(x+1/(2\xi)\big)^2\big]\,dx \le \epsilon/M$, for $|\xi| > 1$ (check!). Therefore we conclude that

$$\int_{|x|>N}\Big|f(x) - f\Big(x+\frac{1}{2\xi}\Big)\Big|\,dx \le \varepsilon.$$

For the integral over $|x| \leq N$, let $g_\xi(x) := f(x) - f(x + 1/(2\xi))$. The functions g_ξ converge uniformly to zero as $|\xi| \to \infty$ on $|x| \leq N$, because f is uniformly continuous on $[-N-1, N+1]$. We can make $g_\xi(x)$ as small as we wish for all $|x| \leq N$, provided ξ is large enough. More precisely, given $\varepsilon > 0$ there exists $K > 0$ such that for all $|\xi| > K$ and for all $|x| \leq N$, $|g_\xi(x)| < \varepsilon/(4N)$. Therefore

$$|\widehat{f}(\xi)| \leq \frac{\varepsilon}{2} + \int_{|x|\leq N} |g_\xi(x)|\, dx \leq \frac{\varepsilon}{2} + 2N\frac{\varepsilon}{4N} = \varepsilon.$$

We conclude that given $\varepsilon > 0$, there exists $K > 0$ such that for all $|\xi| > K$, $|\widehat{f}(\xi)| \leq \varepsilon$. That is, $\lim_{|\xi|\to\infty} |\widehat{f}(\xi)| = 0$, as required. □

In fact this argument works, essentially unchanged, for functions f that are only assumed to be continuous and integrable. Integrability takes care of the integral for $|x| > N$, while continuity controls the integral for $|x| \leq N$. A density argument shows that if $f \in L^1(\mathbb{R})$, then $\lim_{|\xi|\to\infty} \widehat{f}(\xi) = 0$. Furthermore, if $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is continuous; see Lemma 8.51.

The final ingredient needed to prove that the Fourier transform takes $\mathcal{S}(\mathbb{R})$ to itself (Theorem 7.18) is that the Fourier transforms of Schwartz functions decrease rapidly.

Lemma 7.17. *If $f \in \mathcal{S}(\mathbb{R})$, then $\widehat{f}$ is rapidly decreasing.*

Proof. To prove the lemma, we must check that for $f \in \mathcal{S}(\mathbb{R})$,

$$\lim_{\xi\to\infty} |\xi|^n |\widehat{f}(\xi)| = 0 \qquad \text{for all integers } n \geq 0. \tag{7.17}$$

We already know that $\widehat{f}$ is C^∞. The Riemann–Lebesgue Lemma (Lemma 7.16) shows that $\widehat{f}$ vanishes at infinity. We must show that as $|\xi| \to \infty$, $\widehat{f}$ vanishes faster than the reciprocals of all polynomials. By property (g) in Table 7.1, $\widehat{f'}(\xi) = 2\pi i\xi \widehat{f}(\xi)$. Iterating, we get

$$\widehat{f^{(n)}}(\xi) = (2\pi i\xi)^n \widehat{f}(\xi) \qquad \text{for } f \in \mathcal{S}(\mathbb{R}),\ n \geq 0. \tag{7.18}$$

Since $f \in \mathcal{S}(\mathbb{R})$, $f^{(n)}$ is also a Schwartz function. Applying the Riemann–Lebesgue Lemma to $f^{(n)}$, we conclude that for all $n \geq 0$, $\lim_{|\xi|\to\infty} \widehat{f^{(n)}}(\xi) = 0$. Therefore $\lim_{|\xi|\to\infty}(2\pi i\xi)^n \widehat{f}(\xi) = 0$. Hence $\widehat{f}$ is rapidly decreasing, as required. □

With these results in hand, we can now prove that the Schwartz class is closed under the Fourier transform.

Theorem 7.18. *If $f \in \mathcal{S}(\mathbb{R})$, then $\widehat{f} \in \mathcal{S}(\mathbb{R})$.*

Proof. Suppose $f \in \mathcal{S}(\mathbb{R})$. We have shown that $\widehat{f} \in C^\infty(\mathbb{R})$ and

$$(d^\ell/d\xi^\ell)\widehat{f}(\xi) = [(-2\pi i x)^\ell f(x)]^\wedge(\xi).$$

In other words, the ℓ^{th} derivative of $\widehat{f}$ is the Fourier transform of the Schwartz function $g_\ell(x) = (-2\pi i x)^\ell f(x)$. It remains to show that $\widehat{f}$ and all its derivatives are rapidly decreasing. Applying Lemma 7.17 to each $g_\ell \in \mathcal{S}(\mathbb{R})$, we find that $(d^\ell/d\xi^\ell)\widehat{f} = \widehat{g_\ell}$ is rapidly decreasing for each $\ell \geq 0$. We conclude that $\widehat{f} \in \mathcal{S}(\mathbb{R})$, as required. $\square$

Equations (7.18) illustrate an important principle in Fourier analysis already discussed for Fourier series in Chapter 3. *Smoothness of a function is correlated with the fast decay of its Fourier transform at infinity.* In other words, the more times a function f can be differentiated, the faster its Fourier transform $\widehat{f}(\xi)$ goes to zero as $\xi \to \pm\infty$. We will see this principle more clearly in Chapter 8 when considering functions that have less decay at $\pm\infty$ than the Schwartz functions do.

7.5. Convolution and approximations of the identity

The convolution $f * g$ of two functions f, $g \in \mathcal{S}(\mathbb{R})$ is defined by

$$f * g(x) := \int_{\mathbb{R}} f(x-y)g(y)\,dy. \tag{7.19}$$

This integral is well-defined for each $x \in \mathbb{R}$ and for f, $g \in \mathcal{S}(\mathbb{R})$. By change of variables, $f * g = g * f$: convolution is commutative.

The new function $f * g$ defined by the integral in equation (7.19) is rapidly decreasing, because of the smoothness and fast decay of the functions f and g. It turns out that the derivatives of $f * g$ are rapidly decreasing as well, and so the convolution of two Schwartz functions is again a Schwartz function. The key observation is that

$$\frac{d}{dx}\big(f * g\big)(x) = \frac{df}{dx} * g(x) = f * \frac{dg}{dx}(x) \quad \text{for } f,\, g \in \mathcal{S}(\mathbb{R}).$$

The first identity holds because in this setting, we are entitled to interchange differentiation and integration. To see this, break the integral into two pieces, one where the variable of integration is bounded ($|y| \leq M$), the other where it is not ($|y| > M$). On the bounded part, use Theorem 2.53 to exchange the limit and the integral. On the unbounded part, use the decay of the function to show that the integral goes to zero as $M \to \infty$. This is exactly the argument we used to prove property (h) from the time–frequency dictionary.

The second identity holds because convolution is commutative; it can also be proved by integration by parts.

Exercise 7.19 (*The Schwartz Space Is Closed under Convolution*). Verify that if f, $g \in \mathcal{S}(\mathbb{R})$, then $f * g \in \mathcal{S}(\mathbb{R})$. $\Diamond$

Convolution is a *smoothing* operation. The output keeps the best of each input. This heuristic is not completely evident in the case of convolution of Schwartz functions, because we are already in paradise. One can convolve much less regular functions (for example a function that is merely integrable, or even a generalized function or distribution) with a smooth function; then the smoothness will be inherited by the convolution. In fact when convolving Schwartz functions with distributions (as we do in Chapter 8), we will see that the resulting distribution can be identified with a C^∞ function.

Convolutions correspond to *translation-invariant bounded linear transformations* and to *Fourier multipliers*. These, and their generalizations, have been and are the object of intense study. We discussed the discrete analogue of this statement in Section 6.4.2.

Exercise 7.20. Verify properties (i) and (j) in Table 7.1: if f, $g \in \mathcal{S}(\mathbb{R})$, then $\widehat{f * g}\,(\xi) = \widehat{f}\,(\xi)\widehat{g}\,(\xi)$ and $\widehat{fg}\,(\xi) = \widehat{f} * \widehat{g}\,(\xi)$. $\Diamond$

Convolution can be viewed as a binary operation on the Schwartz class $\mathcal{S}(\mathbb{R})$. However, the Schwartz class with convolution does not form a group, since there is no *identity element* in $\mathcal{S}(\mathbb{R})$. This is simple to verify once we have established property (i) from the time–frequency dictionary. Suppose there were some $\varphi \in \mathcal{S}(\mathbb{R})$ such that $f * \varphi = f$ for all $f \in \mathcal{S}(\mathbb{R})$. Taking the Fourier transform on both sides, we find that $\widehat{f}\,(\xi)\,\widehat{\varphi}\,(\xi) = \widehat{f}(\xi)$. In particular we can set $f(x)$

equal to $G(x) = e^{-\pi x^2}$, the Gaussian, which is equal to its own Fourier transform and never vanishes. Hence we can cancel G from the identity $G(\xi)\,\widehat{\varphi}\,(\xi) = G(\xi)$ and conclude that $\widehat{\varphi} \equiv 1$. But we have just shown that if $\varphi \in \mathcal{S}(\mathbb{R})$, then $\widehat{\varphi} \in \mathcal{S}(\mathbb{R})$. Therefore $\widehat{\varphi}$ must decrease rapidly, but the function identically equal to one does not decrease at all. We have reached a contradiction. There cannot be an identity in $\mathcal{S}(\mathbb{R})$ for the convolution operation.

There are two ways to address this lack of an identity element. The first is to introduce the *delta function*, defined in Section 8.4. The delta function can be thought of as a point mass since, roughly speaking, it takes the value zero except at a single point. The delta function is not a Schwartz function and in fact is not a true function at all but is instead a generalized kind of function called a *distribution*. It acts as an identity under convolution (see Exercise 8.32; one must extend the definition of convolution appropriately).

The second substitute for an identity element in $\mathcal{S}(\mathbb{R})$ under convolution builds on the first. The idea is to use a parametrized *sequence*, or *family*, of related functions, that tends to the delta distribution as the parameter approaches some limiting value. Such families are called *approximations of the identity*. We have already seen them in the context of Fourier series; see Section 4.3 and especially Definition 4.19. The functions in the family may actually be Schwartz functions, or they may be less smooth.

Definition 7.21. An *approximation of the identity* in $\mathbb{R}$ is a family $\{K_t\}_{t\in\Lambda}$ of integrable real-valued functions on $\mathbb{R}$, where Λ is an index set, together with a point t_0 that is an accumulation point[3] of Λ, with the following three properties.

(i) The functions K_t have mean value one: $\int_{\mathbb{R}} K_t(x)\,dx = 1$.

(ii) The functions K_t are uniformly integrable in t: there is a constant $C > 0$ such that $\int_{\mathbb{R}} |K_t(x)|\,dx \le C$ for all $t \in \Lambda$.

(iii) The mass[4] of K_t becomes increasingly concentrated at $x = 0$ as $t \to t_0$: $\lim_{t\to t_0} \int_{|x|>\delta} |K_t(x)|\,dx = 0$ for each $\delta > 0$. ◊

[3]That is, we can find points $t \in \Lambda$, with $t \neq t_0$, that are arbitrarily close to t_0.

[4]The mass is the integral of the absolute value of K_t.

The examples below illustrate what the accumulation point t_0 should be. Sometimes $t_0 = 0 < t$ and then $t \to 0^+$. Or one can have $t_0 = \infty$ and then $t \to \infty$, or $t_0 = 1$ and then $t \to 1$. In the periodic case the kernels were indexed by $N \in \mathbb{N}$ (Fejér kernel) or by r in $(0, 1)$ (Poisson kernel), and we considered $N \to \infty$ and $r \to 1^-$, respectively. In the periodic case we deduced the concentration of mass at zero from the uniform convergence of the sequence $K_n(x) \to 0$ for $\delta \leq |x| \leq \pi$. However on the real line, uniform convergence to zero of the sequence $K_t(x) \to 0$ for $|x| \geq \delta$ is not sufficient to obtain property (iii), since the mass can be thinly spreading out to infinity.

Notice that in the definition there is no explicit mention of the delta function. Nevertheless the three properties in the definition do imply that as $t \to t_0$, $K_t(x)$ converges in some sense to the delta function, that is, to the identity under convolution; see Exercise 8.33. This fact is the origin of the term *approximation of the identity*. Another aspect of the same idea, in a language we have already met, is the property $f * K_t \to f$ as $t \to t_0$. See Theorem 7.24.

Exercise 7.22 (*Generating an Approximation of the Identity from a Kernel*)**.** An easy way to produce an approximation of the identity is to start with a nonnegative function (or *kernel*) $K(x) \geq 0$ with mean value one, $\int_{\mathbb{R}} K(x)\,dx = 1$, and to define the family of dilations

$$K_t(x) := t^{-1} K(t^{-1}x) \qquad \text{for all } t > 0. \tag{7.20}$$

Show that $\{K_t\}_{t>0}$ is an approximation of the identity, with $t_0 = 0$. ♢

Example 7.23 (*Gaussian Kernels Form an Approximation of the Identity*)**.** Gaussians are good kernels, and by scaling them we obtain an approximation of the identity. Given $G(x) = e^{-\pi x^2}$, define the *Gauss kernels* $G_t(x)$ by

$$G_t(x) := t^{-1} e^{-\pi(t^{-1}x)^2}, \quad \text{for } t > 0.$$

Since $e^{-\pi x^2} > 0$ and $\int_{\mathbb{R}} e^{-\pi x^2}\,dx = 1$, this family is an approximation of the identity with $t_0 = 0$. As t decreases to 0, the graphs of $G_t(x)$ look more and more like the delta function; see Figure 7.1. ♢

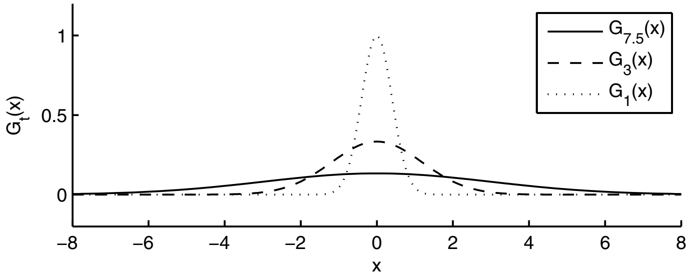


Figure 7.1. Graphs of Gauss kernels $G_t(x)$ for $t = 1$, 3, and 7.5.

Another facet of the approximation-of-the-identity idea is that by convolving against an approximation of the identity, one can approximate a function very well. One instance of this heuristic is given in the next theorem, whose analogue on $\mathbb{T}$ we proved as Theorem 4.23.

Theorem 7.24. *Let $\{K_t\}_{t\in\Lambda}$ be an approximation of the identity as $t \to t_0$. Suppose K_t and f belong to $\mathcal{S}(\mathbb{R})$. Then the convolutions $f * K_t$ converge uniformly to f as $t \to t_0$.*

See also Theorem 7.35 and its proof, for convergence in L^1.

Exercise 7.25. Prove Theorem 7.24. Show that Theorem 7.24 remains true even if f is only assumed to be continuous and integrable, so that the convolution is still well-defined. ◊

Note that because of the smoothing effect of convolution, the approximating functions $f * K_t$ may be much smoother than the original function f. One can use this idea to show that the uniform limit of differentiable functions need not be differentiable.

7.6. The Fourier Inversion Formula and Plancherel

In this subsection we present several fundamental ingredients of the theory of Fourier transforms: the multiplication formula, the Fourier

Inversion Formula, Plancherel's Identity, and the polarization identity. Then we use all these tools in an application to linear differential equations.

It follows as a consequence of differentiation being transformed into polynomial multiplication, and vice versa, that *the Fourier transform maps* $\mathcal{S}(\mathbb{R})$ *into itself.* See Theorem 7.18.

We turn to the *multiplication formula*, relating the inner products of two Schwartz functions with each other's Fourier transforms.

Exercise 7.26 (*The Multiplication Formula*)**.** Verify that

$$\int_{\mathbb{R}} f(s)\,\widehat{g}\,(s)\,ds = \int_{\mathbb{R}} \widehat{f}\,(s)\,g(s)\,ds \quad \text{for } f,\, g \in \mathcal{S}(\mathbb{R}). \tag{7.21}$$

Hint: Use Fubini's Theorem (see the Appendix) to interchange the integrals in an appropriate double integral. ♢

It is interesting to note that there is no analogue of the multiplication formula in the Fourier series setting. The Fourier coefficients of a periodic function are not a function but a sequence of numbers, and so the pairing of, say, f and the Fourier coefficients of g (the analogue of the pairing of f and $\widehat{g}$ in equation (7.21)) would not make sense. However it makes sense in the discrete case, where both the input and the output of the Discrete Fourier Transform are vectors in $\mathbb{C}^N$.

Exercise 7.27. State and prove a multiplication formula, analogous to formula (7.21), that is valid in $\mathbb{C}^N$. ♢

Several of the preceding facts can be woven together to prove the important Fourier Inversion Formula in $\mathcal{S}(\mathbb{R})$.

Theorem 7.28 (Fourier Inversion Formula in $\mathcal{S}(\mathbb{R})$)**.** *If* $f \in \mathcal{S}(\mathbb{R})$, *then for all* $x \in \mathbb{R}$,

$$f(x) = \int_{\mathbb{R}} \widehat{f}\,(\xi) e^{2\pi i \xi x}\, d\xi = (\widehat{f}\,)^{\vee}(x). \tag{7.22}$$

Notice the symmetry between the definition in formula (7.11) of the Fourier transform $\widehat{f}\,(\xi)$ of the function $f(x)$ and the formula (7.22) for $f(x)$ in terms of its Fourier transform $\widehat{f}\,(\xi)$. Formally the only differences are the exchange of the symbols x and ξ and the exchange of the negative or positive signs in the exponent. Whatever conditions are required on f so that formula (7.11) makes sense must also be required on $\widehat{f}$ so that formula (7.22) makes sense. This symmetry is automatic in the Schwartz space, which is why the Fourier theory on $\mathcal{S}(\mathbb{R})$ is so beautiful and self-contained.

Proof of Theorem 7.28. Fix $x \in \mathbb{R}$. The idea is to apply the multiplication formula (7.21) to functions f and g, where f is an arbitrary Schwartz function and g is the specific Schwartz function

$$g(s) := e^{2\pi isx}e^{-\pi|ts|^2} \in \mathcal{S}(\mathbb{R}), \quad \text{for } t > 0.$$

Compute directly or use properties (c) and (d) of the time–frequency dictionary in Section 7.3 to show that

$$\widehat{g}\,(s) = t^{-1}e^{-\pi|t^{-1}(x-s)|^2} = G_t(x-s). \tag{7.23}$$

This expression is a translate of the approximation of the identity generated by the Gaussian. Therefore

$$\begin{aligned} f(x) &= \lim_{t\to 0}\int_{\mathbb{R}} f(s)G_t(x-s)\,ds = \lim_{t\to 0}\int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}e^{-\pi|ts|^2}\,ds \\ &= \int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}\lim_{t\to 0} e^{-\pi|ts|^2}\,ds = \int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}\,ds. \end{aligned}$$

The first equality holds because $f * G_t(x) \to f(x)$ uniformly in x as $t \to 0$, by Exercise 7.25. The second equality holds by equation (7.21). The third equality follows from the Lebesgue Dominated Convergence Theorem, which applies since $|\widehat{f}\,(s)e^{2\pi isx}e^{-\pi|ts|^2}| \le |\widehat{f}\,(s)|$, where $\widehat{f} \in \mathcal{S}(\mathbb{R})$ by Theorem 7.18, and so $|\widehat{f}| \in L^1(\mathbb{R})$. (See Theorem A.59 for a statement of the Lebesgue Dominated Convergence Theorem, in the case of integrands indexed by $n \in \mathbb{N}$ rather than by $t > 0$ as here.) The fourth inequality holds since $\lim_{t\to 0} e^{-\pi|ts|^2} = 1$ for each $s \in \mathbb{R}$.

Lo and behold, $f(x) = (\widehat{f}\,)^{\vee}(x)$, which is the inversion formula we were seeking. □

Exercise 7.29. Use the time–frequency dictionary to check the identity (7.23) used in the proof of the Fourier Inversion Formula. ◊

Exercise 7.30. Justify the interchange of limit and integral to show that

$$\lim_{t\to 0}\int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}e^{-\pi|ts|^2}\,ds = \int_{\mathbb{R}} \widehat{f}\,(s)e^{2\pi isx}\lim_{t\to 0} e^{-\pi|ts|^2}\,ds,$$

for $f \in \mathcal{S}(\mathbb{R})$, without using Lebesgue Dominated Convergence (Theorem A.59). **Hint:** Break the integral into two pieces: where $|s| \leq M$ and where $|s| > M$. For the bounded piece use uniform convergence of the integrands to interchange the limit and the integral. For the unbounded piece use the decay properties of $\widehat{f}$ to show that the integral can be made small for M large enough. $\Diamond$

The inversion formula guarantees that the Fourier transform is in fact a *bijection* from the Schwartz space $\mathcal{S}(\mathbb{R})$ onto itself. More is true: the Fourier transform is an *energy-preserving map*, also called a *unitary transformation*, on $\mathcal{S}(\mathbb{R})$. This is the content of *Plancherel's Identity.* First we must define our notion of the size of a function. We can equip $\mathcal{S}(\mathbb{R})$ with an inner product

$$\langle f, g \rangle := \int_{\mathbb{R}} f(x)\overline{g(x)}\, dx \qquad \text{for all } f,\, g \in \mathcal{S}(\mathbb{R})$$

with associated L^2 norm given by $\|f\|_2^2 = \langle f, f \rangle$. Thus the L^2 norm is defined by $\|f\|_2 := \left(\int_{\mathbb{R}} |f(x)|^2\, dx\right)^{1/2}$. For Schwartz functions f and g, the integral defining the inner product yields a well-defined finite complex number.

Plancherel's Identity says that each Schwartz function f and its Fourier transform $\widehat{f}$ have the same size; in other words, the Fourier transform preserves the L^2 norm.

Theorem 7.31 (Plancherel's Identity)**.** *If* $f \in \mathcal{S}(\mathbb{R})$, *then*

$$\|f\|_2 = \|\widehat{f}\,\|_2.$$

Proof. We use the time–frequency dictionary, the multiplication formula and the Fourier Inversion Formula to give a proof of Plancherel's Identity. Set $\widehat{g} = \overline{f}$ in the multiplication formula (7.21). Then, by the analogue $g = (\overline{f})^{\vee} = \overline{\widehat{f}}$ of property (f) for the inverse Fourier transform, we obtain

$$\int |f|^2 = \int f\,\overline{f} = \int f\,\widehat{g} = \int \widehat{f}\,g = \int \widehat{f}\,\overline{\widehat{f}} = \int |\widehat{f}\,|^2. \qquad \square$$

We end this section with two useful results. The first is the *polarization identity*, which says that the inner product of two Schwartz functions is equal to the inner product of their Fourier transforms.

The second is an application to differential equations that relies crucially on Theorem 7.18.

Exercise 7.32 (*Polarization Identity*)**.** Prove the polarization identity for real-valued functions f and g in $\mathcal{S}(\mathbb{R})$. Namely,

$$\langle f, g\rangle = \langle \widehat{f}, \widehat{g}\,\rangle. \tag{7.24}$$

Hint: Use Plancherel's Identity for $f+g$ and $f-g$ and then add. ♢

Example 7.33 (*An Application to Solving Differential Equations*)**.** Let us find a function f on $\mathbb{R}$ that satisfies the (driven, linear, constant-coefficient) differential equation

$$f^{(3)}(x) + 3f''(x) - 2f'(x) - 6f(x) = e^{-\pi x^2}.$$

Taking the Fourier transform on both sides of the equation and using the time–frequency dictionary, we obtain

$$\widehat{f}\,(\xi)\big[(2\pi i\xi)^3 + 3(2\pi i\xi)^2 - 2(2\pi i\xi) - 6\big] = e^{-\pi\xi^2}. \tag{7.25}$$

Let $P(t) = t^3 + 3t^2 - 2t - 6 = (t+3)(t^2-2)$. Then the polynomial inside the brackets in equation (7.25) is $Q(\xi) := P(2\pi i\xi)$. Solving for $\widehat{f}\,(\xi)$ and using the Fourier Inversion Formula in $\mathcal{S}(\mathbb{R})$, we obtain

$$\widehat{f}\,(\xi) = e^{-\pi\xi^2}/Q(\xi) \qquad \text{and so} \qquad f(x) = \big(e^{-\pi\xi^2}/Q(\xi)\big)^{\vee}(x).$$

(Fortunately $\widehat{f}(\xi)$ does lie in $\mathcal{S}(\mathbb{R})$, since the denominator $Q(\xi)$ has no real zeros; check!) We have expressed the solution $f(x)$ as the inverse Fourier transform of a known function depending only on the driving term and the coefficients. ♢

In general, let $P(x) = \sum_{k=0}^{n} a_k x^k$ be a polynomial of degree n with constant complex coefficients a_k. If $Q(\xi) = P(2\pi i\xi)$ has no real zeros and $u \in \mathcal{S}(\mathbb{R})$, then the linear differential equation

$$P(D)f = \sum_{0\le k\le n} a_k D^k f = u$$

has solution $f = \big(\widehat{u}(\xi)/Q(\xi)\big)^{\vee}$.

Note that since $Q(\xi)$ has no real zeros, the function $\widehat{u}(\xi)/Q(\xi)$ is in the Schwartz class. Therefore we may compute its inverse Fourier transform and, by Theorem 7.18, the result is a function in $\mathcal{S}(\mathbb{R})$. Often $\big(\widehat{u}(\xi)/Q(\xi)\big)^{\vee}$ can be computed explicitly.

7.7. L^p norms on $\mathcal{S}(\mathbb{R})$

In this subsection we define the L^p norm $\|f\|_p$ of a Schwartz function f for $1 \leq p \leq \infty$. The heart of the subsection is the proof of Theorem 7.35, showing that for Schwartz functions f, the convolution of f with a suitable approximation to the identity converges to f in the L^p sense for each real number p such that $1 \leq p < \infty$.

We equip the Schwartz space $\mathcal{S}(\mathbb{R})$ with a family of norms, called the L^p *norms*, indexed by the real numbers p such that $1 \leq p < \infty$ and by $p = \infty$. For $f \in \mathcal{S}(\mathbb{R})$ let $\|f\|_p := \left(\int_{\mathbb{R}} |f(x)|^p \, dx\right)^{1/p}$, for $1 \leq p < \infty$, and $\|f\|_\infty := \sup_{|x| \in \mathbb{R}} |f(x)|$, for $p = \infty$.

Exercise 7.34 (*Schwartz Functions Have Finite L^p Norm*)**.** Verify that if $f \in \mathcal{S}(\mathbb{R})$, then $\|f\|_p < \infty$ for each p with $1 \leq p \leq \infty$. ◊

The Schwartz space $\mathcal{S}(\mathbb{R})$ with the L^p norm is a normed space. In particular, the *Triangle Inequality* holds:

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p,$$

for all f, $g \in \mathcal{S}(\mathbb{R})$ and all p with $1 \leq p \leq \infty$. This Triangle Inequality for the L^p norm is also known as *Minkowski's Inequality*[5]. We prove it in Chapter 12 (Lemma 12.48).

When $\|f\|_p < \infty$, we say that f belongs to $L^p(\mathbb{R})$, or $f \in L^p(\mathbb{R})$. Thus for each p with $1 \leq p \leq \infty$, the function space $L^p(\mathbb{R})$ consists of those functions $f : \mathbb{R} \to \mathbb{R}$ whose L^p norms are finite. To make this precise, we have to appeal to Lebesgue integration on $\mathbb{R}$, as we did when defining $L^p(\mathbb{T})$ in Chapter 2. Exercise 7.34 shows that $\mathcal{S}(\mathbb{R}) \subset L^p(\mathbb{R})$. A shortcut, which we have already used when talking about $L^p(\mathbb{T})$, would be to define $L^p(\mathbb{R})$ as the *completion of* $\mathcal{S}(\mathbb{R})$ *with respect to the L^p norm.* The two definitions are equivalent. The second has the advantage that it immediately implies that the Schwartz class is dense in $L^p(\mathbb{R})$ with respect to the L^p norm.

Given an approximation of the identity $\{K_t\}_{t \in \Lambda}$ in $\mathcal{S}(\mathbb{R})$ and a function $f \in \mathcal{S}(\mathbb{R})$, their convolutions $K_t * f$ converge to f not only uniformly but also in the L^p norm. We state and prove this approximation result in the case $p = 1$.

[5]Named after the German mathematician Hermann Minkowski (1864–1909).

Theorem 7.35. *Let the family* $\{K_t\}_{t\in\Lambda}$ *be an approximation of the identity in* $\mathcal{S}(\mathbb{R})$ *as* $t \to t_0$*. If* $f \in \mathcal{S}(\mathbb{R})$*, then*

$$\lim_{t\to 0} \|K_t * f - f\|_1 = 0.$$

Proof. Define the translation operator τ_y by $\tau_y f(x) := f(x-y)$, for $y \in \mathbb{R}$. The following pointwise inequality holds:

$$\begin{aligned} |K_t * f(x) - f(x)| &= \left| \int_{\mathbb{R}} f(x-y)\, K_t(y)\, dy - f(x) \right| \\ &= \left| \int_{\mathbb{R}} [\tau_y f(x) - f(x)]\, K_t(y)\, dy \right| \\ &\leq \int_{\mathbb{R}} |\tau_y f(x) - f(x)|\, |K_t(y)|\, dy. \end{aligned} \tag{7.26}$$

In the second equality we used property (i) of the kernels K_t, namely that they have integral equal to one, so that $f(x) = \int f(x) K_t(y)\, dy$. The inequality holds because the absolute value of an integral is smaller than the integral of the absolute value (the *Triangle Inequality for integrals*). It can be justified by noticing that the integrals can be approximated by Riemann sums (finite sums), for which the ordinary Triangle Inequality for absolute values can be applied. Now integrate the pointwise inequality just derived and apply Fubini's Theorem (Theorem A.58) to interchange the order of integration:

$$\begin{aligned} \|K_t * f - f\|_1 &\leq \int_{\mathbb{R}} \int_{\mathbb{R}} |\tau_y f(x) - f(x)|\, |K_t(y)|\, dy\, dx \\ &= \int_{\mathbb{R}} \|\tau_y f - f\|_1 |K_t(y)|\, dy. \end{aligned} \tag{7.27}$$

We estimate separately the contributions to the right-hand side of inequality (7.27) for large and small y.

First, the integral of the absolute values of K_t away from zero can be made arbitrarily small for t close to t_0; this is property (iii) of the approximation of the identity. More precisely, given $f \in \mathcal{S}(\mathbb{R})$, $\varepsilon > 0$, and $\delta > 0$, we can choose h small enough so that for all $|t - t_0| < h$,

$$\int_{|y|>\delta} |K_t(y)|\, dy \leq \varepsilon/4\|f\|_1.$$

Second, we can make $\|\tau_y f - f\|_1$ small provided y is small as follows. Since $f \in \mathcal{S}(\mathbb{R})$, we know that f is continuous and $f \in L^1(\mathbb{R})$. For continuous functions in $L^1(\mathbb{R})$,

$$\lim_{y\to 0} \|\tau_y f - f\|_1 = 0. \tag{7.28}$$

In other words, translation is a continuous operation at zero in L^1, at least for continuous functions f. See Exercise 7.37. A density argument shows that the statement is actually true for all functions in $L^1(\mathbb{R})$, not just for those that are continuous. Hence, given $f \in \mathcal{S}(\mathbb{R})$ and $\varepsilon > 0$, there exists $\delta > 0$ (depending on f) such that

$$\|\tau_y f - f\|_1 \leq \varepsilon/2C \qquad \text{for all } |y| < \delta. \tag{7.29}$$

For $|y| < \delta$ we apply estimate (7.29), with C the constant that uniformly controls the L^1 norms of K_t, that is, $\int |K_t(y)|dy \leq C$ for all $t \in \Lambda$. Altogether, we have

$$\begin{aligned}
\|K_t * f - f\|_1 &\leq \int_{|y|<\delta} |K_t(y)|\|\tau_y f - f\|_1\, dy \\
&\quad + \int_{|y|\geq\delta} |K_t(y)|\|\tau_y f - f\|_1\, dy \\
&\leq \frac{\varepsilon}{2C}\int_{|y|<\delta} |K_t(y)|\, dy + 2\|f\|_1 \int_{|y|\geq\delta} |K_t(y)|\, dy \\
&\leq \varepsilon/2 + 2\|f\|_1 \varepsilon/4\|f\|_1 = \varepsilon.
\end{aligned}$$

In the last inequality we have used properties (ii) and (iii) of the approximation of the identity. Theorem 7.35 is proved. □

The proof above can be modified for the case $p = \infty$, as stated in Theorem 7.24, and for $1 < p < \infty$. A key step is to use an integral version of the Triangle Inequality, known as *Minkowski's Integral Inequality*, to prove the L^p version of inequality (7.27). Minkowski's Integral Inequality essentially says that the *norm of the integral is less than or equal to the integral of the norms.* See p. 357 for a precise statement and some applications.

Exercise 7.36. Show that if $f \in \mathcal{S}(\mathbb{R})$ and $y \in \mathbb{R}$, then $\tau_y f \in \mathcal{S}(\mathbb{R})$ and the L^p norm is preserved: $\|f\|_p = \|\tau_y f\|_p$, for all $1 \leq p < \infty$. ◊

Exercise 7.37. Show that we can interchange the limit and integral: $\lim_{y\to 0}\int_{\mathbb{R}}|f(x-y)-f(x)|^p\,dx = \int_{\mathbb{R}}\lim_{y\to 0}|f(x-y)-f(x)|^p\,dx = 0$, for $f\in\mathcal{S}(\mathbb{R})$. ♢

Exercises 7.36 and 7.37 for $p = 1$ show that equation (7.28) is valid.

Exercise 7.38. State and prove an L^p version of Theorem 7.35. ♢

We have not used the fact that $K \in \mathcal{S}(\mathbb{R})$, except to feel comfortable when writing down the integrals. We did use the defining properties of an approximation of an identity. As for f, all that is really required is that f is L^p integrable.

7.8. Project: A bowl of kernels

When we studied summability methods in Sections 4.4 and 4.5, we used the Cesàro and Abel means to improve the convergence of Fourier series, by convolving with the Fejér and Poisson kernels on the circle $\mathbb{T}$. Those kernels happened to be approximations of the identity on $\mathbb{T}$. The analogous Fejér and Poisson kernels on the real line $\mathbb{R}$ can be used to improve the convergence of the Fourier integral. In this project we define the Fejér kernel, the Poisson kernel, the heat kernel, the Dirichlet kernel, and the conjugate Poisson kernel on $\mathbb{R}$, and we ask the reader to explore some of their properties.

The Fejér kernel on $\mathbb{R}$:

$$F_R(x) := R\left[\big(\sin(\pi Rx)\big)/\pi Rx\right]^2 \qquad \text{for } R > 0. \tag{7.30}$$

The Poisson kernel on $\mathbb{R}$:

$$P_y(x) := y/\pi(x^2+y^2) \qquad \text{for } y > 0. \tag{7.31}$$

The heat kernel on $\mathbb{R}$ (a modified Gaussian):

$$H_t(x) := (1/\sqrt{4\pi t})\, e^{-|x|^2/4t} \qquad \text{for } t > 0. \tag{7.32}$$

The Dirichlet kernel on $\mathbb{R}$:

$$D_R(x) := \int_{-R}^{R} e^{2\pi i x\xi}\, d\xi = \frac{\sin(2\pi Rx)}{\pi x} \qquad \text{for } R > 0. \tag{7.33}$$

The conjugate Poisson kernel on $\mathbb{R}$:

(7.34) $$Q_y(x) := x/\pi(x^2+y^2) \qquad \text{for } y>0.$$

Of these kernels, only the heat kernel is a Schwartz function. The Fejér and Poisson kernels are *functions of moderate decrease* (Chapter 8). For such functions we can reproduce essentially all the proofs above concerning approximations of the identity. Since the Dirichlet kernels are not uniformly integrable, the family $\{D_R\}_{R>0}$ does not generate an approximation of the identity. However, the integral averages of the Dirichlet kernels are the Fejèr kernels, which are an approximation of the identity. Investigate the history behind these kernels. See also Chapter 8 for the much larger class of *tempered distributions*, which includes all the kernels above.

(a) Show that the Poisson kernel is a solution of *Laplace's equation*, $\Delta P := (\partial^2/\partial x^2)P + (\partial^2/\partial y^2)P = 0$, for (x,y) in the upper half-plane $(x \in \mathbb{R}, y>0)$. You can start your search in [**SS03**, Section 5.2.2].

(b) Show that the heat kernel is a solution of the *heat equation* on the line: $(\partial/\partial t)H = (\partial^2/\partial x^2)H$. You may benefit from reading [**Kör**, Chapters 54–58] and [**SS03**, Section 5.2.1].

(c) Show that the Fejér kernels $\{F_R\}_{R>0}$, the Poisson kernels $\{P_y\}_{y>0}$, the Gaussian kernels $\{G_\delta\}_{\delta>0}$, and the heat kernels $\{H_t\}_{t>0}$ generate approximations of the identity on $\mathbb{R}$, $R\to\infty$, $y\to 0$, as $\delta\to 0$ and $t\to 0$, respectively. The well-known identities $\int_{\mathbb{R}} (1-\cos x)/x^2\,dx = \pi$ and $\int_{\mathbb{R}} 1/(1+x^2)\,dx = \pi$ may be helpful.

(d) Let S_Rf denote the *partial Fourier integral* of f, defined by $S_Rf(x) := \int_{-R}^{R} \widehat{f}(\xi)e^{2\pi i x\xi}\,d\xi$. Show that $S_Rf = D_R * f$ and that $F_R(x) = \frac{1}{R}\int_0^R D_t(x)\,dt$.

(e) Show that the conjugate Poisson kernels are square integrable but not integrable: $Q_y \notin L^1(\mathbb{R})$, but $Q_y \in L^2(\mathbb{R})$. Also show that the conjugate Poisson kernel satisfies Laplace's equation $\Delta Q = 0$. Finally, show that $P_y(x) + iQ_y(x) = 1/(\pi i z)$, where $z = x+iy$. The function $1/(\pi i z)$ is analytic on the upper half-plane; it is known as the *Cauchy kernel*. Thus the Poisson kernel and the conjugate Poisson kernel are the real and imaginary parts, respectively, of the Cauchy kernel $1/(\pi i z)$.

Chapter 8

Beyond paradise

The Fourier transform and most of its properties are valid in a more general setting than the class $\mathcal{S}(\mathbb{R})$ of Schwartz functions. Both the differentiability and the rapid decrease of the Schwartz functions can be relaxed. In this chapter we extend the Fourier transform first to the class of *continuous functions of moderate decrease*, which contains $\mathcal{S}(\mathbb{R})$ (Section 8.1), and then to the still larger class $\mathcal{S}'(\mathbb{R})$ of *tempered distributions*, which contains the $L^p(\mathbb{R})$ functions for all real p with $1 \leq p \leq \infty$, the delta distribution, the Borel measures, and more (Section 8.2).

Strictly speaking, the tempered distributions are not functions at all, as the example of the delta distribution shows (Section 8.4). However, the distributions still have a Fourier theory, meaning that we can extend to them the fundamental operations discussed in the previous chapter, such as translation, dilation, the Fourier transform, and convolution with Schwartz functions (Section 8.3).

We present three beautiful and simple consequences of Fourier theory: the Poisson summation formula, the Whittaker–Shannon sampling formula, and the Heisenberg uncertainty principle, with important applications in number theory, signal processing, and quantum mechanics (Section 8.5). To finish, we return to tempered distributions and the interplay of the Fourier transform with the Lebesgue spaces $L^p(\mathbb{R})$ (Section 8.6). In the project in Section 8.7 we examine the *principal value distribution* $1/x$, a tempered distribution of a different form: neither a function nor a delta distribution nor a measure. Its significance for us is that it is the heart of the Hilbert transform (Chapter 12).

8.1. Continuous functions of moderate decrease

We extend the Fourier transform from the Schwartz class $\mathcal{S}(\mathbb{R})$ to the class of continuous functions of moderate decrease.

Definition 8.1. A function f is said to be a *continuous function of moderate decrease* if f is continuous and if there are constants A and $\varepsilon > 0$ such that

$$|f(x)| \leq A/(1+|x|^{1+\varepsilon}) \qquad \text{for all } x \in \mathbb{R}. \tag{8.1}$$

◊

The 1 in the denominator means that $|f|$ cannot be too big for x near 0, while the $|x|^{1+\varepsilon}$ in the denominator means that $|f|$ decreases like $|x|^{-(1+\varepsilon)}$ as $x \to \pm\infty$. The continuity will allow us to define the integral over $\mathbb{R}$ of such functions without needing Lebesgue integration theory. The growth condition (8.1) ensures that such an f is bounded and is both integrable and square integrable.

Exercise 8.2. Show that all Schwartz functions are continuous functions of moderate decrease. ◊

Example 8.3 (*Continuous Functions of Moderate Decrease Need Not Be Schwartz*)**.** There are functions such as

$$1/(1+x^{2n}) \quad \text{for } n \geq 1 \qquad \text{and} \qquad e^{-a|x|} \quad \text{for } a > 0$$

that are continuous functions of moderate decrease, although they are not in $\mathcal{S}(\mathbb{R})$. The first of these functions is smooth with slow decay; the second is nonsmooth with rapid decay. ◊

Exercise 8.4. Neither the Poisson kernel (7.31) nor the Fejér kernel (7.30) belong to $\mathcal{S}(\mathbb{R})$. On the other hand, show that for each $R > 0$ and each $y > 0$, the Fejér kernel F_R and the Poisson kernel P_y are continuous functions of moderate decrease with $\varepsilon = 1$. Show also that the conjugate Poisson kernel Q_y is *not* a continuous function of moderate decrease (see the project in Section 7.8). ◊

Exercise 8.5. Verify that the class of continuous functions of moderate decrease is closed under multiplication by bounded continuous functions. ◊

We summarize the integrability and boundedness properties of continuous functions of moderate decrease. Define the *improper integral over* $\mathbb{R}$ *of a continuous function* f *of moderate decrease* as the limit as $N \to \infty$ of the Riemann integrals of f over the compact intervals $[-N, N]$:

$$(8.2) \qquad \int_{-\infty}^{\infty} f(x)\, dx := \lim_{N\to\infty} \int_{-N}^{N} f(x)\, dx < \infty.$$

By Exercise 8.5, if f is a continuous function of moderate decrease, then so too is $e^{-2\pi i x\xi} f(x)$, and the integral over $x \in \mathbb{R}$ of this product is well-defined and finite for each real number ξ. Thus the *Fourier transform for continuous functions of moderate decrease* is well-defined by the formula

$$\widehat{f}(\xi) := \int_{-\infty}^{\infty} f(x) e^{-2\pi i x\xi}\, dx.$$

Exercise 8.6. Show that the Fourier transform of the Fejér kernel F_R is $(1 - |\xi|/R)$ if $|\xi| \leq R$ and zero otherwise. Show that the Fourier transform of the Poisson kernel P_y is $e^{-2\pi|\xi|y}$ (see the project in Section 7.8). ◊

Exercise 8.7. Show that the convolution of two continuous functions f and g of moderate decrease is also a continuous function of moderate decrease. Show that $\widehat{f * g} = \widehat{f}\,\widehat{g}$. ◊

Exercise 8.8. Show that if f is a continuous function of moderate decrease and $\phi \in \mathcal{S}(\mathbb{R})$, then their convolution $f * \phi \in \mathcal{S}(\mathbb{R})$. ◊

Let f be a continuous function of moderate decrease. Then f is bounded, since $|f(x)| \leq A$ for all $x \in \mathbb{R}$. Hence f belongs to the space $L^\infty(\mathbb{R})$ of *essentially bounded* functions (see Section A.3.1), and $\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$. Also, f is *locally integrable on* $\mathbb{R}$, written $f \in L^1_{\text{loc}}(\mathbb{R})$, meaning that for each compact interval $[a, b]$ the integral $\int_a^b |f(x)|\, dx$ is finite. Finally, f lies in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, since the decay condition (8.1) guarantees that f is both integrable ($f \in L^1(\mathbb{R})$) and square integrable ($f \in L^2(\mathbb{R})$). Furthermore,

$$\|f\|_2^2 = \int_{-\infty}^{\infty} |f(x)|^2\,dx \leq \big(\sup_{x\in\mathbb{R}} |f(x)|\big) \int_{-\infty}^{\infty} |f(x)|\,dx = \|f\|_\infty \|f\|_1.$$

Exercise 8.9 (*Continuous Functions of Moderate Decrease Are in* $L^p(\mathbb{R})$**).** Check that continuous functions of moderate decrease are in $L^1(\mathbb{R})$ and indeed in every space $L^p(\mathbb{R})$. In other words, check that $\int_{-\infty}^{\infty} |f(x)|^p\,dx < \infty$, for $1 \leq p < \infty$. ♢

The Fourier transform is well defined for continuous functions of moderate decrease. The Approximation of the Identity Theorem (Theorem 7.35) can be generalized to continuous functions of moderate decrease. In other words, if f and the good kernels K_δ are continuous functions of moderate decrease, then the convolutions $f * K_\delta$ converge uniformly to f, and they converge in L^p to f for $1 \leq p < \infty$. Therefore, the inversion formula and Plancherel's Identity hold if f and $\widehat{f}$ are both continuous functions of moderate decrease. See [**SS03**, Sections 5.1.1 and 5.1.7] for more details.

In fact, in the proofs given in Chapter 7 we did not use the full power of the Schwartz class. What is really needed there is the fast decay (integrability) of the function, to ensure that integrals outside a large interval ($|x| > M$) can be made arbitrarily small for M sufficiently large. Inside the compact interval $|x| \leq M$, one can use uniform convergence to interchange the limit and the integral and take advantage of the integrand being small to guarantee that the integral over the compact interval is arbitrarily small. These arguments still apply if we assume the functions involved are continuous functions of moderate decrease instead of Schwartz functions. Similarly, the inversion formula holds when f and $\widehat{f}$ are both in $L^1(\mathbb{R})$. The proof uses Lebesgue integration theory together with theorems for interchanging limits and integrals, such as the Lebesgue Dominated Convergence Theorem A.59; see for example [**SS05**, Section 2.4].

Exercise 8.10. Verify that the Approximation of the Identity Theorem (Theorem 7.35) holds when the functions and kernels convolved are continuous functions of moderate decrease. ♢

Exercise 8.11. Verify that the inversion formula (7.22) and Plancherel's Identity (Theorem 7.31) hold when f and $\widehat{f}$ are both continuous functions of moderate decrease. ♢

In the project in Section 7.8 we saw that the *partial Fourier integrals* $S_R f(x) := \int_{-R}^{R} \widehat{f}(\xi) e^{2\pi i \xi x}\, d\xi$ are given by convolution with the Dirichlet kernel D_R on $\mathbb{R}$, that the Fejér kernel on $\mathbb{R}$ can be written as an integral mean of Dirichlet kernels,

$$F_R(x) = \frac{1}{R} \int_0^R D_t(x)\, dt,$$

and that the Fejér kernels define an approximation of the identity on $\mathbb{R}$ as $R \to \infty$. Therefore the *integral Cesàro means* of a function f of moderate decrease converge to f as $R \to \infty$; in other words,

$$\sigma_R f(x) := \frac{1}{R} \int_0^R S_t f(x)\, dt = F_R * f(x) \to f(x). \tag{8.3}$$

The convergence is both uniform and in L^p.

Notice the parallel with the results obtained for Fourier series when we discussed summability methods.

In Chapter 12, we prove that the partial Fourier sums $S_N f$ of f converge in $L^p(\mathbb{T})$ to f as $N \to \infty$, this time as a consequence of the boundedness of the periodic Hilbert transform in $L^p(\mathbb{T})$. The same is true in $\mathbb{R}$.

8.2. Tempered distributions

We begin with some definitions. A *linear functional* on $\mathcal{S}(\mathbb{R})$ is a linear mapping from $\mathcal{S}(\mathbb{R})$ to the complex numbers $\mathbb{C}$. We explain below the idea of continuity for linear functionals.

Definition 8.12. A *tempered distribution* T is a continuous linear functional from the space $\mathcal{S}(\mathbb{R})$ of Schwartz functions to the complex numbers:

$$T : \mathcal{S}(\mathbb{R}) \to \mathbb{C}.$$

The space of tempered distributions is denoted by $\mathcal{S}'(\mathbb{R})$. Further, two tempered distributions T and U are said to *coincide in the sense of distributions* if $T(\phi) = U(\phi)$ for all $\phi \in \mathcal{S}(\mathbb{R})$. A sequence of tempered distributions $\{T_n\}_{n\in\mathbb{N}}$ is said to *converge in the sense of distributions* to $T \in \mathcal{S}'(\mathbb{R})$ if for all $\phi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{n\to\infty} T_n(\phi) = T(\phi). \qquad \Diamond$$

The notation $\mathcal{S}'(\mathbb{R})$ is chosen to match the standard notation V' for the collection of all continuous linear functionals $T : V \to \mathbb{C}$ on a so-called topological vector space V. This collection V' is called the *dual* of V. The space of tempered distributions is the dual of the space of Schwartz functions.

To say that a functional T is *linear* means that for all a, $b \in \mathbb{C}$ and all ϕ, $\psi \in \mathcal{S}(\mathbb{R})$,

$$T(a\phi + b\psi) = aT(\phi) + bT(\psi).$$

If T is linear, then $T(0) = 0$.

Continuity requires a bit more explanation. If $\phi \in \mathcal{S}(\mathbb{R})$, then for each k, $\ell \in \mathbb{N}$ the quantities

$$\rho_{k,\ell}(\phi) := \sup_{x\in\mathbb{R}} |x|^k |\phi^{(\ell)}(x)| \tag{8.4}$$

are finite (Exercise 7.2). These quantities are *seminorms*, meaning that they are positive (although not necessarily positive definite, so $\rho_{k,\ell}(\phi) = 0$ does not imply $\phi = 0$), they are homogeneous ($\rho_{k,\ell}(\lambda\phi) = |\lambda|\rho_{k,\ell}(\phi)$), and they satisfy the Triangle Inequality. These seminorms enable us to define a notion of convergence in $\mathcal{S}(\mathbb{R})$.

Definition 8.13. A sequence $\{\phi_n\}_{n\in\mathbb{N}}$ *converges to* ϕ *in* $\mathcal{S}(\mathbb{R})$ if and only if $\lim_{n\to\infty} \rho_{k,\ell}(\phi_n - \phi) = 0$ for all k, $\ell \in \mathbb{N}$. $\Diamond$

Now we can define continuity for tempered distributions.

Definition 8.14. A *linear functional* T *on* $\mathcal{S}(\mathbb{R})$ *is continuous* if whenever a sequence $\{\phi_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}(\mathbb{R})$ converges in $\mathcal{S}(\mathbb{R})$ to some $\phi \in \mathcal{S}(\mathbb{R})$, the sequence $\{T(\phi_n)\}_{n\in\mathbb{N}}$ of complex numbers converges to $T(\phi)$ in $\mathbb{C}$. $\Diamond$

Exercise 8.15 (*It Suffices to Check Continuity at Zero*)**.** Show that for a *linear* functional it is enough to check convergence when $\phi = 0$: a linear functional T on $\mathcal{S}(\mathbb{R})$ is continuous if and only if $T(\phi_n) \to 0$ in $\mathbb{C}$ whenever $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$. $\Diamond$

Exercise 8.16. Show that if $\phi_n \to \phi$ in $\mathcal{S}(\mathbb{R})$, then $\phi_n^{(\ell)} \to \phi^{(\ell)}$ uniformly for all $\ell \in \mathbb{N}$. ♢

We show below that the delta distribution is continuous. Other examples are left as exercises for the reader.

The canonical example of a tempered distribution is the functional T_f given by integration against a reasonable function f:

$$T_f(\phi) := \int_{\mathbb{R}} f(x)\phi(x)\,dx \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}). \tag{8.5}$$

In this case we say for short that the distribution T *is* a function f.

How reasonable must the function f be? For the integral to be well-defined for all $\phi \in \mathcal{S}(\mathbb{R})$, f cannot grow too fast at infinity. For example, exponential growth could overwhelm the decay of a Schwartz function ϕ as $x \to \pm\infty$. However, polynomials or continuous functions that increase no faster than a polynomial are reasonable. Schwartz functions f are as reasonable as they could possibly be since they are rapidly decreasing, and continuous functions of moderate decrease are also reasonable.

Example 8.17 (*Bounded Continuous Functions Induce Tempered Distributions*)**.** If f is bounded and continuous, then the formula (8.5) for T_f defines a continuous linear functional on $\mathcal{S}(\mathbb{R})$. So T_f defines a tempered distribution: $T_f \in \mathcal{S}'(\mathbb{R})$.

In particular T_f is a tempered distribution if $f \in \mathcal{S}(\mathbb{R})$ or if f is a continuous function of moderate decrease.

First, for f bounded and continuous and for $\phi \in \mathcal{S}(\mathbb{R})$, the product $f\phi$ is integrable; hence $T_f(\phi)$ is a well-defined number. Linearity of the functional T_f follows from the linearity of the integral. To show that T_f is continuous, suppose $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$. Then in particular $\lim_{n\to\infty} \rho_{2,0}(\phi_n) = 0$; that is, there is some N_0 such that for all $n > N_0$, we have $|x|^2|\phi_n(x)| \le 1$. By choosing $M > 0$ sufficiently large we can guarantee that for all $n > N_0$,

$$\int_{|x|>M} |\,\phi_n(x)|\,dx \le \int_{|x|>M} x^{-2}\,dx < \varepsilon/2.$$

On the compact interval $[-M, M]$ the functions ϕ_n converge uniformly to zero (because $\lim_{n\to\infty} \rho_{0,0}(\phi_n) = 0$), and by Theorem 2.53

we can interchange the limit and the integral:

$$\lim_{n\to\infty} \int_{|x|\leq M} |\phi_n(x)|\,dx = \int_{|x|\leq M} \lim_{n\to\infty} |\phi_n(x)|\,dx = 0.$$

Therefore, given $\varepsilon > 0$, there is some $N_1 > 0$ so that for all $n > N_1$,

$$\int_{|x|\leq M} |\phi_n(x)|\,dx < \varepsilon/2.$$

Thus for $n > N := \max\{N_0, N_1\}$ we have $\int_{\mathbb{R}} |\phi_n(x)|\,dx \leq \varepsilon$.

Finally, T_f is continuous, since for $n > N$,

$$\begin{aligned} |T_f(\phi_n)| &\leq \int_{\mathbb{R}} |f(x)||\phi_n(x)|\,dx \\ &\leq \sup_{x\in\mathbb{R}} |f(x)| \int_{\mathbb{R}} |\phi_n(x)|\,dx \leq \sup_{x\in\mathbb{R}} |f(x)|\,\varepsilon, \end{aligned}$$

and therefore $\lim_{n\to\infty} T_f(\phi_n) = 0$. ◊

For example, the distribution T_1 induced by the bounded function $f \equiv 1$ is a well-defined tempered distribution.

Exercise 8.18 (*Polynomials Induce Tempered Distributions*)**.** Show that the linear functional defined by integration against the polynomial $f(x) = a_0 + a_1 x + \cdots + a_k x^k$ is a tempered distribution. ◊

Example 8.17 tells us that we can view $\mathcal{S}(\mathbb{R})$ as sitting inside $\mathcal{S}'(\mathbb{R})$, via the one-to-one mapping $\varepsilon : \mathcal{S}(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ given by

$$\varepsilon(f) := T_f \qquad \text{for } f \in \mathcal{S}(\mathbb{R}).$$

The mapping ε is continuous, in the sense that if f_n, $f \in \mathcal{S}(\mathbb{R})$ and $f_n \to f$ in $\mathcal{S}(\mathbb{R})$, then $T_{f_n}(\phi) \to T_f(\phi)$ in $\mathbb{C}$ for all $\phi \in \mathcal{S}(\mathbb{R})$, so that T_{f_n} converges to T_f in the sense of distributions. Such continuous and one-to-one mappings are sometimes called *embeddings*, as in Section 2.4.

Exercise 8.19. Show that the mapping $\varepsilon : \mathcal{S}(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ given by $\varepsilon(f) := T_f$ for $f \in \mathcal{S}(\mathbb{R})$ is one-to-one. ◊

Exercise 8.20. Prove that if f_n, $f \in \mathcal{S}(\mathbb{R})$ and $f_n \to f$ in $\mathcal{S}(\mathbb{R})$, then $T_{f_n}(\phi) \to T_f(\phi)$ in $\mathbb{C}$ for all $\phi \in \mathcal{S}(\mathbb{R})$. ◊

Like bounded functions, unbounded functions that have some integrability properties can also induce tempered distributions via integration, as we show in Section 8.6 for L^p functions. Some tempered distributions are not induced by any function, however, such as the delta distribution (Section 8.4) and the *principal value distribution* $1/x$ (the project in Section 8.7).

In Example 8.17 we used the fact that if $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, then $\|\phi_n\|_{L^1} \to 0$. By a similar argument, the L^p norms also go to zero.

Exercise 8.21. Verify that if $\phi_n \in \mathcal{S}(\mathbb{R})$ and $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, then for all real p with $1 \leq p \leq \infty$, $\|\phi_n\|_p^p := \int_{\mathbb{R}} |\phi_n(x)|^p \, dx \to 0$. ◊

The continuous linear functionals that act on the space of compactly supported C^∞ functions are called *distributions* (unfortunately, this term is easy to confuse with the *tempered distributions* $\phi \in \mathcal{S}'(\mathbb{R})$). We denote by $\mathcal{D}(\mathbb{R})$ the space of compactly supported C^∞ functions. Then $\mathcal{D}(\mathbb{R}) \subset \mathcal{S}(\mathbb{R})$. The space of distributions is the dual $\mathcal{D}'(\mathbb{R})$ of $\mathcal{D}(\mathbb{R})$. Tempered distributions are distributions, so these two spaces are nested: $\mathcal{S}'(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$. Although we sometimes write *distributions* for short, in this book we always mean tempered distributions. The book by Robert Strichartz [**Stri94**] is a delightful introduction to distribution theory, including many applications, at a level comparable to that of this book.

8.3. The time–frequency dictionary for $\mathcal{S}'(\mathbb{R})$

In this section we show how to extend from functions to distributions the fundamental operations discussed so far: translation, dilation, differentiation, the Fourier and inverse Fourier transforms, multiplication by Schwartz functions and by polynomials, and convolution by Schwartz functions. We assume without proof that these extensions of the fundamental operations, and the Fourier transform, are nicely behaved with respect to the notion of continuity in $\mathcal{S}'(\mathbb{R})$.

The key idea is that since tempered distributions are defined by the way they act on the space of Schwartz functions, it makes sense to define an operation on a tempered distribution by pushing the operation across to act on the Schwartz function.

We illustrate with the example of differentiation. When T is of the form T_f for some $f \in \mathcal{S}(\mathbb{R})$, we want its derivative T'_f to be given by the tempered distribution corresponding to the derivative f' of f. Integration by parts gives

$$(8.6) \qquad T_{f'}(\phi) = \int_{\mathbb{R}} f'(x)\phi(x)\,dx = -\int_{\mathbb{R}} f(x)\phi'(x)\,dx = -T_f(\phi').$$

Thus we can define the new distribution T'_f, representing the derivative of T_f, by setting

$$T'_f(\phi) = T_{f'}(\phi) = -T_f(\phi').$$

In this sense, we have pushed the differentiation across to the Schwartz function ϕ. We emphasize, though, that $T'_f(\phi)$ is the *negative* of $T_f(\phi')$; the derivative did not simply migrate unchanged to ϕ.

Next, for a general distribution T, we use the formula that we were led to in the special case $T = T_f$. Namely, we omit the middle of the string of equalities (8.6) and define the *derivative of* T by

$$T'(\phi) := -T(\phi').$$

Given a tempered distribution $T : \mathcal{S}(\mathbb{R}) \to \mathbb{C}$, we use this same idea of pushing the operator across to create nine new tempered distributions defined by their action on functions $\phi \in \mathcal{S}(\mathbb{R})$, as shown in Table 8.1. The new functionals defined in Table 8.1 are linear, and it can be shown that they are also continuous. So we have indeed defined new tempered distributions, starting from a given tempered distribution T.

The function g in the multiplication in item (f) could be a function in the Schwartz class, a bounded function such as the trigonometric functions $e_\xi(x) = e^{2\pi i x\xi}$, or an unbounded function that does not increase too fast, such as a polynomial. Multiplication by the trigonometric function e_ξ is denoted by $M_\xi T := \mathcal{M}_{e_\xi} T$ and is called *modulation*.

Exercise 8.22. Verify that if $T \in \mathcal{S}'(\mathbb{R})$, then the nine linear functionals defined in Table 8.1 are in $\mathcal{S}'(\mathbb{R})$. Notice that verifying continuity reduces to verifying that for example, in the case of (a), if $\phi_n \to \phi$ in $\mathcal{S}(\mathbb{R})$, then $\tau_{-h}\phi_n \to \tau_{-h}\phi$ in $\mathcal{S}(\mathbb{R})$, etc. ♢

Table 8.1. New tempered distributions from old. Here $\phi \in \mathcal{S}(\mathbb{R})$, $T \in \mathcal{S}'(\mathbb{R})$, $h \in \mathbb{R}$, $s > 0$, g is a function such that $g\phi \in \mathcal{S}(\mathbb{R})$, and $\psi \in \mathcal{S}(\mathbb{R})$. FT means Fourier transform.

	Operation on $\mathcal{S}(\mathbb{R})$	**Operation on $\mathcal{S}'(\mathbb{R})$**
(a)	translation $\tau_h\phi(x) = \phi(x-h)$	translate $\tau_h T$ of T $\tau_h T(\phi) := T(\tau_{-h}\phi)$
(b)	dilation $D_s\phi(x) = s\phi(sx)$	dilation D_sT of T $D_sT(\phi) := T(sD_{s^{-1}}\phi)$
(c)	derivative $\phi'(x) = (d/dx)\phi(x)$	derivative T' of T $T'(\phi) := -T(\phi')$
(d)	Fourier transform $\widehat{\phi}(\xi) = \int_{-\infty}^{\infty} \phi(x)e^{-ix\xi}dx$	Fourier transform $\widehat{T}$ of T $\widehat{T}(\phi) := T(\widehat{\phi})$
(e)	inverse FT $\phi^\vee(\xi) = \int_{-\infty}^{\infty} \phi(x)e^{ix\xi}dx$	inverse FT $T^\vee$ of T $T^\vee(\phi) := T(\phi^\vee)$
(f)	product with g $\mathcal{M}_g\phi(x) = g(x)\phi(x)$	product $\mathcal{M}_gT$ of T and g $\mathcal{M}_gT(\phi) := T(\mathcal{M}_g\phi)$
(g)	reflection $\widetilde{\phi}(x) = \phi(-x)$	reflection $\widetilde{T}$ of T $\widetilde{T}(\phi) := T(\widetilde{\phi})$
(h)	conjugate $\overline{\phi}(x) = \overline{\phi(x)}$	conjugate $\overline{T}$ of T $\overline{T}(\phi) := T(\overline{\phi})$
(i)	convolution $\psi * \phi(x) = \int_{-\infty}^{\infty} \psi(x-y)\phi(y)dy$	convolution $\psi * T$ of T, ψ $\psi * T(\phi) := T(\widetilde{\psi} * \phi)$

The mapping $f \mapsto T_f$ interacts as one would expect with the operations of translation of f, dilation of f, and so on. For example, the translate $\tau_h T_f$ of T_f is the same as the tempered distribution $T_{\tau_h f}$ induced by the translate $\tau_h f$ of f: $\tau_h T_f = T_{\tau_h f}$.

Note that in lines (d)–(h) of Table 8.1, the operation migrates unchanged from T to ϕ, while in each of lines (a)–(c) and (i) the operation suffers some change on the way. For example, for the Fourier transform $\widehat{T}(\phi) := T(\widehat{\phi})$, while for the translation $\tau_h T(\phi) := T(\tau_{-h}\phi)$. The interested reader may like to investigate further; the key point is that the operations in (d)–(h) are what is called *selfadjoint*, while those in the other lines are not.

Example 8.23 and Exercise 8.24 offer some practice in working with the definition of tempered distributions, as we work out the details of this interaction for each of the nine new distributions.

Example 8.23 (*The Mapping $f \mapsto T_f$ Respects Convolution*)**.** We verify that for Schwartz functions f and ψ, the convolution of ψ with the tempered distribution T_f is the same as the tempered distribution induced by the convolution of ψ with f:

$$\psi * T_f = T_{\psi * f} \qquad \text{for all } f \in \mathcal{S}(\mathbb{R}),\ \psi \in \mathcal{S}(\mathbb{R}).$$

For f, ψ, and $\phi \in \mathcal{S}(\mathbb{R})$,

$$\begin{aligned} T_{\psi * f}(\phi) &= \int_{\mathbb{R}} \phi(y)(\psi * f)(y)\,dy \\ &= \int_{\mathbb{R}} \phi(y) \left(\int_{\mathbb{R}} f(x)\psi(y-x)\,dx \right) dy \\ &= \int_{\mathbb{R}} f(x) \left(\int_{\mathbb{R}} \widetilde{\psi}\,(x-y)\phi(y)\,dy \right) dx \\ &= \int_{\mathbb{R}} f(x)(\widetilde{\psi} * \phi)(x)\,dx \;=\; T_f(\widetilde{\psi} * \phi) \;=\; (\psi * T_f)(\phi), \end{aligned}$$

where in the third equality we interchanged the order of integration and replaced $\psi(y-x)$ by $\widetilde{\psi}\,(x-y)$. ♢

Notice that the above calculation leads us to the correct definition of $\psi * T(\phi) = T(\widetilde{\psi} * \phi)$. It turns out that the object obtained by convolution of $\psi \in \mathcal{S}(\mathbb{R})$ with $T \in \mathcal{S}'(\mathbb{R})$ can be identified with an infinitely differentiable function that grows slower than some polynomial. Specifically, $f(y) = T(\tau_y \widetilde{\psi}\,)$ [**Graf08**, pp. 116–117].

Exercise 8.24 (*The Mapping $f \mapsto T_f$ Respects the Fundamental Operations*)**.** Show that for all $f \in \mathcal{S}(\mathbb{R})$, $h \in \mathbb{R}$, $s > 0$, and $g \in C^\infty(\mathbb{R})$ such that $gf \in \mathcal{S}(\mathbb{R})$ for all $f \in \mathcal{S}(\mathbb{R})$, we have

$$\tau_h(T_f) = T_{\tau_h f}, \quad D_s(T_f) = T_{D_s f}, \quad \widehat{T_f} = T_{\widehat{f}}, \quad (T_f\,)^\vee = T_{(f)^\vee},$$

$$\widetilde{T_f} = T_{\widetilde{f}}, \quad \overline{T_f} = T_{\overline{f}}, \quad \text{and} \quad (\mathcal{M}_g T_f) = T_{\mathcal{M}_g f}. \qquad ♢$$

Table 8.2. The time–frequency dictionary in $\mathcal{S}'(\mathbb{R})$. Here T, $U \in \mathcal{S}'(\mathbb{R})$, a, b, $h \in \mathbb{R}$, $s > 0$, x, $\xi \in \mathbb{R}$, and ϕ, $\psi \in \mathcal{S}(\mathbb{R})$.

	Time	**Frequency**
(a)	linear properties $aT + bU$	linear properties $a\widehat{T} + b\widehat{U}$
(b)	translation $\tau_h T(\phi) := T(\tau_{-h}\phi)$	modulation $\widehat{\tau_h T} = M_{-h}\widehat{T}$
(c)	modulation by $e_h(x) = e^{2\pi i h x}$ $M_h T(\phi) := T(M_h\phi)$	translation $\widehat{M_h T} = \tau_h \widehat{T}$
(d)	dilation $D_s T(\phi) := T(sD_{s^{-1}}\phi)$	inverse dilation $\widehat{D_s T} = sD_{s^{-1}}\widehat{T}$
(e)	reflection $\widetilde{T}(\phi) := T(\widetilde{\phi})$	reflection $\widehat{\widetilde{T}} = -\widetilde{\widehat{T}}$
(f)	conjugate $\overline{T}(\phi) := T(\overline{\phi})$	conjugate reflection $\widehat{\overline{T}} = \overline{[T]^{\vee}}$
(g)	derivative $T'(\phi) := -T(\phi')$	multiply by polynomial $\widehat{T'} = [2\pi i\xi]\widehat{T} = \mathcal{M}_{2\pi i\xi}\widehat{T}$
(h)	multiply by polynomial $\mathcal{M}_{-2\pi i x}T(\phi) := T(-2\pi i x\phi(x))$	derivative $[\mathcal{M}_{-2\pi i x}T]^{\wedge} = (\widehat{T})'$
(i)	convolution $\psi * T(\phi) := T(\widetilde{\psi} * \phi)$	product $\widehat{\psi * T} = \mathcal{M}_{\widehat{\psi}}\widehat{T} = \widehat{\psi}\,\widehat{T}$

Definitions (d) and (e) in Table 8.1 show that the Fourier transform can be extended to be a bijection on $\mathcal{S}'(\mathbb{R})$, because it is a bijection on $\mathcal{S}(\mathbb{R})$. To see this, observe that for all $\phi \in \mathcal{S}(\mathbb{R})$, $[\widehat{T}]^{\vee}(\phi) = T([\widehat{\phi}]^{\vee}) = T(\phi)$. Thus $[\widehat{T}]^{\vee} = T$. Similarly $\widehat{[T]^{\vee}} = T$.

We can now build a time–frequency dictionary for $\mathcal{S}'(\mathbb{R})$, displayed in Table 8.2, that is in one-to-one correspondence with the time–frequency dictionary we built for $\mathcal{S}(\mathbb{R})$ (Table 7.1).

In the time column, we recall the definitions of the basic operations with tempered distributions; in the frequency column, we recall how they interact with the Fourier transform.

Example 8.25. As an example of how to derive the time–frequency dictionary encoded in Table 8.2, we derive (i), the multiplication formula for the convolution. For ψ, $\phi \in \mathcal{S}(\mathbb{R})$, apply the definition of the Fourier transform of a distribution and the definition of convolution with a distribution to obtain

$$\widehat{\psi * T}(\phi) = (\psi * T)(\widehat{\psi}) = T(\widetilde{\psi} * \widehat{\phi}). \tag{8.7}$$

Let us look more closely at the argument on the right-hand side of equation (8.7). We see that it is the Fourier transform of $\widehat{\psi}\phi$:

$$\begin{aligned}
\widetilde{\psi} * \widehat{\phi}(\xi) &= \int_{\mathbb{R}} \widetilde{\psi}(s)\widehat{\phi}(\xi - s)\, ds \\
&= \int_{\mathbb{R}} \widetilde{\psi}(s) \left(\int_{\mathbb{R}} \phi(x) e^{-2\pi i x(\xi - s)}\, dx \right) ds \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \psi(-s) e^{-2\pi i x(-s)}\, ds \right) \phi(x) e^{-2\pi i x \xi}\, dx \\
&= \int_{\mathbb{R}} \widehat{\psi}(x)\phi(x) e^{-2\pi i x \xi}\, dx \; = \; \widehat{\widehat{\psi}\phi}\,(\xi).
\end{aligned}$$

Now we can continue, using again the definition of the Fourier transform of a distribution and the definition of multiplication of a distribution by the function $\widehat{\psi}$. We see that for all $\psi \in \mathcal{S}(\mathbb{R})$,

$$\widehat{\psi * T}(\phi) = T(\widehat{\widehat{\psi}\phi}) = \widehat{T}(\widehat{\psi}\phi) = \mathcal{M}_{\widehat{\psi}}\widehat{T}(\phi).$$

This is exactly the multiplication formula we were seeking:

$$\widehat{\psi * T} = \mathcal{M}_{\widehat{\psi}}\widehat{T} = \widehat{\psi}\,\widehat{T}. \qquad \lozenge$$

Exercise 8.26. Derive the formulas in the frequency column of Table 8.2. It will be useful to consult Table 7.1, the corresponding time–frequency dictionary for the Schwartz class. $\lozenge$

8.4. The delta distribution

There are objects in $\mathcal{S}'(\mathbb{R})$ other than functions. The *delta distribution* is the canonical example of a tempered distribution that is not a function. Informally, the delta distribution takes the value infinity at $x = 0$ and zero everywhere else. Rigorously, the delta distribution is the Fourier transform of the tempered distribution induced by the function that is identically equal to one. Computationally, the delta distribution has a very simple effect on Schwartz functions: it evaluates them at the point 0, as we now show.

By Example 8.17, the bounded function $f(x) \equiv 1$ induces a tempered distribution T_1 via integration against the function 1:

$$T_1(\phi) := \int_{\mathbb{R}} \phi(x)\,dx \qquad \text{for } \phi \in \mathcal{S}(\mathbb{R}).$$

The Fourier transform of T_1 is the new tempered distribution $\widehat{T_1}$ given by

$$\widehat{T_1}(\phi) := T_1(\widehat{\phi}\,) = \int_{\mathbb{R}} \widehat{\phi}\,(\xi)\,d\xi = \int_{\mathbb{R}} \widehat{\phi}\,(\xi)e^{2\pi i\xi 0}\,d\xi = (\widehat{\phi}\,)^{\vee}(0) = \phi(0)$$

for $\phi \in \mathcal{S}(\mathbb{R})$. The last equality holds by the Fourier Inversion Formula for Schwartz functions (Theorem 7.28).

Definition 8.27. The *delta distribution* $\delta : \mathcal{S}(\mathbb{R}) \to \mathbb{C}$ is defined by

$$\delta(\phi) := \widehat{T_1}(\phi\,) = \phi(0) \qquad \text{for } \phi \in \mathcal{S}(\mathbb{R}). \qquad \diamondsuit$$

We verify that δ is a tempered distribution. It is linear, since $(a\phi_1+b\phi_2)(0) = a\phi_1(0)+b\phi_2(0)$. To check continuity, suppose $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$. Then in particular $\rho_{0,0}(\phi_n) = \sup_{x\in\mathbb{R}} |\phi_n(x)| \to 0$. Since $|\phi_n(0)| \leq \rho_{0,0}(\phi_n)$, we have $\delta(\phi_n) = \phi_n(0) \to 0$, as required.

What is the Fourier transform $\widehat{\delta}$ of δ? For $\phi \in \mathcal{S}(\mathbb{R})$,

$$\widehat{\delta}\,(\phi) = \delta(\widehat{\phi}\,) = \widehat{\phi}\,(0) = \int_{\mathbb{R}} \phi(x)\,dx = T_1(\phi).$$

Thus the Fourier transform $\widehat{\delta}$ is the tempered distribution T_1 induced by the function $f(x) \equiv 1$. We write $\widehat{\delta} = T_1$, or just $\widehat{\delta} = 1$ for short.

Exercise 8.28. Show that the *inverse* Fourier transform of δ is also given by T_1: $(\delta)^{\vee}(\phi) = T_1(\phi) = \widehat{\delta}\,(\phi)$ for all $\phi \in \mathcal{S}(\mathbb{R})$. $\diamondsuit$

Exercise 8.29. Verify that the Fourier transform of $\tau_n\delta$ is the tempered distribution induced by the function $e_n(x) = e^{-2\pi inx}$. $\diamondsuit$

Why isn't the delta distribution a true function? For a moment let us pretend that there is a genuine function $\delta(x)$, called the "*delta function*", that induces the delta distribution by integration. Then

$$\int_{\mathbb{R}} \delta(x)\phi(x)\,dx = \phi(0) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}). \tag{8.8}$$

In particular, $\int_{\mathbb{R}} \delta(x)\phi(x)\,dx = 0$ for all $\phi \in \mathcal{S}(\mathbb{R})$ such that $\phi(0) = 0$. So $\delta(x)$ must be zero for most x; if for instance $\delta(x)$ were positive on some interval not including the point $x = 0$, we could build a Schwartz function supported on that interval whose integral against $\delta(x)$ would not be zero, contradicting equation (8.8). One can use measure theory to show that in fact $\delta(x) = 0$ for all x except $x = 0$.

Now, what value can we assign to $\delta(0)$ so that integration against $\delta(x)$ will pick out the value of ϕ at 0? If we assign a finite value to $\delta(0)$, then $\int_{\mathbb{R}} \delta(x)\phi(x)\,dx = 0$ instead of the $\phi(0)$ we want. Therefore $\delta(0)$ cannot be a finite number, and so $\delta(0)$ must be infinite. Thus the delta function is not a true function.

Fortunately, all is not lost: one can interpret δ as a distribution as we do here, or as a *point-mass measure* in the language of measure theory [**Fol**].

It is interesting that the support of the delta distribution is as small as it could possibly be without being the empty set, while the support of its Fourier transform is as large as it could possibly be. Compare this point with the Uncertainty Principle (Section 8.5.3).

Exercise 8.30 (*The Derivatives and the Antiderivative of the Delta Distribution*)**.** Find the derivatives $(d^k/dx^k)\delta$, $k \geq 1$, of the delta distribution. Check that their Fourier transforms $((d^k/dx^k)\delta)^\wedge$ can be identified with the polynomials $(2\pi i x)^k$. Check that δ coincides, in the sense of distributions, with the derivative H' of the *Heaviside function* H defined by $H(x) = 0$ if $x \leq 0$, $H(x) = 1$ if $x > 0$. ◊

Remark 8.31. We see that although the delta distribution is not a function, it can be seen as the derivative of the Heaviside function. Similarly, it turns out that for *each tempered distribution* T, no matter how wild, there is some function f such that T is the derivative of some order (in the sense of distributions) of f. Another example is the distribution given by the k^{th} derivative of the delta distribution.

By Exercise 8.30, $(d^k/dx^k)\delta$ coincides with the $(k+1)^{\text{st}}$ derivative of the Heaviside function. See [**Stri94**, Section 6.2]. $\diamondsuit$

We now return to the ideas at the start of Section 7.5, where we lamented that there is no identity element in $\mathcal{S}(\mathbb{R})$ and suggested that the delta distribution, and more generally an approximation to the identity, could act in that rôle.

Exercise 8.32 (*The Delta Distribution Acts as an Identity Under Convolution*)**.** Show that the delta distribution acts as an identity for the operation of convolution: for all $\phi \in \mathcal{S}(\mathbb{R})$, $\phi * \delta = \phi$, meaning as usual that as distributions, $\phi * \delta = T_\phi$. $\diamondsuit$

Exercise 8.33 (*Approximations of the Identity Converge to the Delta Distribution*)**.** Let $\{K_t(x)\}_{t\in\Lambda}$, with t_0 a limit point of Λ, be an approximation of the identity in $\mathbb{R}$. Prove that as $t \to t_0$, the distribution induced by K_t converges in the sense of distributions to the delta distribution: for all $\phi \in \mathcal{S}(\mathbb{R})$, $\lim_{t\to t_0} T_{K_t}(\phi) = \delta(\phi) = \phi(0)$. $\diamondsuit$

For those who have seen some measure theory, we note that every finite Borel measure μ on $\mathbb{R}$ defines a tempered distribution T_μ, via integration, similarly to the way in which equation (8.5) defines the tempered distribution induced by a function: $T_\mu(\phi) := \int_{\mathbb{R}} \phi(x)\, d\mu$ for $\phi \in \mathcal{S}(\mathbb{R})$. This formula embeds the finite Borel measures into $\mathcal{S}'(\mathbb{R})$.

8.5. Three applications of the Fourier transform

We present here three of the most classical applications of the Fourier transform. The formulas and inequalities discussed in the next pages have remarkable consequences in a range of subjects: from number theory (the Poisson Summation Formula), to signal processing (the Whittaker–Shannon Sampling Formula), to quantum mechanics (the Heisenberg Uncertainty Principle). The proofs are beautiful and elegant. For many other applications to physics, partial differential equations, probability, and more, see the books by Strichartz [**Stri94**], Dym and McKean [**DM**], Körner [**Kör**], Stein and Shakarchi [**SS03**], and Reed and Simon [**RS**].

8.5.1. The Poisson Summation Formula. Given a function f on the line, how can we construct from f a periodic function of period 1?

There are two methods. The first, called periodization, uses a sum over the integer translates of the function f, and the second uses a Fourier series whose Fourier coefficients are given by the Fourier transform of f evaluated at the integers. The Poisson Summation Formula gives conditions under which these two methods coincide.

The *periodization* of a function ϕ on the line is constructed by summing over its integer translates:

$$P_1\phi(x) = \sum_{n\in\mathbb{Z}} \phi(x+n) = \lim_{N\to\infty} \sum_{|n|\le N} \phi(x+n),$$

where the series is supposed to converge in the sense of symmetric partial sums. When ϕ is in $\mathcal{S}(\mathbb{R})$ or is a continuous function of moderate decrease, the sum in the periodization is absolutely convergent.

Exercise 8.34. Verify that for $\phi \in \mathcal{S}(\mathbb{R})$ the function $P_1\phi$ is periodic with period 1. $\diamondsuit$

Example 8.35. Let H_t be the *heat kernel* on the line, defined by formula (7.32). Its periodization is the *periodic heat kernel*

$$P_1 H_t(x) = (4\pi t)^{-1/2} \sum_{n\in\mathbb{Z}} e^{-(x+n)^2/4t}. \qquad \diamondsuit$$

The second method assigns to a given function $\phi \in \mathcal{S}(\mathbb{R})$ a periodic function of period 1 defined by

$$P_2\phi(x) = \sum_{n\in\mathbb{Z}} \widehat{\phi}(n) e^{2\pi i x n}.$$

The series is well-defined and convergent because $\phi \in \mathcal{S}(\mathbb{R})$.

Example 8.36. If $f \in L^1(0,1)$ and if it is defined to be zero outside $(0,1)$, then both periodizations are well-defined. They coincide with the periodic extension of f to the whole line. In particular, $P_1 f = P_2 f$. $\diamondsuit$

The phenomenon shown in this example is not an accident. The *Poisson Summation Formula*[1] states the idea precisely.

[1] Named after the same Poisson as the Poisson kernel in Chapter 4.

Theorem 8.37 (Poisson Summation Formula)**.** *If $\phi \in \mathcal{S}(\mathbb{R})$, then*

$$P_1\phi(x) = \sum_{n\in\mathbb{Z}} \phi(x+n) = \sum_{n\in\mathbb{Z}} \widehat{\phi}(n)e^{2\pi inx} = P_2\phi(x). \tag{8.9}$$

In particular, setting $x = 0$,

$$\sum_{n\in\mathbb{Z}} \phi(n) = \sum_{n\in\mathbb{Z}} \widehat{\phi}(n). \tag{8.10}$$

Proof. It suffices to check that the right- and left-hand sides of equation (8.9) have the same Fourier coefficients. The n^{th} Fourier coefficient on the right-hand side is $\widehat{\phi}(n)$. Computing the n^{th} Fourier coefficient for the left-hand side, we get

$$\begin{aligned}\int_0^1 \Big(\sum_{m\in\mathbb{Z}} \phi(x+m)\Big)e^{-2\pi inx}\,dx &= \sum_{m\in\mathbb{Z}} \int_0^1 \phi(x+m)e^{-2\pi inx}\,dx \\ = \sum_{m\in\mathbb{Z}} \int_m^{m+1} \phi(y)e^{-2\pi iny}\,dy &= \int_{-\infty}^{\infty} \phi(y)e^{-2\pi iny}\,dy = \widehat{\phi}(n).\end{aligned}$$

We may interchange the sum and the integral in the first step since ϕ is rapidly decreasing. □

Example 8.38. For the heat kernel H_t, all the conditions in the Poisson Summation Formula are satisfied. Furthermore $\widehat{H_t}(n) = e^{-4\pi^2 tn^2}$, and so the periodic heat kernel has a natural Fourier series in terms of separated solutions of the heat equation:

$$(4\pi t)^{-1/2} \sum_{n\in\mathbb{Z}} e^{-(x+n)^2/4t} = \sum_{n\in\mathbb{Z}} e^{-4\pi^2 tn^2} e^{2\pi inx}. \qquad \Diamond$$

The above calculation can be justified for continuous functions of moderate decrease, and so the Poisson Summation Formula also holds for that larger class of functions.

Exercise 8.39. Check that the Poisson Summation Formula is valid for continuous functions of moderate decrease. Use this result to show that the periodizations of the Poisson and Fejér kernels on the line (the project in Section 7.8) coincide with the 1-periodic Poisson and Fejér kernels, introduced in Sections 4.4 and 4.5. ◊

We give an interpretation of the Poisson Summation Formula in the language of distributions. Recall that $\tau_n\delta(\phi) = \phi(n)$ for all $\phi \in \mathcal{S}(\mathbb{R})$. The left-hand side of (8.10) can be written as

$$\sum_{n\in\mathbb{Z}} \phi(n) = \sum_{n\in\mathbb{Z}} \tau_n\delta(\phi).$$

Similarly, for the right-hand side, $\sum_{n\in\mathbb{Z}} \widehat{\phi}(n) = \sum_{n\in\mathbb{Z}} \widehat{\tau_n\delta}(\phi)$, where in the last equality we have used the definition of the Fourier transform for distributions. What the Poisson Summation Formula says is that the Fourier transform of the distribution $\sum_{n\in\mathbb{Z}} \tau_n\delta$ is the same distribution, that is, $(\sum_{n\in\mathbb{Z}} \tau_n\delta)^\wedge = \sum_{n\in\mathbb{Z}} \tau_n\delta$.

Exercise 8.40. Verify that $\sum_{n\in\mathbb{Z}} \tau_n\delta$ is a tempered distribution and that it coincides with its own Fourier transform. ♢

Notice that the distribution above is like an infinite comb with point-masses at the integers. Its Fourier transform can be calculated directly, by noticing that $\widehat{\tau_n\delta}$ is the distribution induced by the function $e^{-2\pi inx}$ (Exercise 8.29), and so the apparently very complicated distribution $\sum_{n\in\mathbb{Z}} e^{2\pi inx}$ is simply the infinite comb $\sum_{n\in\mathbb{Z}} \tau_n\delta$.

See [**Stri94**, Section 7.3] for a beautiful application of the Poisson Summation Formula in two dimensions to crystals and quasicrystals.

8.5.2. The Sampling Formula.

The *Whittaker–Shannon Sampling Formula*[2] states that *band-limited functions* (functions whose Fourier transforms are supported on a compact interval) can be perfectly reconstructed from appropriate samples. We will need the *kernel function* sinc, defined by

$$\operatorname{sinc}(x) := \sin(\pi x)/(\pi x) \quad \text{for } x \neq 0; \quad \operatorname{sinc}(0) := 1.$$

Theorem 8.41 (Whittaker–Shannon Sampling Formula)**.** *Suppose f is a continuous function of moderate decrease whose Fourier transform is supported on the interval $[-1/2, 1/2]$. Then the values $f(n)$ at the integers completely determine f; more precisely,*

$$f(x) = \sum_{n\in\mathbb{Z}} f(n) \operatorname{sinc}(x - n).$$

[2]Named after British mathematician Edmund Taylor Whittaker (1873–1956) and American mathematician and electronic engineer Claude Elwood Shannon (1916–2001).

Proof. The support of $\widehat{f}$ is the interval $[-1/2, 1/2]$. Consider the periodization of $\widehat{f}$ to $\mathbb{R}$, as a 1-periodic function. We compute its Fourier coefficients:

$$(\widehat{f}\,)^\wedge(n) = \int_{-1/2}^{1/2} \widehat{f}\,(x) e^{-2\pi i n x}\, dx = \int_{-\infty}^{\infty} \widehat{f}\,(x) e^{2\pi i(-n)x}\, dx = f(-n).$$

Here $\widehat{f}$ is the Fourier transform in $\mathbb{R}$, and we have used the hat notation to also denote the Fourier coefficients of a 1-periodic function $(g)^\wedge(n)$. We have also used the fact that $\widehat{f}$ is supported on $[-1/2, 1/2]$ and, for the last equality, the Fourier Inversion Formula on $\mathbb{R}$. (The Fourier Inversion Formula does hold for continuous functions f of moderate decrease whose Fourier transforms $\widehat{f}$ are also of moderate decrease; we omit the proof.)

Therefore $\widehat{f}$ coincides with its Fourier series for $\xi \in [-1/2, 1/2]$:

$$\widehat{f}\,(\xi) = \sum_{n\in\mathbb{Z}} f(-n) e^{2\pi i n \xi} = \sum_{n\in\mathbb{Z}} f(n) e^{-2\pi i n \xi}. \tag{8.11}$$

The Fourier Inversion Formula on $\mathbb{R}$ also tells us that we can recover f by integrating over $\mathbb{R}$ its Fourier transform $\widehat{f}$ times $e^{2\pi i \xi n}$. Using the compact support of $\widehat{f}$ to replace the integral over $\mathbb{R}$ by the integral over $[-2\pi, 2\pi]$ and substituting in formula (8.11), we find that

$$\begin{aligned} f(x) &= \int_{-1/2}^{1/2} \widehat{f}\,(\xi) e^{2\pi i \xi x}\, d\xi = \int_{-1/2}^{1/2} \sum_{n\in\mathbb{Z}} f(n) e^{-2\pi i n\xi} e^{2\pi i \xi x}\, d\xi \\ &= \sum_{n\in\mathbb{Z}} f(n) \int_{-1/2}^{1/2} e^{2\pi i (x-n)\xi}\, d\xi = \sum_{n\in\mathbb{Z}} f(n)\, \operatorname{sinc}(x-n), \end{aligned}$$

as required. The interchange of sum and integral is justified because of the uniform convergence in ξ of $g_N(x,\xi)$; see Exercise 8.42.

The last equality holds because $\operatorname{sinc}(x) = \int_{-1/2}^{1/2} e^{2\pi i x\xi}\, d\xi$. $\square$

Exercise 8.42. Use the Weierstrass M-test to show that for each x the functions $g_N(x,\xi) := \sum_{|n|\le N} f(n) e^{2\pi i(x-n)\xi}$ converge uniformly in $\xi \in [-1/2, 1/2]$ as $N \to \infty$, whenever the function f is a continuous function of moderate decrease. $\Diamond$

The support of the Fourier transform need not have length 1. As long as the Fourier transform is supported on a compact interval, say of length L, we can carry out similar calculations. We would use the Fourier theory for L-periodic functions; see Section 1.3.2.

Exercise 8.43. Suppose that f is a continuous function of moderate decrease and $\operatorname{supp}\widehat{f} \subset [-L/2, L/2]$. Show that

$$f(x) = \sum_{n\in\mathbb{Z}} f(n/L)\ \operatorname{sinc}\big((x-n)/L\big). \qquad \Diamond$$

The larger the support of $\widehat{f}$, the more we have to sample.

This statement is reasonable, for if only those frequencies with $|\xi| < L/2$ are present in f, then it is plausible that sampling at the $1/L$ rate will suffice. Sampling at a lower rate would not suffice.

This formula has many applications in signal processing (see [**Sha**]).

8.5.3. The Heisenberg Uncertainty Principle.

It is impossible to find a function that is simultaneously well-localized in time and in frequency. Heisenberg's Uncertainty Principle[3] is an inequality that expresses this principle. The idea can be carried further: there is no square-integrable function that is both compactly supported and band-limited. See [**SS03**, Section 5.4].

Theorem 8.44 (Heisenberg Uncertainty Principle)**.** *Suppose that $\psi \in \mathcal{S}(\mathbb{R})$ and that ψ is normalized in $L^2(\mathbb{R})$: $\|\psi\|_2 = 1$. Then*

$$\left(\int_{\mathbb{R}} x^2|\psi(x)|^2\,dx\right)\left(\int_{\mathbb{R}} \xi^2|\widehat{\psi}(\xi)|^2\,d\xi\right) \geq 1/(16\pi^2).$$

Proof. By hypothesis, $1 = \int_{\mathbb{R}} |\psi(x)|^2\,dx$. We integrate by parts, setting $u = |\psi(x)|^2$ and $dv = dx$ and taking advantage of the fast decay of ψ to get rid of the boundary terms. Note that

$$(d/dx)|\psi|^2 = (d/dx)\big(\psi\overline{\psi}\big) = \overline{\psi}\psi' + \overline{\overline{\psi}\psi'} = 2\mathrm{Re}(\overline{\psi}\psi').$$

[3] Named after the German theoretical physicist Werner Heisenberg (1901–1976).

Also $\operatorname{Re} z \leq |z|$, and $|\int f| \leq \int |f|$. Hence

$$\begin{aligned} 1 &= 2\int_{\mathbb{R}} x\operatorname{Re}\big[\overline{\psi(x)}\psi'(x)\big]\,dx \leq 2\int_{\mathbb{R}} |x||\psi(x)||\psi'(x)|\,dx \\ &\leq 2\Big(\int_{\mathbb{R}} |x\psi(x)|^2\,dx\Big)^{1/2}\Big(\int_{\mathbb{R}} |\psi'(x)|^2\,dx\Big)^{1/2} \\ &\leq 2\Big(\int_{\mathbb{R}} |x\psi(x)|^2\,dx\Big)^{1/2}\Big(4\pi^2\int_{\mathbb{R}} \xi^2|\widehat{\psi}(\xi)|^2\,d\xi\Big)^{1/2}. \end{aligned}$$

We have used the Cauchy–Schwarz Inequality, Plancherel's Identity, and property (g) of the time–frequency dictionary. □

The Uncertainty Principle holds for the class of continuous functions of moderate decrease and for the even larger class $L^2(\mathbb{R})$ of square-integrable functions.

Exercise 8.45. Check that equality holds in Heisenberg's Uncertainty Principle if and only if $\psi(x) = \sqrt{2B/\pi}\, e^{-Bx^2}$ for some $B > 0$. **Hint:** Equality holds in the Cauchy–Schwarz Inequality if and only if one of the functions is a multiple of the other. ♢

Exercise 8.46. Check that under the conditions of Theorem 8.44, $\big(\int_{\mathbb{R}} (x - x_0)^2|\psi(x)|^2\,dx\big)\Big(\int_{\mathbb{R}} (\xi - \xi_0)^2|\widehat{\psi}(\xi)|^2\,d\xi\Big) \geq 1/(16\pi^2)$ for all $x_0, \xi_0 \in \mathbb{R}$. **Hint:** Use the time–frequency dictionary for $\psi_{x_0,\xi_0}(x) = e^{2\pi i x\xi_0}\psi(x + x_0) \in \mathcal{S}(\mathbb{R})$ and Heisenberg's Uncertainty Principle. ♢

Let us use the Heisenberg Uncertainty Principle to show that *if* a function $f \in \mathcal{S}(\mathbb{R})$ is supported on the interval $[-a, a]$, *if* its Fourier transform is supported on $[-b, b]$, and *if* $\|f\|_2 = \|\widehat{f}\|_2 = 1$, then $ab \geq 1/(4\pi)$. It follows that if the support of f is small, then the support of $\widehat{f}$ has to be large, and vice versa. First,

$$\int_{\mathbb{R}} x^2|f(x)|^2\,dx \leq \int_{-a}^{a} x^2|f(x)|^2\,dx \leq a^2\|f\|_2^2 = a^2.$$

Likewise, $\int_{\mathbb{R}} \xi^2|\widehat{f(\xi)}|^2\,d\xi \leq b^2$. Applying the Heisenberg Uncertainty Principle and taking the square root, we conclude that $ab \geq 1/(4\pi)$, as claimed.

Those are big *if*'s. In fact, there is no function that is compactly supported both in time and in frequency; see Exercise 8.47. One can instead consider an approximate support: an interval on which most but not all of the L^2 norm is concentrated. See Remark 8.49.

Exercise 8.47. Suppose f is continuous on $\mathbb{R}$. Show that f and $\widehat{f}$ cannot both be compactly supported unless $f \equiv 0$. **Hint:** Suppose that f is supported on $[0, 1/2]$ and $\widehat{f}$ is compactly supported. Expand f in a Fourier series on $[0, 1]$, and observe that f must be a trigonometric polynomial. But trigonometric polynomials cannot vanish on an interval. ♢

The trigonometric functions are not in $L^2(\mathbb{R})$. However, viewing them as distributions, we can compute their Fourier transforms.

Exercise 8.48. Check that the Fourier transform of $e_\xi(x) = e^{2\pi i x\xi}$ in the sense of distributions is the shifted delta distribution $\tau_\xi\delta$ defined by $\tau_\xi\delta(\phi) = \phi(\xi)$. ♢

The preceding exercise also illustrates the time–frequency localization principle. In general, if the support of a function is essentially localized on an interval of length d, then its Fourier transform is essentially localized on an interval of length d^{-1}. Thus the support in the *time–frequency plane* is essentially a *rectangle* of area 1 and dimensions $d \times d^{-1}$.

Remark 8.49. The Uncertainty Principle has an interpretation in quantum mechanics, from which the name derives, as an *uncertainty about the position and the momentum associated to a free particle.* One cannot measure precisely both position and momentum. The better one measurement is, the worse the other will be. Mathematically, the state of a free particle is described by a function $\psi \in L^2(\mathbb{R})$ with $\|\psi\|_2 = 1$. The probability density that this particle is located at position x is given by $|\psi(x)|^2$. The probability density that its momentum is ξ is given by $|\widehat{\psi}(\xi)|^2$. The *average location* u and the *average momentum* ω of the particle are given, respectively, by

$$u = \int_{\mathbb{R}} x\,|\psi(x)|^2\,dx, \qquad \omega = \int_{\mathbb{R}} \xi\,|\widehat{\psi}(\xi)|^2\,d\xi.$$

The variance, or *spread*, σ_x^2 around the average location and the variance σ_ξ^2 around the average momentum are given by

$$\sigma_x^2 = \int_{\mathbb{R}} (x-u)^2 |\psi(x)|^2\, dx, \qquad \sigma_\xi^2 = \int_{\mathbb{R}} (\xi-\omega)^2 |\widehat{\psi}(\xi)|^2\, d\xi.$$

These numbers measure how far the particle's actual position or momentum can be from the average position and momentum. The larger σ_x is, the larger the uncertainty about its position is. The larger σ_ξ is, the larger the uncertainty about its momentum is. The Heisenberg Principle (or Exercise 8.46) says that $\sigma_x^2\sigma_\xi^2 \geq 1/(16\pi^2)$; therefore σ_x and σ_ξ cannot both be arbitrarily small.

A *Heisenberg box*, or *time–frequency box*, for a function ψ with $\|\psi\|_2 = 1$ is defined to be a rectangle centered at (u, ω), with dimensions $\sigma_x \times \sigma_\xi \geq 1/(4\pi)$. Although there are no functions that are simultaneously compactly supported in time and in frequency, there are functions with bounded Heisenberg boxes, which tell us about the time–frequency localization of the function on average. ♢

8.6. $L^p(\mathbb{R})$ as distributions

Functions in the *Lebesgue spaces* $L^p(\mathbb{R})$, for real numbers p with $1 \leq p < \infty$, induce tempered distributions via integration (equation (8.5)). Recall that the space $L^p(\mathbb{R})$ consists of those functions such that $\int_{\mathbb{R}} |f(x)|^p\, dx < \infty$; see Section 7.7 and the Appendix. Here the integral is in the sense of Lebesgue. Each $L^p(\mathbb{R})$ space is *normed*, with norm given by $\|f\|_p = \big(\int_{\mathbb{R}} |f(x)|^p\, dx\big)^{1/p}$. We summarize in Table 8.3 on page 215 the boundedness properties of the Fourier transform on each L^p space and on $\mathcal{S}(\mathbb{R})$ and $\mathcal{S}'(\mathbb{R})$.

The Schwartz class is *dense* in each $L^p(\mathbb{R})$ (see the Appendix). In other words, Schwartz functions are in $L^p(\mathbb{R})$ and 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: $\lim_{n\to\infty} \|f - \phi_n\|_p = 0$.

Functions $f \in L^p(\mathbb{R})$ define tempered distributions T_f by integration. The first thing to worry about is that the integral in equation (8.5) must be well-defined for all $f \in L^p(\mathbb{R})$ and $\phi \in \mathcal{S}(\mathbb{R})$. In fact the integral $\int_{\mathbb{R}} f(x)\phi(x)\, dx$ is well-defined (in the sense of Lebesgue) as long as $f \in L^p(\mathbb{R})$ and $\phi \in L^q(\mathbb{R})$ for p, q *dual or conjugate exponents*, that is, such that $1/p + 1/q = 1$ (in particular if $\psi \in \mathcal{S}(\mathbb{R}) \subset L^q(\mathbb{R})$ for all $1 \leq q \leq \infty$). This is a consequence of the fundamental *Hölder Inequality* (proved in Section 12.6). Hölder's Inequality on the line says that if $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$ with $1/p + 1/q = 1$, then $fg \in L^1(\mathbb{R})$. Furthermore, $\|fg\|_1 = \int_{\mathbb{R}} |f(x)g(x)|\,dx \leq \|f\|_p \|g\|_q$. It follows that $\int_{\mathbb{R}} f(x)g(x)\,dx$ is well-defined and finite.

The linearity of the integral guarantees the linearity of the functional T_f. Hölder's Inequality can be used to prove the continuity of the functional T_f, since the L^p norms of a function $\phi \in \mathcal{S}(\mathbb{R})$ can be controlled by the Schwartz seminorms; see Exercise 8.21.

Exercise 8.50. Show that if $f \in L^p(\mathbb{R})$, then the linear functional T_f defined by integration against f is continuous. ♢

Now that we have identified $f \in L^p(\mathbb{R})$ with its tempered distribution T_f, we can consider the Fourier transforms of L^p functions in the sense of distributions. Can we identify these Fourier transforms with reasonable functions? We have already looked at the special cases $p = 1$, 2, at least for $f \in \mathcal{S}(\mathbb{R})$.

Case $p = 1$: If $f \in \mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R})$, then $\|\widehat{f}\,\|_\infty \leq \|f\|_1$, since

$$|\widehat{f}\,(\xi)| = \left| \int_{\mathbb{R}} f(x) e^{-2\pi i x\xi}\,dx \right| \leq \int_{\mathbb{R}} |f(x)|\,dx.$$

Case $p = 2$: If $f \in \mathcal{S}(\mathbb{R}) \subset L^2(\mathbb{R})$, then Plancherel's Identity holds:

$$\|\widehat{f}\,\|_2 = \|f\|_2.$$

It turns out that the Fourier transform is a bijection on the full space $L^2(\mathbb{R})$. Also, using the Riemann–Lebesgue Lemma, the Fourier transform takes L^1 functions into (but not onto) a class of bounded functions, namely the subset $C_0(\mathbb{R})$ (continuous functions vanishing at infinity) of $L^\infty(\mathbb{R})$.

Lemma 8.51 (Riemann–Lebesgue Lemma)**.** *Suppose $f \in L^1(\mathbb{R})$. Then $\widehat{f}$ is continuous and $\lim_{|\xi|\to\infty} \widehat{f}(\xi) = 0$. That is, the Fourier transform maps $L^1(\mathbb{R})$ into $C_0(\mathbb{R})$.*

However, the Fourier transform is not a surjective map from $L^1(\mathbb{R})$ to $C_0(\mathbb{R})$; there is a function in $C_0(\mathbb{R})$ that is not the Fourier transform of any function in $L^1(\mathbb{R})$. See [**Kran**, Proposition 2.3.15].

Exercise 8.52. Prove the Riemann–Lebesgue Lemma in the special case when f is a function of moderate decrease. In particular, $\widehat{f}$ is then bounded. Show also that the inequality $\|\widehat{f}\|_\infty \leq \|f\|_1$ holds for continuous functions f of moderate decrease. **Hint:** The proof we used for $f \in \mathcal{S}(\mathbb{R})$ goes through for continuous functions of moderate decrease and also for functions in $L^1(\mathbb{R})$, using the theory of Lebesgue integration. ♢

Table 8.3. Effect of the Fourier transform on L^p spaces.

ϕ	$\to$	$\widehat{\phi}$
$\mathcal{S}(\mathbb{R})$	unitary bijection	$\mathcal{S}(\mathbb{R})$
$L^1(\mathbb{R})$	bounded map $\|\widehat{f}\|_\infty \leq \|f\|_1$ (Riemann–Lebesgue Lemma)	$C_0(\mathbb{R}) \subset L^\infty(\mathbb{R})$
$L^p(\mathbb{R})$ $1 < p < 2$	bounded map $\|\widehat{f}\|_q \leq C_p\|f\|_p$ (Hausdorff–Young Inequality)	$L^q(\mathbb{R})$ $1/p + 1/q = 1$
$L^2(\mathbb{R})$	isometry $\|\widehat{f}\|_2 = \|f\|_2$ (Plancherel's Identity)	$L^2(\mathbb{R})$
$\mathcal{S}'(\mathbb{R})$	bijection	$\mathcal{S}'(\mathbb{R})$

Warning: The class of continuous functions of moderate decrease is *not* closed under the Fourier transform.

Exercise 8.53 (*The Hat of the Hat*)**.** Let f be the *hat function*, defined by $f(x) = 1 - |x|$ for $|x| \leq 1$ and $f(x) = 0$ otherwise. Show that f is a continuous function of moderate decrease but that its Fourier transform $\widehat{f}$ is not. $\Diamond$

The Fourier transform is well-defined for $f \in L^1(\mathbb{R})$, but since $\widehat{f}$ may not be in $L^1(\mathbb{R})$, one has difficulties with the inversion formula. This is why we have chosen to present the theory on $\mathcal{S}(\mathbb{R})$. Since $\mathcal{S}(\mathbb{R})$ is dense in $L^2(\mathbb{R})$ (see the Appendix), one can extend the Fourier transform (and Plancherel's Identity) to $L^2(\mathbb{R})$, by continuity. Does this extension coincide with the Fourier transform defined via the integral formula for functions f in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$? Yes. It also coincides with the definition of the Fourier transform in the sense of distributions.

Notice that on $\mathbb{R}$, $L^1(\mathbb{R}) \not\subseteq L^2(\mathbb{R})$ and $L^1(\mathbb{R}) \not\supseteq L^2(\mathbb{R})$ (Exercise 8.54). Compare with the ladder of nested L^p spaces on the circle $\mathbb{T}$ (Figure 2.3).

For $1 < p < 2$, the Fourier transform is a bounded map from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$, for p, q dual exponents ($1/p + 1/q = 1$). This fact follows from the Hausdorff–Young Inequality, which we prove in Corollary 12.42. For $p > 2$, however, the Fourier transform does not map $L^p(\mathbb{R})$ into any nice functional space. If $f \in L^p(\mathbb{R})$ for $p > 2$, then its Fourier transform is a tempered distribution, but unless more information is given about f, we cannot say more. What is so special about $L^p(\mathbb{R})$ for $1 < p < 2$? It turns out that for p in this range we have $L^p(\mathbb{R}) \subset L^1(\mathbb{R}) + L^2(\mathbb{R})$, and one can then use the $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ theory. See [**Kat**, Sections VI.3–VI.4].

Exercise 8.54 ($L^p(\mathbb{R})$ *and* $L^q(\mathbb{R})$ *Are Not Nested*)**.** Suppose $1 \leq p < q \leq \infty$. Find a function in $L^q(\mathbb{R})$ that is not in $L^p(\mathbb{R})$. Find a function that is in $L^p(\mathbb{R})$ but not in $L^q(\mathbb{R})$. $\Diamond$

Exercise 8.55 ($L^p(\mathbb{R})$ *Is Contained in* $L^1(\mathbb{R}) + L^2(\mathbb{R})$)**.** Suppose $f \in L^p(\mathbb{R})$ for $1 < p < 2$. Let $f_\lambda(x) := f(x)$ if $|f(x)| \leq \lambda$ and $f_\lambda(x) := 0$ if $|f(x)| > \lambda$; define $f^\lambda(x) := f(x) - f_\lambda(x)$. Show that (i) $f_\lambda \in L^2(\mathbb{R})$, (ii) $f^\lambda \in L^1(\mathbb{R})$, and (iii) $L^p(\mathbb{R}) \subset L^1(\mathbb{R}) + L^2(\mathbb{R})$ for $1 < p < 2$. $\Diamond$

Even for f in $\mathcal{S}(\mathbb{R})$, if we hope for an inequality of the type

$$\|\widehat{f}\,\|_q \leq C\|f\|_p \qquad \text{for all } f \in \mathcal{S}(\mathbb{R}), \tag{8.12}$$

then p and q must be conjugate exponents: $1/p + 1/q = 1$ (see Example 8.56). Although it is not obvious from the example, such an inequality can only hold for $1 \leq p \leq 2$.

Example 8.56. Consider the following one-parameter family of functions in $\mathcal{S}(\mathbb{R})$, together with their Fourier transforms:

$$g_t(x) = e^{-\pi t x^2}, \qquad \widehat{g_t}(\xi) = e^{-\pi x^2/t}/\sqrt{t}.$$

Recall that $\int_{\mathbb{R}} e^{-\pi x^2}\,dx = 1$, and so $\int_{\mathbb{R}} e^{-\pi A x^2}\,dx = 1/\sqrt{A}$. It follows that the L^p norm of g_t and the L^q norm of $\widehat{g_t}$ are given by

$$\|g_t\|_p^p = 1/\sqrt{tp}, \qquad \|\widehat{g_t}\|_q^q = (\sqrt{t})^{1-q}/\sqrt{q}.$$

(Note that $\|g_t\|_p^p = \|\widehat{g_t}\|_q^q$ when $p = q = 2$, as Plancherel's Identity ensures.) If inequality (8.12) holds, then we must be able to bound the ratios $\|\widehat{g_t}\|_q/\|g_t\|_p$ independently of t. But the ratio

$$\|\widehat{g_t}\|_q/\|g_t\|_p = (\sqrt{t})^{\frac{1}{p}+\frac{1}{q}-1}(\sqrt{p})^{1/p}/(\sqrt{q})^{1/q}$$

is bounded independently of t if and only if the exponent of $\sqrt{t}$ is zero, in other words, when p and q are conjugate exponents. $\diamondsuit$

8.7. Project: Principal value distribution $1/x$

In this project we present an example of a distribution that is neither a function in $L^p(\mathbb{R})$ nor a point-mass (like the delta distribution) nor a measure. It is the *principal value distribution* $1/x$, the building block of the Hilbert transform, discussed in more detail in Chapter 12. Below are some exercises to guide you. Search for additional historical facts to include in your essay.

(a) (*The Principal Value Distribution* $1/x$) The linear functional $H_0 : \mathcal{S}(\mathbb{R}) \to \mathbb{C}$ is given by

$$H_0(\phi) = \text{p.v.} \int_{\mathbb{R}} \frac{\phi(y)}{y}\,dy := \lim_{\varepsilon \to 0} \int_{|y|>\varepsilon} \frac{\phi(y)}{y}\,dy. \tag{8.13}$$

Prove that the functional H_0 is continuous by showing that if $\phi_k \to 0$ in $\mathcal{S}(\mathbb{R})$ as $k \to \infty$, then the complex numbers $H_0(\phi_k) \to 0$ as $k \to \infty$, as follows. We need to take advantage of some hidden cancellation. Observe that for each $\varepsilon > 0$, $\int_{\varepsilon \leq |y| < 1} \frac{1}{y}\,dy = 0$. Now add zero to equation (8.13) to show that $\left|\int_{|y|>\varepsilon} \frac{\phi_k(y)}{y}\,dy\right| \leq 2\big(\rho_{1,0}(\phi_k) + \rho_{0,1}(\phi_k)\big)$, where $\rho_{1,0}$ and $\rho_{0,1}$ are the seminorms introduced in formula (8.4). Finally verify that the right-hand side can be made arbitrarily small for k large enough. You have shown that H_0 is a tempered distribution, namely the *principal value distribution* $H_0 = \text{p.v. } 1/x$. Remark 8.31 states that there must be a function f such that H_0 is the k^{th} derivative (in the sense of distributions) of f, for some k. Find f and k.

(b) For each $\varepsilon > 0$, the function $x^{-1}\chi_{\{|x|>\varepsilon\}}(x)$ is bounded and defines a tempered distribution, which we call H_0^ε. Check that the limit of H_0^ε in the sense of distributions as $\varepsilon \to 0$ is H_0, meaning that

$$\lim_{\varepsilon\to 0} H_0^\varepsilon(\phi) = H_0(\phi) \qquad \text{for each } \phi \in \mathcal{S}(\mathbb{R}).$$

(c) For each $x \in \mathbb{R}$ define a new tempered distribution by appropriately translating and reflecting H_0, as follows.

Definition 8.57 (*The Hilbert Transform as a Distribution*)**.** Given $x \in \mathbb{R}$, the *Hilbert transform* $H_x(\phi)$ of ϕ at x, also written $H\phi(x)$, is a tempered distribution acting on $\phi \in \mathcal{S}(\mathbb{R})$ and defined by

$$H_x(\phi) = H\phi(x) := (\tau_{-x}H_0)^{\sim}(\phi)/\pi. \qquad \Diamond$$

Verify that $H\phi(x) = \frac{1}{\pi}\lim_{\varepsilon\to 0}\int_{|x-t|>\varepsilon}\frac{\phi(t)}{x-t}\,dt$.

8.8. Project: Uncertainty and primes

In this project we ask the reader to research and present a proof of a discrete uncertainty principle and to understand its applications to the distribution of prime numbers (to get started, see [**SS03**, Chapter 8]).

(a) Using the definitions given in the project in Section 6.9, define the support of a function $f \in L^2(G)$ to be $\text{supp}(f) := \{x \in G : f(x) \neq 0\}$. Verify the following inequality, which encodes a *discrete uncertainty principle*:

$$\#\,\text{supp}(f) \times \#\,\text{supp}(\widehat{f}) \geq \#G = n.$$

(b) Recently T. Tao proved a refinement of this result: Let G be a cyclic group of prime order p and take $f \in L^2(G)$. Then

$$\#\text{supp}(f) + \#\text{supp}(\widehat{f}) \geq p + 1,$$

and this bound is sharp. As an application, Green and Tao prove that there are arbitrarily long arithmetic progressions of primes. An *arithmetic progression of primes* is a set of the form $\{p, p+d, p+2d, p+3d, \ldots, p+nd\}$, where p and d are fixed and each element is a prime number. At the time of writing, the longest known arithmetic progression of primes has only twenty-six primes[4]. This fact makes their theorem even more remarkable. The Tao–Green Theorem is a very difficult result. We list references of an expository nature, by the authors and others: [**Gre**], [**Kra**], [**Tao05**], [**Tao06c**].

[4] You can find the longest known arithmetic progressions listed at `http://en.wikipedia.org/wiki/Primes\_in\_arithmetic\_progression`.

Chapter 9

From Fourier to wavelets, emphasizing Haar

There are many things we would like to record, compress, transmit, or reconstruct: for instance, audio signals such as music or speech, video signals such as TV or still photos, and data in the form of words or numbers. Harmonic analysis provides the theoretical tools to do all these tasks effectively. Nowadays, in the toolbox of every engineer, statistician, applied mathematician, or other practitioner interested in signal or image processing, you will find not only classical Fourier analysis tools but also modern wavelet analysis tools.

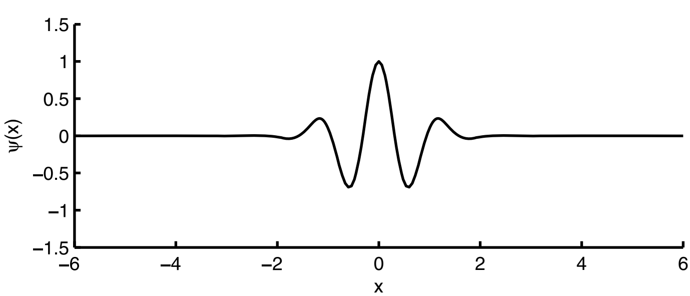


Figure 9.1. A Morlet wavelet given by $\psi(x) = e^{-t^2/2}\cos(5t)$.

Now, what is a wavelet? It is a *little wave that starts and stops*, such as the Morlet[1] wavelet shown in Figure 9.1. It differs from sines and cosines, which go on forever, and from truncated sines and cosines, which go on for the whole length of the truncation window.

The key idea of wavelets is to express functions or signals as sums of these little waves and of their translations and dilations. Wavelets play the rôle here that sines and cosines do in Fourier analysis. They can be more efficient, especially if the signal lasts only a finite time or behaves differently in different time periods. A third type of analysis, the windowed Fourier transform, unites some features of Fourier analysis with the localization properties of wavelets. Each of these three types of analysis has its pros and cons.

In this chapter we survey the windowed Fourier transform and its generalization known as the Gabor transform (Section 9.2) and introduce the newest member of the family, the wavelet transform (Section 9.3). We develop the first and most important example of a wavelet basis, the Haar basis (Section 9.4). We show that the Haar system is a complete orthonormal system in $L^2(\mathbb{R})$. We question the plausibility of this statement, and then we prove the statement. To do so, we introduce dyadic intervals, expectation (averaging) operators, and difference operators. We compare Fourier and Haar analysis and touch on some dyadic operators, whose boundedness properties hold the key to the unconditionality of the Haar basis in $L^p(\mathbb{R})$ (Section 9.5).

9.1. Strang's symphony analogy

Let us begin with the central idea of wavelet theory, through a musical analogy developed by Gilbert Strang; the quotations in this section are from his article [**Stra94**].

Imagine we are listening to an orchestra playing a symphony. It has a rich, complex, changing sound that involves dynamics, tempi, pitches, and so on. How do we write it down? How can we notate this sound experience? Musical notation is one way. Recording on

[1]Named after French geophysicist Jean Morlet (1931–2007), who coined the term *wavelet* to describe the functions he used. In 1981, Morlet and Croatian-French physicist Alex Grossman (born 1930) developed what we now call the wavelet transform.

CD is another. These two methods do not contain exactly the same information. We are expressing the symphony in different ways.

We consider three further ways of analyzing a symphony: in terms of (1) cosine waves (traditional Fourier series); (2) pieces of cosine waves (windowed Fourier series); and (3) wavelets.

(1) *Traditional Fourier analysis* "separate[s] the whole symphony into pure harmonics. Each instrument plays one steady note", with a specific loudness; for example the B flat above middle C at a given volume, or the F sharp below middle C at some other volume. The signal, meaning the whole symphony, becomes a sum $a_0 + a_1 \cos t + a_2 \cos 2t + \cdots$ of cosine waves. To store the signal, we need to record only the list of *amplitudes* $(a_0, a_1, a_2, \dots)$ and corresponding *frequencies* $(0, 1, 2, \dots)$.

In practice we cannot retain all these terms, but we may need to keep many of them to get a high-fidelity reproduction. Also, if the symphony has been playing forever and continues forever, then we need a continuum of frequencies (uncountably many musicians), and instead of a sum we get an integral, namely the Fourier transform.

(2) *Windowed Fourier analysis*, also known as the short-time Fourier transform, looks at short segments of the symphony individually. "In each segment, the signal is separated into cosine waves as before. The musicians still play one note each, but they change amplitude in each segment. This is the way most long signals are carved up." Now we need to record lists of amplitudes and frequencies *for each segment.*

"One disadvantage: There are sudden breaks between segments." In an audio signal, we might be able to hear these breaks, while in a visual image, we might see an edge. This kind of artificial discontinuity is called a *blocking artifact.* An example is shown in the JPEG fingerprint image in Figure 10.3.

(3) In *wavelet analysis*, "[i]nstead of cosine waves that go on forever or get chopped off, the new building blocks are '*wavelets*'. These are little waves that start and stop", like the Morlet wavelet in Figure 9.1. In the symphony analogy "[t]hey all come from one basic wavelet $w(t)$,

which gives the sound level of the standard tune at time t.” In practice, the Daubechies wavelet shown in Figure 9.4 is a typical standard tune.

We make up the symphony from different versions of this tune, played with specific amplitudes and starting times and at specific speeds. The *double basses* play the standard tune $w(t)$. “The *cellos* play the same tune but in half the time, at doubled frequency. Mathematically, this speed-up replaces the time variable t by $2t$. The first bass plays $b_1w(t)$ and the first cello plays $c_1w(2t)$, both starting at time $t = 0$. The next bass and cello play $b_2w(t-1)$ and $c_2w(2t-1)$, with amplitudes b_2 and c_2. The bass starts when $t - 1 = 0$, at time $t = 1$. The cello starts earlier, when $2t - 1 = 0$ and $t = 1/2$. There are twice as many cellos as basses, to fill the length of the symphony. Violas and violins play the same passage but faster and faster and all overlapping. At every level the frequency is doubled (up one octave) and there is new richness of detail. ‘Hyperviolins’ are playing at 16 and 32 and 64 times the bass frequency, probably with very small amplitudes. Somehow these wavelets add up to a complete symphony.”

It is as if each possible signal (symphony) could be achieved by an orchestra where each musician plays only the song “Happy Birthday” but at his or her own speed and start time and with his or her own volume chosen according to the signal required. To store the signal, we need to record only the list of amplitudes, in order. At every level the frequency is doubled, which means that the pitch goes up by an octave. The rôle played by the cosines and sines in Fourier analysis is played here by *dilates and translates of a single function* $w(t)$. Here we see the fundamental idea of wavelets.

For a delightful introduction to these ideas for a broad audience, see Barbara Burke Hubbard’s book [**Burke**].

9.2. The windowed Fourier and Gabor bases

We construct an orthonormal basis on the line by pasting together copies of the trigonometric basis on intervals (*windows*) that partition the line. Generalizing this idea, we obtain the Gabor bases.

9.2.1. The windowed Fourier transform. The continuous Fourier transform gives us a tool for analyzing functions that are defined on the whole real line $\mathbb{R}$, rather than on the circle $\mathbb{T}$. However, the trigonometric functions $\{e^{2\pi i\xi x}\}_{\xi\in\mathbb{R}}$ no longer form a countable basis, since there is one for each real number ξ. Also, for each fixed $\xi \in \mathbb{R}$, the function $e^{2\pi i\xi x}$ is not in $L^2(\mathbb{R})$. The windowed Fourier transform addresses the problem of finding a basis for $L^2(\mathbb{R})$ to fill the rôle played by the trigonometric basis for $L^2(\mathbb{T})$.

How can we obtain an *orthonormal basis* for $L^2(\mathbb{R})$? A simple solution is to split the line into unit segments $[k, k+1)$ indexed by $k \in \mathbb{Z}$ and on each segment use the periodic Fourier basis for that segment.

Theorem 9.1. *The functions*

$$g_{n,k}(x) = e^{2\pi inx}\chi_{[k,k+1)}(x) \qquad \text{for } n, k \in \mathbb{Z}$$

form an orthonormal basis for $L^2(\mathbb{R})$, where $\chi_{[k,k+1)}$ is the characteristic function of the interval $[k, k+1)$. The corresponding reconstruction formula *holds, with equality in the L^2 sense:*

$$f(x) = \sum_{n,k\in\mathbb{Z}} \langle f, g_{n,k}\rangle g_{n,k}(x).$$

Exercise 9.2. Prove Theorem 9.1. **Hint:** Use Lemma A.51. ♢

Definition 9.3. The *windowed Fourier transform* is the map G that assigns to each function in $L^2(\mathbb{R})$ the sequence of coefficients with respect to the windowed Fourier basis $\{g_{n,k}\}_{j,k\in\mathbb{Z}}$ defined in Theorem 9.1. More precisely, $G : L^2(\mathbb{R}) \to \ell^2(\mathbb{Z}^2)$ is defined by

$$Gf(n, k) := \langle f, g_{n,k}\rangle = \int_k^{k+1} f(x)e^{-2\pi inx}\, dx.$$

Here our signal f is a function of the continuous variable $x \in \mathbb{R}$, while its windowed Fourier transform Gf is a function of two discrete (integer) variables, n and k. ♢

We can think of each characteristic function $\chi_{[k,k+1)}(x)$ as giving us a window through which to view the behavior of f on the interval $[k, k+1)$. Both the function $\chi_{[k,k+1)}(x)$ and the associated interval $[k, k+1)$ are commonly called *windows*. These windows all have unit length and they are integer translates of the fixed window $[0,1)$, so that $\chi_{[k,k+1)}(x) = \chi_{[0,1)}(x-k)$.

Instead, we could use windows of varying sizes. Given an arbitrary partition $\{a_k\}_{k\in\mathbb{Z}}$ of $\mathbb{R}$ into bounded intervals $[a_k, a_{k+1})$, $k \in \mathbb{Z}$, let $L_k = a_{k+1} - a_k$, and on each window $[a_k, a_{k+1})$ use the corresponding L_k Fourier basis. Then the functions

$$(1/\sqrt{L_k})e^{-2\pi inx/L_k}\chi_{[a_k,a_{k+1})}(x) \qquad \text{for } n, k \in \mathbb{Z}$$

form an orthonormal basis of $L^2(\mathbb{R})$. This generalization lets us adapt the basis to the function to be analyzed. For instance, if the behavior of f changes a lot in some region of $\mathbb{R}$, we may want to use many small windows there, while wherever the function does not fluctuate much, we may use wider windows.

We hope to get a fairly accurate reconstruction of f while retaining only a few coefficients. On the wider windows, a few low frequencies should contain most of the information. On the smaller windows, retaining a few big coefficients may not be very accurate, but that may not matter much if the windows are small and few in number. Alternatively one may have many small windows, on each of which few coefficients are retained.

One seeks a balance between the chosen partition of $\mathbb{R}$ and the number of significant coefficients to keep per window. This preprocessing may require extra information about the function and/or extra computations which may or may not be affordable for a specific application. On the other hand, by adapting the windows to the function, we lose the translation structure that arises from having all the windows the same size.

Exercise 9.4 (*The Gibbs Phenomenon for the Windowed Fourier Transform*)**.** Another problem arises from the discontinuity of the windows at the endpoints: the Gibbs phenomenon (see also the project in Section 3.4). The *hat function* is defined to be $f(x) := 1 - |x|$ if $-1 \le x < 1$ and $f(x) := 0$ otherwise. Compute the windowed Fourier transform of the hat function using windows on the intervals $[k, k+1)$, for $k \in \mathbb{Z}$, and plot using MATLAB. You should see corners at $x = -1$, 0, 1. Do the same with windows $[2k-1, 2k+1)$. Do you see any corners? How about with windows $[k/2, (k+1)/2)$? ♢

9.2.2. Gabor bases. In numerical calculations, the *sharp windows* $\chi_{[k,k+1)}(x)$ used in the windowed Fourier transform produce at the edges the same *artifacts* that are seen when analyzing periodic functions at discontinuity points (the Gibbs phenomenon, or *corners* at the divisions between windows). So smoother windows are desirable. The sharp windows given by $\chi_{[0,1)}(x)$ and its modulated integer translates $e^{2\pi inx}\chi_{[0,1)}(x-k)$ can be replaced by a *smooth window* g and its modulated integer translates.

Definition 9.5. A function $g \in L^2(\mathbb{R})$ is a *Gabor*[2] *function* if the family of its modulated integer translates

$$g_{n,k}(x) = g(x-k)e^{2\pi inx} \qquad \text{for } n, k \in \mathbb{Z} \tag{9.1}$$

is an orthonormal basis for $L^2(\mathbb{R})$. Such a basis is a *Gabor basis.* ♢

Example 9.6. The sharp window $g(x) = \chi_{[0,1)}(x)$ is a Gabor function. Figure 9.2 shows the imaginary part of $g_{n,k}(x)$ with the sharp window $g(x)$ and the parameters $n = 6$, $k = 0$. ♢

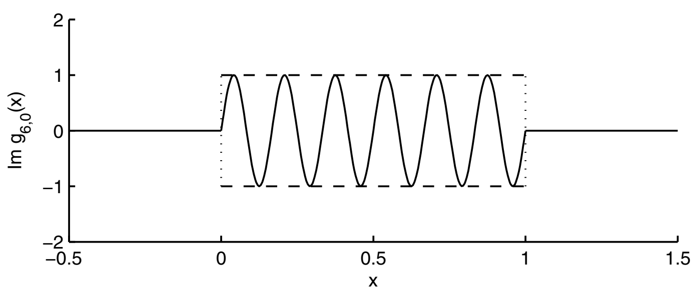


Figure 9.2. Graph (solid line) of the imaginary part $\operatorname{Im} g_{6,0}(x) = \sin(12\pi x)\chi_{[0,1)}(x)$ of the Gabor function $g_{6,0}(x) = \exp\{2\pi i6x\}\chi_{[0,1)}(x)$, for the sharp window $g = \chi_{[0,1)}$. Dashed lines show the envelope formed by g and $-g$.

In 1946, Gabor proposed using systems of this kind in communication theory [**Gab**]. More specifically, he proposed using integer

[2]Named after Dennis Gabor (1900–1979), a Hungarian electrical engineer and inventor. He is most notable for inventing holography, for which he received the 1971 Nobel Prize in physics.

translates and modulations of the Gaussian function as a "basis" for $L^2(\mathbb{R})$, though unfortunately the Gaussian is not a Gabor function.

Lemma 9.7. *A family $\{\psi_n\}_{n\in\mathbb{N}}$ of functions is an orthonormal basis for $L^2(\mathbb{R})$ if and only if its Fourier transforms $\{\widehat{\psi_n}\}_{n\in\mathbb{N}}$ form an orthonormal basis for $L^2(\mathbb{R})$.*

Proof. Orthonormality holds on one side if and only if it holds on the other, because by the polarization formula (equation (7.24)), $\langle\psi_n, \psi_m\rangle = \langle\widehat{\psi_n}, \widehat{\psi_m}\rangle$ for all n, $m \in \mathbb{N}$. The same is true for completeness, because $f \perp \psi_n$ if and only if $\widehat{f} \perp \widehat{\psi_n}$. □

In particular, given a Gabor basis $\{g_{n,k}\}_{n,k\in\mathbb{Z}}$ as in equation (9.1), the Fourier transforms $\{\widehat{g_{n,k}}\}_{n,k\in\mathbb{Z}}$ form an *orthonormal* basis. Remarkably, they also form a *Gabor* basis. We see on closer examination that the form of the modulated integer translates in equation (9.1) is exactly what is needed here, since the Fourier transform converts translation to modulation, and vice versa.

Exercise 9.8. Use the time–frequency dictionary to show that the Fourier transforms of the Gabor basis elements are

$$\widehat{g_{n,k}}(\xi) = \widehat{g}\,(\xi - n)e^{-2\pi ik\xi} = (\widehat{g}\,)_{-k,n}(\xi). \tag{9.2}$$ ◊

To sum up, we have proved the following lemma.

Lemma 9.9. *A function $g \in L^2(\mathbb{R})$ generates a Gabor basis, meaning that $\{g_{n,k}\}_{n,k\in\mathbb{Z}}$ forms an orthonormal basis in $L^2(\mathbb{R})$, if and only if $\widehat{g} \in L^2(\mathbb{R})$ generates a Gabor basis, meaning that $\{(\widehat{g})_{n,k}\}_{n,k\in\mathbb{Z}}$ forms an orthonormal basis in $L^2(\mathbb{R})$.*

Example 9.10. Since $g = \chi_{[0,1)}$ generates a Gabor basis, so does its Fourier transform

$$\widehat{g}\,(\xi) = (\chi_{[0,1)})^\wedge(\xi) = e^{-i\pi\xi}\left(\sin(\pi\xi)/\pi\xi\right) = e^{-i\pi\xi}\,\mathrm{sinc}(\xi).$$

This window $\widehat{g}$ is differentiable, in contrast to g which is not even continuous. However $\widehat{g}$ is not compactly supported, unlike g. ◊

Can we find a Gabor function that is simultaneously smooth and compactly supported? The answer is NO. The limitations of the Gabor analysis are explained by the following result.

Theorem 9.11 (Balian–Low[3] Theorem)**.** *If $g \in L^2(\mathbb{R})$ is a Gabor function, then either*

$$\int_{\mathbb{R}} x^2 |g(x)|^2 \, dx = \infty \quad \text{or} \quad \int_{\mathbb{R}} \xi^2 |\widehat{g}\,(\xi)|^2 \, d\xi = \int_{\mathbb{R}} |g'(x)|^2 \, dx = \infty.$$

Exercise 9.12 (*Examples and Nonexamples of Balian–Low*)**.** Verify the Balian–Low Theorem for the two examples discussed so far: $g(x) = e^{2\pi inx}\chi_{[0,1)}(x)$ and $g(x) = e^{-i\pi x} \operatorname{sinc} x$. Show also that the Gaussian $G(x) = e^{-\pi x^2}$ is not a Gabor function. ♢

A proof of Theorem 9.11 can be found in [**Dau92**, p. 108, Theorem 4.1.1]. The theorem implies that a Gabor window, or *bell*, cannot be simultaneously smooth and compactly supported. Our first example, $g(x) = \chi_{[0,1)}(x)$, is not even continuous but is perfectly localized in time, while our second example, $g(x) = e^{-i\pi x} \operatorname{sinc}(x)$, is the opposite. In particular the slow decay of the sinc function reflects the lack of smoothness of the characteristic function $\chi_{[0,1)}(x)$. This phenomenon is an incarnation of *Heisenberg's Uncertainty Principle*.

Exercise 9.13. Show that the Balian–Low Theorem implies that a Gabor function cannot be both smooth and compactly supported. ♢

However, if the exponentials are replaced by appropriate cosines and sines, one can obtain Gabor-type bases with smooth bell functions. These are the so-called *local cosine and sine bases*, first discovered by Malvar [**Malv**] and later described by Coifman and Meyer [**CMe**]. See the discussion in [**HW**] and the project in Section 9.6.

Another way to get around the Balian–Low Theorem is to accept some redundancy and use translations and modulations with less-than-integer increments, obtaining *frames* in place of orthogonal bases. Gabor would have been pleased.

There is a *continuous Gabor transform* as well, where the parameters are now real numbers instead of integers. Let g be a real and symmetric window with $\|g\|_2 = 1$. The Gabor transform is given by

$$Gf(\xi, u) = \int_{\mathbb{R}} f(x) g(x-u) e^{-2\pi i \xi x} \, dx = \langle f, g_{\xi,u} \rangle,$$

[3] Named after French physicist Roger Balian (born 1933) and American theoretical physicist Francis E. Low (1921–2007).

where $g_{\xi,u}(x) := g(x-u)e^{2\pi i\xi x}$ for u, $\xi \in \mathbb{R}$. The multiplication by the translated window localizes the Fourier integral in a neighborhood of u. The following *inversion formula* holds for $f \in L^2(\mathbb{R})$:

$$f(x) \quad = \quad \int_{\mathbb{R}^2} \langle f, g_{\xi,u}\rangle g_{\xi,u}(x)\, d\xi\, du. \tag{9.3}$$

These formulas are similar in spirit to the Fourier transform and the inverse Fourier transform integral formulas on $\mathbb{R}$.

Exercise 9.14. Verify the inversion formula (9.3) for f, $g \in \mathcal{S}(\mathbb{R})$, where g is a real-valued even function with $\|g\|_2 = 1$. **Hint:** Using the time–frequency dictionary, show that (9.3) is equivalent to $f(x) = \int \big(g_{\xi,0} * g_{\xi,0} * f\big)(x)\, d\xi$ and that the Fourier transform of the right-hand side is $\widehat{f}$. As usual, $*$ denotes convolution on $\mathbb{R}$. ♢

Gabor bases give partial answers to questions about localization. One problem is that the sizes of the windows are fixed. Variable widths applicable to different functions, while staying within a single basis, are the new ingredient added by wavelet analysis.

9.3. The wavelet transform

The wavelet transform involves translations (as in the Gabor basis) and scalings (instead of modulations). These translates and dilates introduce a natural zooming mechanism. The idea is to express a discrete signal in terms of its *average* value a_1 and successive levels of *detail*, d_1, d_2, ... Similarly, in archery, one first sees the entire *target* and then resolves the details of the *bull's-eye* painted on it. When using binoculars, one first locates the object (finds its average position) and then adjusts the focus until the fine details of the object jump into view. The zooming mechanism is mathematically encoded in the *multiresolution* structure of these bases; see Chapter 10.

Definition 9.15. A function $\psi \in L^2(\mathbb{R})$ is a *wavelet* if the family

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \quad \text{for } \ j,k \in \mathbb{Z} \tag{9.4}$$

forms an orthonormal basis of $L^2(\mathbb{R})$. If so, the basis is called a *wavelet basis*. ♢

The family of Fourier transforms of a wavelet basis is another orthonormal basis, but it is not a wavelet basis. It is generated from one function $\widehat{\psi}$ by scalings and modulations, rather than by scalings and translations.

Exercise 9.16 (*Fourier Transform of a Wavelet*)**.** Use the time–frequency dictionary (Table 7.1) to find $\widehat{\psi_{j,k}}(\xi)$, for $\psi \in L^2(\mathbb{R})$. ♢

The following reconstruction result follows immediately from the completeness of an orthonormal basis. The hard part is to identify functions ψ that are wavelets, so that the hypothesis of Proposition 9.17 holds. We tackle that issue in Chapter 10.

Proposition 9.17. *If $\psi \in L^2(\mathbb{R})$ is a wavelet, then the following* reconstruction formula *holds in the L^2 sense:*

$$f(x) = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \, \psi_{j,k}(x) \qquad \textit{for all } f \in L^2(\mathbb{R}).$$

Definition 9.18. The *orthogonal wavelet transform* is the map $W : L^2(\mathbb{R}) \to \ell^2(\mathbb{Z}^2)$ that assigns to each function in $L^2(\mathbb{R})$ the sequence of its wavelet coefficients:

$$Wf(j,k) := \langle f, \psi_{j,k}\rangle = \int_{\mathbb{R}} f(x)\overline{\psi_{j,k}(x)}\, dx. \qquad ♢$$

The earliest known wavelet basis is the *Haar basis* on $L^2([0,1))$, introduced by Alfréd Haar in 1910 [**Haa**]. For the Haar basis, unlike the trigonometric basis, the partial sums for continuous functions converge uniformly.

Example 9.19 (*The Haar Wavelet*)**.** The *Haar wavelet* $h(x)$ on the unit interval is given by

$$h(x) := -\chi_{[0,1/2)}(x) + \chi_{[1/2,1)}(x).$$

See Figure 9.3. The family $\{h_{j,k}(x) := 2^{j/2}h(2^jx - k)\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$, as we will see in Section 9.4. ♢

Exercise 9.20. Show that $\{h_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal set; that is, verify that $\langle h_{j,k}, h_{j',k'}\rangle = 1$ if $j = j'$ and $k = k'$, and $\langle h_{j,k}, h_{j',k'}\rangle = 0$ otherwise. First show that the functions $h_{j,k}$ have zero integral: $\int h_{j,k} = 0$. ♢

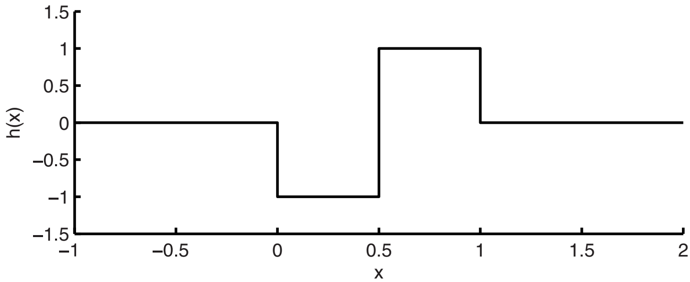


Figure 9.3. The Haar wavelet $h(x)$.

The reader may wonder how an arbitrary function in $L^2(\mathbb{R})$, which need not have zero integral, can be written in terms of the Haar functions, which all have zero integral. To resolve this puzzle, see Section 9.4.3 and the project in Section 9.7.

An important part of wavelet theory is the search for smoother wavelets. The Haar wavelet is discontinuous. It is also perfectly localized in time and therefore not perfectly localized in frequency. At the other end of the spectrum, one finds the *Shannon wavelet*[4] which is localized in frequency but not in time.

Example 9.21 (*The Shannon Wavelet*)**.** Let ψ be given on the Fourier side by

$$\widehat{\psi}(\xi) := e^{2\pi i\xi}\chi_{[-1,-1/2)\cup[1/2,1)}(\xi).$$

The family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$. The function ψ is the *Shannon wavelet*, and the corresponding basis is the *Shannon basis.* ♢

Exercise 9.22. Show that the Shannon functions $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ form an orthonormal set and each has zero integral. Furthermore, they are a basis. **Hint:** Work on the Fourier side using the polarization formula (7.24). On the Fourier side we are dealing with a win-

[4]Named after the same Shannon as the sampling theorem in Section 8.5.2.

dowed Fourier basis, with *double-paned windows* $F_j := [-2^j, -2^{j-1}) \cup [2^{j-1}, 2^j)$ of combined length 2^j, for $j \in \mathbb{Z}$, that are congruent[5] modulo 2^j to the interval $[0, 2^j)$. The collection of double-paned windows F_j forms a partition of $\mathbb{R} \setminus \{0\}$. On each double-paned window, the trigonometric functions $2^{-j/2} e^{2\pi i k x 2^{-j}} \chi_{F_j}(x)$, for $k \in \mathbb{Z}$ and a fixed j, form an orthonormal basis of $L^2(F_j)$. Now use Lemma A.51. ◊

The Shannon wavelet is perfectly localized in frequency, therefore not in time. Compact support on the frequency side translates into smoothness (C^∞) of the Shannon wavelet on the time side. Thus the Shannon wavelet is an example of a C^∞ wavelet without compact support. Can one find compactly supported wavelets that are smooth? YES. In a fundamental paper on wavelet theory [**Dau88**], I. Daubechies constructed compactly supported wavelets with arbitrary (but finite) smoothness; they are in C^k. However, it is impossible to construct wavelets that are both compactly supported and C^∞.

We can associate to most wavelets a sequence of numbers, known as a *filter*, and a companion *scaling function*.

A *finite impulse response (FIR) filter* has only finitely many nonzero entries. If a wavelet has a finite impulse response filter, then the wavelet and its scaling function are compactly supported. The converse is false; see the project in Subsection 10.5.2.

One of Daubechies' key contributions was her discovery of the family of wavelets later named after her, each of which has finitely many nonzero coefficients h_k and wavelet and scaling functions ψ and ϕ (Figure 9.4) having preselected regularity (meaning smoothness, or degree of differentiability) and compact support. As noted in [**Mey**, Chapter 1], eighty years separated Haar's work and its natural extension by Daubechies. IT'S AMAZING! Something as irregular as the spiky function in Figure 9.4 is a wavelet! No wonder it took so long to find these continuous, compactly supported wavelets.

---

[5]A subset A of $\mathbb{R}$ is *congruent* to an interval I if for each point $x \in A$ there exists a unique integer k such that $x + k|I|$ is in I and for each point $y \in I$ there is a unique $x \in A$ and $k \in \mathbb{Z}$ so that $y = x + k|I|$. Here is a nonexample: The set $[-1/2, 0) \cup [1/2, 1)$ is not congruent modulo 1 to the interval $[0, 1)$.

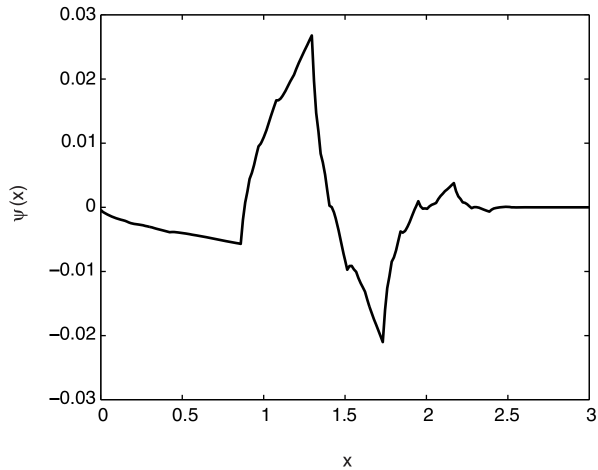


Figure 9.4. The Daubechies wavelet function ψ, for the $db2$ wavelet. The shape is reminiscent of the profile of the Sydney Opera House.

The more derivatives a wavelet has, the longer the filter and the longer the support of the wavelet. The shorter the filter, the better for implementation, so there is a trade-off. Filter bank theory and the practicability of implementing FIR filters opened the door to widespread use of wavelets in applications. See Section 10.3.4 and Chapter 11.

One can develop the theory of wavelets in $\mathbb{R}^N$ or in $\mathbb{C}^N$ via linear algebra, in the same way that we built a finite Fourier theory and introduced the discrete Haar basis in Chapter 6. See [**Fra**]. There is a *Fast Haar Transform* (FHT) (Section 6.7), and it generalizes to a *Fast Wavelet Transform* (FWT) (Chapter 11), which has been instrumental in the success of wavelets in practice.

As in the Fourier and Gabor cases, there is also a *continuous wavelet transform*. It uses continuous translation and scaling parameters, $u \in \mathbb{R}$ and $s > 0$, and a family of *time–frequency atoms* that is obtained from a normalized wavelet $\psi \in L^2(\mathbb{R})$ with zero average

($\int \psi = 0$) by shifting by u and rescaling by s. Let

$$\psi_{s,u}(x) = \frac{1}{\sqrt{s}}\,\psi\left(\frac{x-u}{s}\right), \quad \text{so} \quad (\psi_{s,u})^\wedge(\xi) = \sqrt{s}\,e^{-2\pi i u\xi}\,\widehat{\psi}\,(s\xi).$$

The *continuous wavelet transform* is then defined by

$$Wf(s,u) = \langle f, \psi_{s,u}\rangle = \int_{\mathbb{R}} f(x)\,\overline{\psi_{s,u}(x)}\,dx.$$

If ψ is real-valued and localized near 0 with spread 1, then $\psi_{s,u}$ is localized near u with spread s. The wavelet transform measures the variation of f near u at scale s (in the discrete case, $u = k2^{-j}$ and $s = 2^{-j}$). As the scale s goes to zero (j goes to infinity), the decay of the wavelet coefficients characterizes the regularity of f near u. Also, if $\widehat{\psi}$ is localized near 0 with spread 1, then $\widehat{\psi_{s,u}}$ is localized near 0 with spread $1/s$. That is, the Heisenberg boxes of the wavelets are rectangles of area 1 and dimensions $s \times 1/s$. (See Section 8.5.3.)

Under very mild assumptions on the wavelet ψ, we obtain a reconstruction formula. Each $f \in L^2(\mathbb{R})$ can be written as

$$f(x) = \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^{+\infty} Wf(s,u)\psi_{s,u}(x)\,\frac{du\,ds}{s^2}, \tag{9.5}$$

provided that ψ satisfies *Calderón's admissibility condition* [**Cal**]:

$$C_\psi := \int_0^\infty \frac{|\widehat{\psi}\,(\xi)|^2}{\xi}\,d\xi < \infty.$$

This reconstruction formula can be traced back to the famous *Calderón reproducing formula*[6]:

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \big(\psi_{s,0} * \overline{\widetilde{\psi_{s,0}}} * f\big)(x)\,\frac{ds}{s^2}, \tag{9.6}$$

where $\widetilde{\psi}(x) = \psi(-x)$ and $*$ denotes convolution in $\mathbb{R}$.

Exercise 9.23. Show that the Calderón reproducing formula (9.6) and the reconstruction formula (9.5) are the same. Show that equation (9.6) holds for f, $\psi \in \mathcal{S}(\mathbb{R})$ such that $C_\psi < \infty$, by checking that the Fourier transform of the right-hand side coincides with $\widehat{f}$. ♢

[6] Named after Argentinian mathematician Alberto Pedro Calderón (1920–1998).

9.4. Haar analysis

In this section we show that the Haar functions form a complete orthonormal system. Verifying the orthonormality of the system reduces to understanding the geometry of the dyadic intervals. Verifying the completeness of the system reduces to understanding that the limit in $L^2(\mathbb{R})$ of the averaging operators over intervals as the intervals shrink to a point $x \in \mathbb{R}$ is the identity operator and that the limit as the intervals grow to be infinitely long is the zero operator.

We first define the dyadic intervals and describe their geometry. We define the Haar function associated to an interval and observe that the family of Haar functions indexed by the dyadic intervals coincides with the Haar basis from Example 9.19. We show that the Haar functions are orthonormal. Next, we introduce the expectation (or averaging) and difference operators, relate the completeness of the Haar system to the limiting behavior of the averaging operators, and prove the required limit results for continuous functions and compactly supported functions. Finally, an approximation argument coupled with some uniform bounds gives the desired result: the completeness of the Haar system. Along the way we mention the Lebesgue Differentiation Theorem and the Uniform Boundedness Principle and some of their applications.

9.4.1. The dyadic intervals. Given an interval $I = [a, b)$ in $\mathbb{R}$, the *left half*, or *left child*, of I is the interval $I_l := [a, (a + b)/2)$, and the *right half*, or *right child*, of I is the interval $I_r := [(a + b)/2, b)$ (Figure 9.5). Let $|I|$ denote the length of an interval I.

Given a locally integrable function f, let $m_I f$ denote the *average value*[7] *of f over I*:

$$m_I f := \frac{1}{|I|} \int_I f(x)\, dx. \tag{9.7}$$

Definition 9.24 (*Dyadic Intervals*). The *dyadic intervals* are the half-open intervals of the form

$$I_{j,k} = [k2^{-j}, (k+1)2^{-j}) \qquad \text{for integers } j,\ k.$$

[7]Other popular notations for $m_I f$ are $\langle f \rangle_I$ and f_I.

Let $\mathcal{D}$ denote the set of all dyadic intervals in $\mathbb{R}$, and $\mathcal{D}_j$ the set of intervals $I \in \mathcal{D}$ of length 2^{-j}, also called the j^{th} *generation*. Then $\mathcal{D} = \bigcup_{j\in\mathbb{Z}} \mathcal{D}_j$. Each $\mathcal{D}_j$ forms a partition of the real line. ♢

For instance, $[5/8, 3/4)$ and $[-16, -12)$ are dyadic intervals, while $[3/8, 5/8)$ is not. If I is a dyadic interval, then its children, grandchildren, parents, grandparents, and so on are also dyadic intervals.

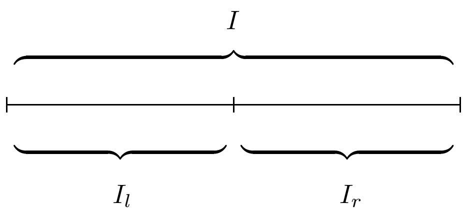


Figure 9.5. Parent interval I and its children I_l and I_r.

Each dyadic interval I belongs to a unique generation $\mathcal{D}_j$, and the next generation $\mathcal{D}_{j+1}$ contains exactly two subintervals of I, namely I_l and I_r. Given two distinct intervals I, $J \in \mathcal{D}$, either I and J are disjoint or one is contained in the other. This nestedness property of the dyadic intervals is so important that we highlight it as a lemma.

Lemma 9.25 (*Dyadic Intervals Are Nested or Disjoint*)**.** *If I, $J \in \mathcal{D}$, then exactly one of the following holds: $I \cap J = \emptyset$ or $I \subseteq J$ or $J \subsetneq I$. Moreover if $J \subsetneq I$, then J is a subset of the left or the right child of I.*

Exercise 9.26. Prove Lemma 9.25. Also show that its conclusion need not hold for nondyadic intervals. ♢

Given $x \in \mathbb{R}$ and $j \in \mathbb{Z}$, there is a unique interval in $\mathcal{D}_j$ that contains x. We denote this unique interval by $I_j(x)$. Figure 9.8 on page 244, shows several dyadic intervals $I_j(x)$ for a fixed point x; they form a *dyadic tower* containing the point x. The intervals $I_j(x)$ shrink to the set $\{x\}$ as $j \to \infty$.

9.4.2. The Haar basis. We associate to each dyadic interval I a step function h_I that is supported on I, is constant on each of I_l

and I_r, has zero integral average, and has L^2 norm equal to one. The zero integral average makes the value of h_I on the left child be the negative of the value on the right child. The L^2 normalization makes the absolute value of the function on I be exactly $1/\sqrt{|I|}$. We make the convention that the value is negative on I_l and thus positive on I_r. These considerations uniquely determine h_I.

Definition 9.27. The *Haar function associated to the interval I* is the step function h_I defined by

$$h_I(x) := (1/\sqrt{|I|})\big(\chi_{I_r}(x) - \chi_{I_l}(x)\big). \qquad \diamond$$

The Haar wavelet h defined in Example 9.19 coincides with the Haar function $h_{[0,1)}$ associated to the unit interval $[0,1)$. The Haar wavelets $h_{j,k}$ coincide with $h_{I_{j,k}}$, where $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$. Figure 9.6 shows the graphs of two Haar functions.

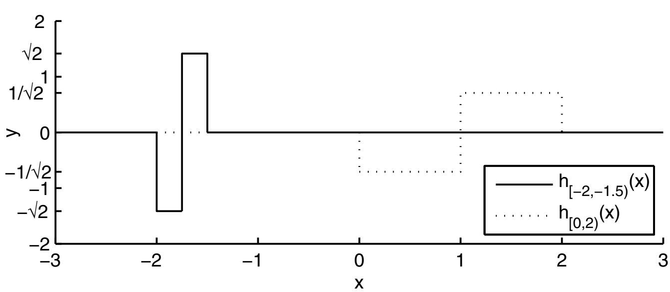


Figure 9.6. Graphs of the two Haar functions defined by $h_{[-2,-1.5)}(x) = \sqrt{2}\,\big[\chi_{[-1.75,-1.5)}(x) - \chi_{[-2,-1.75)}(x)\big]$ and $h_{[0,2)}(x) = (1/\sqrt{2})\,\big[\chi_{[1,2)}(x) - \chi_{[0,1)}(x)\big]$.

Exercise 9.28. Show that

$$h_{I_{j,k}}(x) = 2^{j/2}h(2^j x - k) = h_{j,k}(x), \qquad \text{where } h = h_{[0,1)}. \qquad \diamond$$

By the results of Exercises 9.20 and 9.28, the set $\{h_I\}_{I\in\mathcal{D}}$ is an orthonormal set in $L^2(\mathbb{R})$. We can also prove directly that the Haar functions indexed by the dyadic intervals form an orthonormal family, using the fact that the dyadic intervals are nested or disjoint.

Lemma 9.29. *The family of Haar functions $\{h_I\}_{I\in\mathcal{D}}$ indexed by the dyadic intervals is an orthonormal family.*

Proof. Consider $I,\ J \in \mathcal{D}$ with $I \neq J$. Either they are disjoint or one is strictly contained in the other, by Lemma 9.25. If I and J are disjoint, then clearly $\langle h_I, h_J\rangle = 0$, because the supports are disjoint. If I is strictly contained in J, then h_J is constant on I, and in fact

$$\langle h_I, h_J\rangle = \int_I h_I(x)h_J(x)\,dx = \frac{1}{|J|^{1/2}}\int h_I(x)\,dx = 0.$$

The case $J \subsetneq I$ is similar. This proves the orthogonality of the family. For the normality, when $I = J$ we have

$$\langle h_I, h_I\rangle = \|h_I\|_{L^2(\mathbb{R})} = \int_I |h_I(x)|^2\,dx = \frac{1}{|I|}\int_I dx = 1. \qquad \square$$

The Haar system is not only orthonormal but also a complete orthonormal system and hence a basis for $L^2(\mathbb{R})$, as we now show.

Theorem 9.30. *The Haar functions $\{h_I\}_{I\in\mathcal{D}}$ form an orthonormal basis for $L^2(\mathbb{R})$.*

In an N-dimensional space, an orthonormal set of N vectors is automatically a basis. Thus in Chapter 6 to show that the Haar vectors form a basis for $\mathbb{C}^N$, it was enough to show that there are N orthonormal Haar vectors. In infinite-dimensional space we have to do more than simply counting an orthonormal set. To prove Theorem 9.30, we must make sure the set is *complete*. In other words, for all $f \in L^2(\mathbb{R})$, the identity

$$f = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I \tag{9.8}$$

must hold in the L^2 sense[8]. We must show that each $L^2(\mathbb{R})$ function f can be written as a (possibly infinite) sum of Haar functions, weighted by coefficients given by the inner products of f with h_I. (See for example the lower figure on the cover, where a simple signal is broken into a weighted sum of Haar functions.)

An alternative proof shows that the only function in $L^2(\mathbb{R})$ orthogonal to all Haar functions is the zero function (see Theorem A.41).

[8]Recall that equation (9.8) holds in the L^2 sense if $\|f - \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I\|_2 = 0$.

We will show that both arguments boil down to checking some limit properties of the expectation operators defined in Section 9.4.4.

9.4.3. Devil's advocate. Before we prove the completeness of the Haar system, let us play devil's advocate.

- First, for the function $f(x) \equiv 1$ we have $\langle f, h_I \rangle = \int h_I = 0$ for all $I \in \mathcal{D}$. Thus f is orthogonal to all the Haar functions, so how can the Haar system be complete? Are we contradicting Theorem 9.30? No, because f is not in $L^2(\mathbb{R})$!
- Second, how can functions that all have zero integral (the Haar functions) reconstruct functions that do not have zero integral?[9]

If the Haar system is complete, then for each $f \in L^2(\mathbb{R})$, equation (9.8) holds. Integrating on both sides and interchanging the sum and the integral, we see that

$$\text{``}\int_{\mathbb{R}} f(x)\,dx = \int_{\mathbb{R}} \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I(x)\,dx = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle \int_{\mathbb{R}} h_I(x)\,dx = 0\text{''}.$$

The last equality holds because the Haar functions have integral zero. It seems that all functions in $L^2(\mathbb{R})$ must themselves have integral zero. But that is not true, since for example $\chi_{[0,1)}$ belongs to $L^2(\mathbb{R})$ yet has integral one. What's wrong? Perhaps the Haar system is not complete after all. Or is there something wrong in our calculation? The Haar system *is* complete; it turns out that what is illegal above is the interchange of sum and integral. See the project in Section 9.7.

9.4.4. The expectation and difference operators, P_j and Q_j. We introduce two important *operators*[10] that will help us to understand the zooming properties of the Haar basis.

[9] This question was posed by Lindsay Crowl, a student in the 2004 Program for Women in Mathematics, where we gave the lectures that led to this book. It was a natural concern and a good example of the dangers of interchanging limit operations!

[10] An *operator* is a mapping from a space of functions into another space of functions. The input is a function and so is the output. The Fourier transform is an example of an operator.

Definition 9.31. The *expectation* operators $P_j : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $j \in \mathbb{Z}$, act by taking averages over dyadic intervals at generation j:

$$P_j f(x) := \frac{1}{|I_j|} \int_{I_j} f(t)\, dt,$$

where $I_j = I_j(x)$ is the unique interval of length 2^{-j} that contains x. ◊

The new function $P_j f$ is a step function that is constant on each dyadic interval $I \in \mathcal{D}_j$ in the j^{th} generation. Furthermore the value of the function $P_j f$ on an interval $I \in \mathcal{D}_j$ is the integral average of the function f over the interval I: for each $x \in I$,

$$P_j f(x) \equiv m_I f := \frac{1}{|I|} \int_I f(y)\, dy. \tag{9.9}$$

Figure 9.7 (upper plot) shows the graph of a particular function f, together with the graphs of $P_j f$ and $P_{j+1} f$, for $j = -1$.

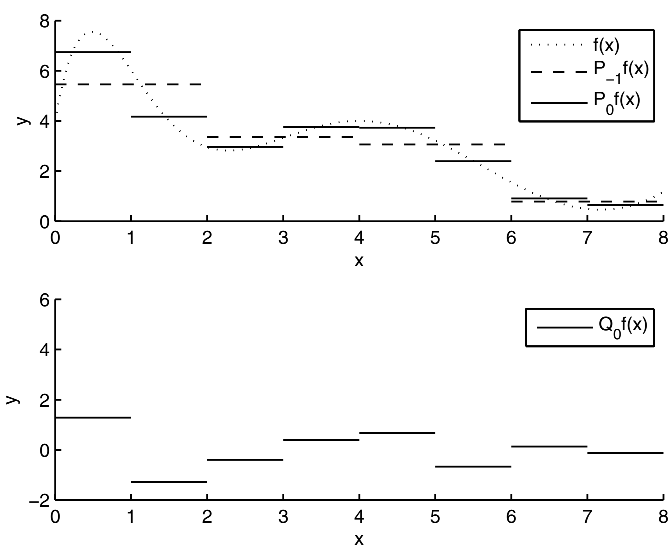


Figure 9.7. Graphs of f, $P_{-1}f$, $P_0 f$, and $Q_0 f$. We have used the function $f(x) = 4 + x(x-1.5)(x-4)^2(x-9) \times e^{-x/2.5}/12$.

Exercise 9.32. Verify that $P_j f(x) = \sum_{I\in\mathcal{D}_j} m_I f\, \chi_I(x)$. ♢

As $j \to \infty$, the length $|I| = 2^{-j}$ of the steps goes to zero, and we expect $P_j f$ to be a better and better approximation of f. We make this idea precise in Section 9.4.5 below.

Definition 9.33. The *difference operators* $Q_j : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $j \in \mathbb{Z}$, are given by $\quad Q_j f(x) := P_{j+1} f(x) - P_j f(x)$. ♢

These operators Q_j encode the information necessary to go from the approximation $P_j f$ at resolution j of f to the better approximation $P_{j+1} f$ at resolution $j+1$. Figure 9.7 (lower plot) shows the graph of $Q_j f$ for $j = 0$.

Notice that when we superimpose the pictures of $P_{j+1} f$ and $P_j f$, the averages at the coarser scale j seem to be sitting exactly halfway between the averages at the finer scale $j + 1$, so that $Q_j f$ seems to be a linear combination of the Haar functions at scale j. Lemma 9.35 makes this fact precise. It is implied by the following useful relationship between integral averages on nested dyadic intervals.

Exercise 9.34. Show that $m_I f = (m_{I_l} f + m_{I_r} f)/2$. In words, the integral average over a parent interval is the average of the integral averages over its children. ♢

Lemma 9.35. *For $f \in L^2(\mathbb{R})$, $Q_j f(x) = \sum_{I\in\mathcal{D}_j} \langle f, h_I\rangle h_I(x)$.*

Proof. By definition of h_I and since $|I| = 2|I_r| = 2|I_l|$, we have

$$\begin{aligned}\langle f, h_I\rangle h_I(x) &= \frac{\sqrt{|I|}}{2}\Big(\frac{1}{|I_r|}\int_{I_r} f - \frac{1}{|I_l|}\int_{I_l} f\Big) h_I(x)\\ &= \sqrt{|I|}/2\,(m_{I_r} f - m_{I_l} f)\, h_I(x).\end{aligned}$$

Since $h_I(x) = 1/\sqrt{|I|}$ for $x \in I_r$ and $h_I(x) = -1/\sqrt{|I|}$ for $x \in I_l$, we conclude that if $x \in I$, then

$$\langle f, h_I\rangle h_I(x) = \begin{cases} (m_{I_r} f - m_{I_l} f)/2, & \text{if } x \in I_r;\\ -(m_{I_r} f - m_{I_l} f)/2, & \text{if } x \in I_l.\end{cases} \tag{9.10}$$

On the other hand, if $x \in I \in \mathcal{D}_j$, then $P_j f(x) = m_I f$; if $x \in I_r$, then $P_{j+1} f(x) = m_{I_r} f$; and if $x \in I_l$, then $P_{j+1} f(x) = m_{I_l} f$. Hence

$$Q_j f(x) = \begin{cases} m_{I_r} f - m_I f, & \text{if } x \in I_r; \\ m_{I_l} f - m_I f, & \text{if } x \in I_l. \end{cases}$$

The averaging property in Exercise 9.34 implies that

$$m_{I_r} f - m_I f = (m_{I_r} f - m_{I_l} f)/2 = m_I f - m_{I_l} f.$$

We conclude that $Q_j f(x) = \sum_{I \in \mathcal{D}_j} \langle f, h_I \rangle h_I(x)$ for $x \in I \in \mathcal{D}_j$. □

9.4.5. Completeness of the Haar system. To prove that the Haar system of functions is complete, we must show that for all $f \in L^2(\mathbb{R})$, we have

$$f(x) = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_I(x).$$

By Lemma 9.35, this condition is equivalent to the condition

$$f(x) = \lim_{M,N \to \infty} \sum_{-M \leq j < N} Q_j f(x).$$

A telescoping series argument shows that

$$\begin{aligned} P_N f(x) - P_M f(x) &= \sum_{M \leq j < N} (P_{j+1} f(x) - P_j f(x)) \\ &= \sum_{M \leq j < N} Q_j f(x). \end{aligned} \tag{9.11}$$

Therefore, verifying completeness reduces to checking that

$$f(x) = \lim_{N \to \infty} P_N f(x) - \lim_{M \to -\infty} P_M f(x),$$

where all the above equalities hold in the L^2 sense. It suffices to prove the following theorem.

Theorem 9.36. *For $f \in L^2(\mathbb{R})$, we have*

$$\lim_{M \to -\infty} \|P_M f\|_2 = 0 \quad \text{and} \tag{9.12}$$

$$\lim_{N \to \infty} \|P_N f - f\|_2 = 0. \tag{9.13}$$

Exercise 9.37. Use Theorem 9.36 to show that if $f \in L^2(\mathbb{R})$ is orthogonal to all Haar functions, then f must be zero in $L^2(\mathbb{R})$. ◊

Aside 9.38. Before proving Theorem 9.36, we pause to develop the Lebesgue Differentiation Theorem on $\mathbb{R}$. Here $I_j(x)$ is the unique dyadic interval in $\mathcal{D}_j$ that contains x, as shown in Figure 9.8. Equation (9.13) says that given $x \in \mathbb{R}$, the averages $P_N f$ of the function f over the dyadic intervals $\{I_j(x)\}_{j\in\mathbb{Z}}$ converge to $f(x)$ in the L^2 sense as the intervals shrink: $\lim_{j\to\infty} \frac{1}{|I_j(x)|}\int_{I_j(x)} f(t)\,dt = f(x)$. In fact, the convergence also holds *almost everywhere*[11] (a.e.).

Theorem 9.39 (Lebesgue Differentiation Theorem on $\mathbb{R}$)**.** *Suppose f is locally integrable: $f \in L^1_{\text{loc}}(\mathbb{R})$. Let I denote any interval, dyadic or not, that contains x. Then*

$$\lim_{x\in I, |I|\to 0} \frac{1}{|I|}\int_I f(y)\,dy = f(x) \quad \textit{for a.e. } x \in \mathbb{R}.$$

For a proof see [**SS05**, Chapter 3]. In Chapter 4 we stated the special case with intervals $[x-h, x+h]$ centered at x, as $h \to 0$. ♢

Exercise 9.40. Prove the *Lebesgue Differentiation Theorem* for continuous functions, and show that for continuous functions the pointwise convergence holds everywhere. That is, for all $x \in \mathbb{R}$ and intervals $[a, b]$ containing x, $\lim_{[a,b]\to\{x\}} \frac{1}{b-a}\int_a^b f(t)\,dt = f(x)$. ♢

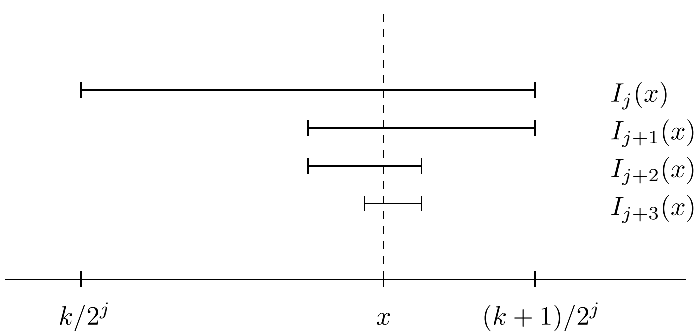


Figure 9.8. The tower $\cdots \supset I_j(x) \supset I_{j+1}(x) \supset I_{j+2}(x) \supset I_{j+3}(x) \supset \cdots$ of nested dyadic intervals containing the point x, with $k/2^j \le x \le (k+1)/2^j$. Here j and k are integers.

[11]In other words, the convergence holds except possibly on a set of measure zero.

Theorem 9.36 is a consequence of the following three lemmas.

Lemma 9.41. *The operators P_j are* uniformly bounded *in $L^2(\mathbb{R})$. Specifically, for each function $f \in L^2(\mathbb{R})$ and for every integer j,*

$$\|P_j f\|_2 \leq \|f\|_2.$$

Lemma 9.42. *If g is continuous and has compact support on the interval $[-K, K]$, then Theorem 9.36 holds for g.*

Lemma 9.43. *The continuous functions with compact support are* dense *in $L^2(\mathbb{R})$. In other words, given $f \in L^2(\mathbb{R})$, for any $\varepsilon > 0$ there exist functions g and h, such that $f = g + h$, where g is a continuous function with compact support on an interval $[-K, K]$ and $h \in L^2(\mathbb{R})$ with small L^2 norm, $\|h\|_2 < \varepsilon$.*

We first deduce the theorem from the lemmas and then after some comments prove the lemmas.

Proof of Theorem 9.36. By Lemma 9.43, given $\varepsilon > 0$, we can decompose f as $f = g + h$, where g is continuous with compact support on $[-K, K]$ and $h \in L^2(\mathbb{R})$ with $\|h\|_2 < \varepsilon/4$. By Lemma 9.42, we can choose N large enough so that for all $j > N$,

$$\|P_{-j} g\|_2 \leq \varepsilon/2.$$

The expectation operators P_j are linear operators, so $P_j(g + h) = P_j g + P_j h$. Now by the Triangle Inequality and Lemma 9.41 we have

$$\|P_{-j} f\|_2 \leq \|P_{-j} g\|_2 + \|P_{-j} h\|_2 \leq \frac{\varepsilon}{2} + \|h\|_2 \leq \varepsilon.$$

This proves equation (9.12). Similarly, by Lemma 9.42, we can choose N large enough that for all $j > N$,

$$\|P_j g - g\|_2 \leq \varepsilon/2.$$

Now using the Triangle Inequality (twice) and Lemma 9.41,

$$\|P_j f - f\|_2 \leq \|P_j g - g\|_2 + \|P_j h - h\|_2 \leq \frac{\varepsilon}{2} + 2\|h\|_2 \leq \varepsilon.$$

This proves equation (9.13). □

Exercise 9.44. Show that the operators P_j are linear. ♢

We have twice used a very important principle from functional analysis: *If a sequence of linear operators is uniformly bounded on a Banach space and the sequence converges to a bounded operator on a dense subset of the Banach space, then it converges on the whole space to the same operator.* For us the Banach space is $L^2(\mathbb{R})$, the dense subset is the set of continuous functions with compact support, the linear operators are P_j, and the uniform bounds are provided by Lemma 9.41. The operators converge to the zero operator as $j \to -\infty$, and they converge to the identity operator as $j \to \infty$.

A related principle is the *Uniform Boundedness Principle*, or the *Banach–Steinhaus Theorem*, from functional analysis. See for example [**Sch**, Chapter III]. This principle gives sufficient conditions for a family of operators to be uniformly bounded. A beautiful application of the Uniform Boundedness Principle is to show the existence of a real-valued continuous periodic function whose Fourier series diverges at a given point x_0. See Aside 9.48.

We now prove the three lemmas that implied Theorem 9.36, which in turn gave us the completeness of the Haar system.

Proof of Lemma 9.41. We can estimate for $x \in I \in \mathcal{D}_j$,

$$|P_j f(x)|^2 = \Big|\frac{1}{|I|}\int_I f(t)\,dt\Big|^2 \le \frac{1}{|I|^2}\Big(\int_I 1^2\,dt\Big)\Big(\int_I |f(t)|^2\,dt\Big)$$
$$= \frac{1}{|I|}\int_I |f(t)|^2\,dt.$$

The inequality is a consequence of the Cauchy–Schwarz Inequality.

Now integrate over the interval I to obtain

$$\int_I |P_j f(x)|^2\,dx \le \int_I |f(t)|^2\,dt,$$

and sum over all intervals in $\mathcal{D}_j$ (this is a disjoint family that covers the whole line!):

$$\int_{\mathbb{R}} |P_j f(x)|^2\,dx = \sum_{I\in\mathcal{D}_n}\int_I |P_j f(x)|^2\,dx \le \sum_{I\in\mathcal{D}_n}\int_I |f(t)|^2\,dt$$
$$= \int_{\mathbb{R}} |f(t)|^2\,dt,$$

as required. We have used Lemma A.51 twice here. □

Proof of Lemma 9.42. Suppose that the function g is continuous and that it is supported on the interval $[-K, K]$. If j is large enough that $K < 2^j$ and if $x \in [0, 2^j) \in \mathcal{D}_{-j}$, then $|P_{-j}g(x)| = \frac{1}{2^j}\int_0^K |g(t)|dt$, and applying the Cauchy–Schwarz Inequality, we get

$$|P_{-j}g(x)| \leq \frac{1}{2^j}\Big(\int_0^K 1^2\,dt\Big)^{1/2}\Big(\int_0^K |g(t)|^2\,dt\Big)^{1/2} \leq \frac{\sqrt{K}}{2^j}\|g\|_2.$$

The same inequality holds for $x < 0$. Also, if $|x| \geq 2^j$, then $P_{-j}g(x) = 0$, because the interval in $\mathcal{D}_j$ that contains x is disjoint with the support of g. We can now estimate the L^2 norm of $P_{-j}g$:

$$\|P_{-j}g\|_2^2 = \int_{-2^j}^{2^j} |P_{-j}g(x)|^2\,dx \leq \frac{1}{2^{2j}}K\|g\|_2^2\int_{-2^j}^{2^j} 1\,dx = 2^{-j+1}K\|g\|_2^2.$$

By choosing N large enough, we can make $2^{-N+1}K\|g\|_2^2 < \varepsilon^2$. That is, given $\varepsilon > 0$, there is an $N > 0$ such that for all $j > N$,

$$\|P_{-j}g\|_2 \leq \varepsilon.$$

This proves equation (9.12) for continuous functions with compact support.

We are assuming that g is continuous and that it is supported on the compact interval $[-K, K] \subset [-2^M, 2^M]$. But then g is uniformly continuous. So given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|g(y) - g(x)| < \varepsilon/\sqrt{2^{M+1}} \quad \text{whenever } |y - x| < \delta.$$

Now choose $N > M$ large enough that $2^{-j} < \delta$ for all $j > N$. Each point x is contained in a unique $I \in \mathcal{D}_j$, with $|I| = 2^{-j} < \delta$. Therefore $|y - x| \leq \delta$ for all $y \in I$, and

$$|P_jg(x) - g(x)| \leq \frac{1}{|I|}\int_I |g(y) - g(x)|\,dy \leq \frac{\varepsilon}{\sqrt{2^{M+1}}}.$$

Squaring and integrating over $\mathbb{R}$, we get

$$\begin{aligned}\int_{\mathbb{R}} |P_jg(x) - g(x)|^2\,dx &= \int_{|x|\leq 2^M} |P_jg(x) - g(x)|^2\,dx \\ &< \frac{\varepsilon^2}{2^{M+1}}\int_{|x|\leq 2^M} 1\,dx = \varepsilon^2.\end{aligned}$$

Notice that if $|x| > 2^M$, then for $j > N \geq M$, $P_jg(x)$ is the average over an interval $I \in \mathcal{D}_j$ that is completely outside the support of g. For such x and j, $P_jg(x) = 0$, and therefore there is zero contribution to the integral from $|x| > 2^M$. Lo and behold, we have shown that given $\varepsilon > 0$, there is an $N > 0$ such that for all $n > N$,

$$\|P_j g - g\|_2 \leq \varepsilon.$$

This proves equation (9.13) for continuous functions with compact support. □

Proof of Lemma 9.43. This lemma is an example of an approximation theorem in $L^2(\mathbb{R})$. First, we choose K large enough so that the tail of f has very small L^2 norm, in other words, $\|f\chi_{\{x\in\mathbb{R}:|x|>K\}}\|_2 \leq \varepsilon/3$. Second, we recall that on compact intervals, the continuous functions are dense in $L^2([-K,K])$; see Theorem 2.75. (For example, polynomials are dense, and trigonometric polynomials are also dense, by the Weierstrass Approximation Theorem (Theorem 3.4).) Now choose g_1 continuous on $[-K,K]$ so that $\|(f-g_1)\chi_{[-K,K]}\|_2 \leq \varepsilon/3$. Third, it could happen that g_1 is continuous on $[-K,K]$, but when extended to be zero outside the interval, it is not continuous on the line. That can be fixed by giving yourself some margin at the endpoints: define g to coincide with g_1 on $[-K+\delta, K-\delta]$ and to be zero outside $[-K,K]$, and connect these pieces with straight segments, so that g is continuous on $\mathbb{R}$. Finally, choose δ small enough so that $\|g_1 - g\|_2 \leq \varepsilon/3$. Now let

$$h = f - g = f\chi_{\{x\in\mathbb{R}:|x|>K\}} + (f-g_1)\chi_{[-K,K]} + (g_1 - g)\chi_{[-K,K]}.$$

By the Triangle Inequality,

$$\|h\|_2 \leq \|f\chi_{\{x\in\mathbb{R}:|x|>K\}}\|_2 + \|(f-g_1)\chi_{[-K,K]}\|_2 + \|g_1 - g\|_2 \leq \varepsilon. \quad \square$$

We have shown (Lemma 9.42) that the step functions can approximate continuous functions with compact support in the L^2 norm. Lemma 9.43 shows that we can approximate L^2 functions by continuous functions with compact support. Therefore, we can approximate L^2 functions by step functions, in the L^2 norm. Furthermore, we can choose the steps to be dyadic intervals of a fixed generation, for any prescribed accuracy.

Exercise 9.45 (*Approximation by Step Functions*)**.** Show that continuous functions with compact support can be approximated in the uniform norm by step functions (with compact support). Furthermore, one can choose the intervals where the approximating function is constant to be dyadic intervals of a fixed generation for any prescribed accuracy. More precisely, show that given f continuous and compactly supported on $\mathbb{R}$ and $\varepsilon > 0$, there exists $N > 0$ such that for all $j > N$ and for all $x \in \mathbb{R}$, $|P_j f(x) - f(x)| < \varepsilon$. ♢

Aside 9.46. The Lebesgue Differentiation Theorem (Exercise 9.40) holds for continuous functions, as a consequence of the Fundamental Theorem of Calculus. A related function, defined for all $x \in \mathbb{R}$ and each locally integrable function f (possibly as ∞), is the *Hardy–Littlewood maximal function*[12] denoted Mf and defined by

$$Mf(x) = \sup_{x\in I} \frac{1}{|I|} \int_I |f(t)|\,dt. \tag{9.14}$$

The supremum in the definition of $Mf(x)$ is taken over all intervals containing x. Notice that the supremum is always defined, unlike the limit in the Lebesgue Differentiation Theorem as the intervals shrink to the point x; *a priori* that limit might not exist. The Lebesgue Differentiation Theorem guarantees both that the limit does exist and that it equals the value of the function at the point x, for almost every x. Boundedness and weak boundedness results for the maximal function (see the project in Section 12.9.2) can be used to deduce the Lebesgue Differentiation Theorem; see [**SS05**, Chapter 3, Section 1].

Maximal functions appear throughout harmonic analysis as controllers for other operators T, in the sense that $\|T\| \leq \|M\|$. It is important to understand their boundedness properties. For example, all convolution operators with good radial kernels can be bounded pointwise by a constant multiple of $Mf(x)$.

Exercise 9.47. Let $k_t \in L^1(\mathbb{R})$ be defined for $t > 0$ by $k_t(x) := t^{-1}K(|y|/t)$, where $K : [0,\infty) \to [0,\infty)$ is a monotone decreasing nonnegative function. This is a family of good kernels in $\mathbb{R}$; see Section 7.8. Show that the following pointwise estimate holds:

$$|K_t * f(x)| \leq Mf(x)\,\|k_1\|_1. \qquad ♢$$

[12] Named after the British mathematicians Godfrey Harold (G. H.) Hardy (1877–1947) and John Edensor Littlewood (1885–1977).

As examples of such kernels, consider the *heat kernel* and the *Poisson kernel* on $\mathbb{R}$, introduced in Section 7.8. In these examples, $K(x) = e^{-|x|^2/4}/\sqrt{4\pi}$ for the heat kernel, and $K(x) = 1/(1+|x|^2)$ for the *Poisson kernel.* A similar argument to the one that gives the Lebesgue Differentiation Theorem implies that if k_t is the heat kernel or the Poisson kernel, then $\lim_{t\to 0} k_t * f(x) = f(x)$ for almost every $x \in \mathbb{R}$ and for every $f \in L^1(\mathbb{R})$. See [**Pin**, Sections 3.5.1 and 3.5.2]. We already knew this fact for continuous functions, since in that case uniform convergence to f was proven in Chapter 7. $\lozenge$

Exercise 9.48. We outline an existence proof (but not a constructive proof) that there is a continuous periodic function whose Fourier series diverges at some point. See the project in Section 2.5 for a constructive proof.

Theorem 9.49 (Uniform Boundedness Principle)**.** *Let W be a family of bounded linear operators $T: X \to Y$ from a Banach space X into a normed space Y, such that for each $x \in X$, $\sup_{T\in W} \|Tx\|_Y < \infty$. Then the operators are uniformly bounded: there is a constant $C > 0$ such that for all $T \in W$ and all $x \in X$, $\|Tx\|_Y \leq C\|x\|_X$.*

Now let X be the Banach space $C(\mathbb{T})$ of all real-valued continuous functions of period 2π with uniform norm, and let $Y = \mathbb{C}$. For $f \in C(\mathbb{T})$, define $T_N(f) := S_N f(0) \in \mathbb{C}$, where $S_N f$ is the N^{th} partial Fourier sum of f. Then $S_N f = D_N * f$ where D_N is the periodic Dirichlet kernel. Show that if $|T_N f| \leq C_N \|f\|_\infty$, then necessarily $C_N \geq c\|D_N\|_{L^1(\mathbb{T})}$. But as we showed in Chapter 4, $\|D_N\|_{L^1(\mathbb{T})} \approx \log N$. It follows that the operators T_N cannot be uniformly bounded, so there must be some $f \in C(\mathbb{T})$ such that $\sup_{N\geq 0} |S_N f(0)| = \infty$. The partial Fourier sums of this function f diverge at $x = 0$. $\lozenge$

9.5. Haar vs. Fourier

We give two examples to illustrate how the Haar basis can outperform the Fourier basis. The first deals with localized data. The second is related to the unconditionality of the Haar basis in $L^p(\mathbb{R})$ for all $1 < p < \infty$, in contrast to the trigonometric basis which is not an unconditional basis for $L^p([0,1]$ except when $p = 2$. We take the opportunity to introduce some operators that are important in harmonic analysis: the *martingale transform*, the *square function*, and *Petermichl's shift operator*[13]. We deduce the unconditionality in L^p of the Haar basis from boundedness properties of these dyadic operators.

9.5.1. Localized data. The first example is a caricature of the problem: What is the most localized "function" we could consider? The answer is the delta distribution. If we could find the Fourier series of a periodic delta distribution, we would see that it has a very slowly decreasing tail that extends well beyond the highly localized support of the delta function. However, its Haar transform is very localized; although the Haar transform still has a tail, the tail decays faster than that of the Fourier series. We try to make this impressionistic comment more precise in what follows.

Consider the following approximation of the delta distribution:

$$f_N(x) = 2^N \chi_{[0,2^{-N}]}.$$

Each of these functions has mass 1, and they converge in the sense of distributions to the delta distribution:

$$\lim_{N\to\infty} T_{f_N}(\phi) = \lim_{N\to\infty} \int f_N(x)\phi(x)\,dx = \phi(0) = \delta(\phi),$$

by the Lebesgue Differentiation Theorem for continuous functions (Exercise 9.40).

Exercise 9.50. Compute the Fourier transform of f_N. Show that if we view f_N as a periodic function on $[0,1)$, with M^{th} partial Fourier sum $S_M(f_N)(x) = \sum_{|m|\le M} \widehat{f_N}(m)\, e^{2\pi i m x}$, then the following rate of decay holds: $\|f_N - S_M(f_N)\|_{L^2([0,1))} \sim 1/\sqrt{M}$. $\Diamond$

We want to compare to the Haar decomposition, working on the interval $[0,1)$. We have to be a little careful in view of the next exercise.

Exercise 9.51. Show that the set $\{h_I\}_{I\in\mathcal{D}([0,1])}$ is not a complete set in $L^2([0,1])$. What are we missing? Can you complete the set? You must show that the set is now complete. $\Diamond$

[13] Named after German mathematician Stefanie Petermichl (born 1971).

The Haar basis on the interval $[0,1)$ consists of all the Haar functions indexed on dyadic subtintervals of $[0,1)$ and the characteristic function $\chi_{[0,1)}$ of the interval $[0,1)$. The function $\chi_{[0,1)}$ is orthonormal to all the Haar functions selected, so it must be part of an orthonormal basis.

Exercise 9.52. Compute the Haar coefficients of f_N viewed as a function on $[0,1)$. Check that the partial Haar sum $P_M(f_N)$, defined by $P_M(f_N) := \sum_{0\leq j<M} Q_j(f_N) + P_0(f_N)$, decays exponentially in M: $\|f_N - P_M(f_N)\|_{L^2([0,1))} = 2^{-M/2}$. ♢

The exponential decay rate seen in Exercise 9.52 is much better than the square-root decay in Exercise 9.50. Suppose we want to approximate F_N with an L^2 error of magnitude less than 10^{-5}. In the Fourier case we would need a partial Fourier sum of order $M \sim 10^{10}$. In the Haar case it suffices to consider $M = 10\log_2 10 < 40$. However, beware: to move from trigonometric polynomials of degree M to degree $M+1$, we just add two functions $e^{\pm 2\pi i(M+1)}$, while to move from generation M to generation $M+1$, we need 2^M Haar functions, so this calculation is deceptive.

However, the localization properties of the Haar functions allow us to worry only about the Haar functions that are supported near where the action occurs in the function. For example, if the function is constant on an interval, then the Haar coefficients corresponding to the Haar functions supported in the interval vanish, because they have the zero integral property. This is not the case with the trigonometric functions, whose support spreads over the whole real line. In the example discussed, the function f_N is supported on the interval $[0, 2^{-N})$ and is constant there. So in fact the number of Haar functions in a given generation whose support intersects the support of f_N and which are not completely inside it is just one. In both the Fourier and the Haar cases, the parameter M counts the same number of basis functions contributing to the estimate. Now the difference in the estimate of the error is dramatic when comparing the rates of convergence. The same phenomenon is observed for compactly supported wavelets other than the Haar wavelet.

See the project in Section 10.6 for more on linear and nonlinear approximations. The lesson from this example is that when there is good localization, the wavelet basis will perform better than the Fourier basis.

9.5.2. Unconditional bases and dyadic operators. The Haar basis is an *unconditional basis* for $L^p(\mathbb{R})$, for $1 < p < \infty$. See the Appendix for the formal definition. Informally, we can approximate a function in the L^p norm with an infinite linear combination of Haar functions (basis), and the order of summation doesn't matter (conditional convergence). Further, the coefficient of h_I must be $\langle f, h_I\rangle$, and we can recover the L^p norm of the function from knowledge about the *absolute value* of these coefficients, that is, using some formula involving only $|\langle f, h_I\rangle|$ for each I. No information about the sign or argument of $\langle f, h_I\rangle$ is necessary. In particular, if $f \in L^p(\mathbb{R})$ and

$$f(x) = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I(x),$$

then the new functions defined by

$$T_\sigma f(x) := \sum_{I\in\mathcal{D}} \sigma_I \langle f, h_I\rangle h_I(x), \quad \text{where } \sigma_I = \pm 1, \tag{9.15}$$

are also in $L^p(\mathbb{R})$ and their norms are comparable to that of f. We will make the statement precise in the next theorem.

Definition 9.53. For a given sequence $\sigma = \{\sigma_I\}_{I\in\mathcal{D}}$, the operator T_σ in equation (9.15) is called the *martingale transform.* ◊

Theorem 9.54. *Let $\sigma = \{\sigma_I\}_{I\in\mathcal{D}}$ be a sequence of plus and minus ones, and let T_σ be its associated martingale transform. There exist constants c, $C > 0$ depending only on $1 < p < \infty$, such that for all choices σ of signs and for all $f \in L^p(\mathbb{R})$ and for $1 < p < \infty$,*

$$c\|f\|_p \le \|T_\sigma f\|_p \le C\|f\|_p. \tag{9.16}$$

The first inequality is deduced from the second one with $c = 1/C$ after noting that $T_\sigma(T_\sigma f) = f$. The second we will deduce from similar inequalities valid for the dyadic square function.

Let us illustrate Theorem 9.54 in the case $p = 2$. We know that the Haar functions provide an orthonormal basis in $L^2(\mathbb{R})$. In particular, Plancherel's Identity holds:

$$\|f\|_2^2 = \sum_{I\in\mathcal{D}} |\langle f, h_I\rangle|^2.$$

To compute the L^2 norm of the function, we add the squares of the absolute values of the Haar coefficients. Since each $|\sigma_I|^2 = 1$, we have

$$\|T_\sigma f\|_2 = \|f\|_2,$$

and so inequality (9.16) holds with $c = C = 1$. Therefore the martingale transform is an isometry in $L^2(\mathbb{R})$. In $L^p(\mathbb{R})$ we do not have an isometry, but we have the next best thing, which is encoded in the norm equivalence given by inequalities (9.16). We state another norm equivalence for yet another operator, the (nonlinear) *dyadic square function* S^d. It will imply Theorem 9.54.

Definition 9.55. The *dyadic square function* $S^d f$ is defined for functions $f \in L^2(\mathbb{R})$ by

$$S^d(f)(x) = \Big(\sum_{I\in\mathcal{D}} \frac{|\langle f, h_I\rangle|^2}{|I|}\chi_I(x)\Big)^{1/2}. \tag{9.17}$$ ◊

It turns out that in the case of the dyadic square function, we also have a norm equivalence in $L^p(\mathbb{R})$.

Theorem 9.56. *There exist positive constants* c_p *and* C_p*, depending only on* $1 < p < \infty$*, such that for all* $f \in L^p(\mathbb{R})$

$$c_p\|f\|_p \leq \|S^d(f)\|_p \leq C_p\|f\|_p. \tag{9.18}$$

This norm equivalence can be considered as an L^p substitute for Plancherel's Identity for Haar functions. Theorem 9.56 tells us that we can recover the L^p norm of f from the L^p norm of $S^d(f)$. In the definition of the dyadic square function only the absolute values of the Haar coefficients of f are used, and so that information is all that is required to decide whether f is in $L^p(\mathbb{R})$. The case $p = 2$ is left as an exercise. The case $p \neq 2$ can be found in [**Graf08**, Appendix C2] and is based on a beautiful probabilistic result called Khinchine's Inequality[14] that is worth exploring and understanding. See the project in Section 9.8.

[14] Named after Russian mathematician Aleksandr Yakovlevich Khinchine (1894–1959).

Exercise 9.57 (*The Dyadic Square Function Is an Isometry on* $L^2(\mathbb{R})$)**.** Verify that $\|S^d(f)\|_2 = \|f\|_2$. ◊

Exercise 9.58 (*The Dyadic Square Function in Terms of Difference Operators*)**.** Show that $\big[S^d(f)(x)\big]^2 = \sum_{j\in\mathbb{Z}} |Q_j f(x)|^2$, where Q_j is the difference operator defined in Definition 9.33. Lemma 9.35 will be useful. ◊

Proof of Theorem 9.54. The definition of the dyadic square function (9.17) implies that for all choices of signs σ,

$$S^d(f) = S^d(T_\sigma f).$$

Theorem 9.56 implies the second inequailty in the norm equivalence (9.16), because

$$c_p\|T_\sigma f\|_p \le \|S^d(T_\sigma f)\|_p = \|S^d(f)\|_p \le C_p\|f\|_p.$$

Dividing by $c_p > 0$, we conclude that $\|T_\sigma f\|_p \le C\|f\|_p$, as desired. The proof of the lower bound is similar. □

We can argue in a similar fashion to show that the following dyadic operator is bounded on $L^p(\mathbb{R})$ for all $1 < p < \infty$.

Definition 9.59. *Petermichl's shift operator* is defined for each function $f \in L^2(\mathbb{R})$ by

$$Шf(x) = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle \frac{1}{\sqrt{2}}\big(h_{I_r}(x) - h_{I_l}(x)\big). \qquad \Diamond$$

The operator Ш (a letter in the Cyrillic alphabet, pronounced "sha") was introduced by Stefanie Petermichl [**Pet**] in connection with the Hilbert transform (see Chapter 12).

Exercise 9.60 (*Petermichl's Shift Operator*)**.** (i) Let $\operatorname{sgn}(I,\tilde{I}) = 1$ when I is the right daughter of $\tilde{I}$, and $\operatorname{sgn}(I,\tilde{I}) = -1$ when I is the left daughter. Show that $Шf(x) = \sum_{I\in\mathcal{D}} \big(\operatorname{sgn}(I,\tilde{I})/\sqrt{2}\big)\langle f, h_{\tilde{I}}\rangle h_I(x)$.

(ii) Show that $\|Шf\|_2 = \|f\|_2$.

(iii) Show that there are constants c_p, $C_p > 0$ such that for all $f \in L^p(\mathbb{R})$, $c_p\|f\|_p \le \|Шf\|_p \le C_p\|f\|_p$. **Hint:** Calculate $SШf$, and then argue as we did for the martingale transform in the proof of Theorem 9.54. ◊

For general wavelets we also have *averaging* and *difference operators*, P_j and Q_j, and a corresponding *square function.* The same norm equivalence (9.18) holds in $L^p(\mathbb{R})$. As it turns out, wavelet bases provide unconditional bases for a whole zoo of function spaces (Sobolev spaces[15], Hölder spaces, etc.). See for example [**Dau92**, Chapter 9].

The trigonometric system is an orthonormal basis for $L^2([0,1])$. However it does not provide an unconditional basis in $L^p([0,1])$ for $p \neq 2$; see [**Woj91**, II.D.9]. There is a square function that plays the same rôle that the dyadic square function plays for the Haar basis:

$$Sf(\theta) := \Big(\sum_{j\geq 0} |\Delta_j f(\theta)|^2\Big)^{1/2}.$$

Here $\Delta_j f$ is the projection of f onto the subspace of trigonometric polynomials of degree n where $2^{j-1} \leq |n| < 2^j$ for $j \geq 1$:

$$\Delta_j f(\theta) := \sum_{2^{j-1}\leq|n|<2^j} \widehat{f}(j)\, e^{2\pi inx} \quad \text{and} \quad \Delta_0 f(\theta) = \widehat{f}(0).$$

It is true that $\|f\|_p$ is comparable to $\|S(f)\|_p$ in the sense of inequality (9.18). We are allowed to change the signs of the Fourier coefficients of f on *the dyadic blocks* of frequency. Denoting by $\mathcal{T}_\delta f$ the function reconstructed with the modified coefficients, that is,

$$\mathcal{T}_\delta f(\theta) := \sum_{j\geq 0} \delta_j \Delta_j f(\theta), \qquad \delta_j = \pm 1,$$

we have $S(\mathcal{T}_\delta f) = S(f)$, and so their L^p norms are the same and are both equivalent to $\|f\|_p$. But there is no guarantee that the same will be true if we change some but not all signs *inside* a given dyadic block! In that case, $S(f)$ does not have to coincide with $S(\mathcal{T}_\delta f)$.

The study of square functions is known as *Littlewood–Paley theory*[16]; it is a widely used tool in harmonic analysis. For an introduction to dyadic harmonic analysis, see the lecture notes by the first author [**Per01**]. The books [**Duo**], [**Graf08**], [**Ste70**], and [**Tor**] all discuss this important topic.

---

[15]Named after Russian mathematician Sergei Lvovich Sobolev (1908–1989).

[16]Named after British mathematicians John Edensor Littlewood (1885–1977) and Raymond Edward Alan Christopher Paley (1907–1933).

9.6. Project: Local cosine and sine bases

Local cosine (or sine) bases are orthonormal bases for $L^2(\mathbb{R})$, consisting of functions that are both smooth and compactly supported. In this sense they could be said to defeat the Balian–Low Theorem (Theorem 9.11), which implies that the functions in a Gabor basis for $L^2(\mathbb{R})$ cannot be both smooth and compactly supported. Local cosine bases were first discovered by Malvar [**Malv**] and are sometimes called Malvar–Wilson wavelets[17]. See [**MP**, Section 2.3], [**CMe**], [**HW**, Chapter 1], and [**JMR**, Chapter 6].

(a) Understand the construction of local cosine and sine bases, using these sources or others.

(b) Why is it useful to have a basis with the properties stated above? Find some applications in the literature.

(c) Find some software implementations of local cosine bases, and explore how they work. Or write your own implementation.

(d) Clarify the relationship of local cosine and sine bases with the Balian–Low Theorem (Theorem 9.11) and Heisenberg's Uncertainty Principle (Theorem 8.44).

9.7. Project: Devil's advocate

Investigate whether the Haar series $\sum_{I\in\mathcal{D}}\langle\chi_{[0,1)}, h_I\rangle h_I(x)$ for the function $\chi_{[0,1)}$ converges to $\chi_{[0,1)}$ pointwise (yes), uniformly (yes), in $L^1(\mathbb{R})$ (no), or in $L^2(\mathbb{R})$ (yes). See also Section 9.4.3.

(a) Show that $\langle\chi_{[0,1)}, h_I\rangle = -1/\sqrt{2^n}$ if $I = I^n := [0, 2^n)$, for $n \geq 1$. Show that for $I \in \mathcal{D}$ and for $I \neq I_n$ for every $n \geq 1$, $\langle\chi_{[0,1)}, h_I\rangle = 0$. Hence show that for each $x \in \mathbb{R}$,

$$(9.19)\quad \sum_{I\in\mathcal{D}}\langle\chi_{[0,1]}, h_I\rangle h_I(x) = \sum_{n=1}^{\infty}\frac{1}{2^n}\big[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\big].$$

[17] Named after the Brazilian engineer Henrique Malvar (born 1957) and the American physicist Kenneth Wilson (born 1936). Wilson received the 1982 Nobel Prize in physics.

(b) Show that the right-hand side of equation (9.19) is equal to $\chi_{[0,1)}(x)$ for each $x \in \mathbb{R}$. Consider separately the cases $x < 0$, $0 \leq x < 1$, and $2^k \leq x < 2^{k+1}$ for $k \geq 0$. For instance for $2^k \leq x < 2^{k+1}$,

$$\sum_{n=1}^{\infty} \frac{1}{2^n}\left[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\right] = -\frac{1}{2^k} + \sum_{n=k+1}^{\infty} \frac{1}{2^n} = 0.$$

(c) The partial Haar sums f_N of f are defined to be $f_N(x) := \sum_{n=1}^{N} \frac{1}{2^n}\left[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\right]$. Show that $f_N(x) = 1 - 2^{-N}$ if $x \in [0,1)$, $f_N(x) = -2^{-N}$ if $x \in [1, 2^N)$, and $f_N(x) = 0$ otherwise. Show that $\int_{\mathbb{R}} f_N(x)\,dx = 0$ for all $N \in \mathbb{N}$. Show that $\{f_N\}_{N\in\mathbb{N}}$ converges uniformly to $\chi_{[0,1)}(x)$ on $\mathbb{R}$. Explain why we cannot interchange the limit and the integral, despite the uniform convergence of f_N. Indeed, if we interchange them, we will conclude that

$$\lim_{N\to\infty} \int_{\mathbb{R}} f_N(x)\,dx = 0 \neq 1 = \int_{\mathbb{R}} \chi_{[0,1)}(x)\,dx = \int_{\mathbb{R}} \lim_{N\to\infty} f_N(x)\,dx.$$

Hint: The real line $\mathbb{R}$ is not compact.

(d) Show that $\{f_N\}_{N\in\mathbb{N}}$ does not converge to $\chi_{[0,1)}$ in $L^1(\mathbb{R})$. Show that $\{f_N\}_{N\in\mathbb{N}}$ converges to $\chi_{[0,1)}$ in $L^2(\mathbb{R})$.

(e) The Lebesgue Dominated Convergence Theorem (see Chapter 2 and Theorem A.59) implies that the interchange of limit and integral would be valid if there were an integrable function g such that $|f_N(x)| \leq g(x)$ for all $N \in \mathbb{N}$. Deduce constructively that there can be no such dominating function $g \in L^1(\mathbb{R})$. **Hint:** Calculate explicitly $g_1(x) = \sup_{N\geq 1} |f_N(x)|$. Any dominant function g must be larger than or equal to g_1. Verify that g_1 is not integrable.

9.8. Project: Khinchine's Inequality

This project deals with *Khinchine's Inequality*, which is the key to proving the norm equivalence in $L^p(\mathbb{R})$ between the square function Sf and the function f (Theorem 9.56). First one needs to get acquainted with the *Rademacher functions*[18] and to be comfortable with the fact that these functions are independent random variables. You will find enough to get you started in [**Graf08**, Appendix C] and [**Woj91**, Section I.B.8]. See also [**Ste70**, Appendix] and [**Zyg59**].

[18]Named after German mathematician Hans Adolph Rademacher (1892–1969).

(a) The Rademacher functions $\{r_n(t)\}_{n=1}^{\infty}$ are defined on $[0,1]$ by $r_n(t) = \operatorname{sgn}\big(\sin(2^n t\pi)\big)$. Show that this definition is equivalent to the following recursive definition: $r_1(t) = 1$ if $0 \le t \le 1/2$, $r_1(t) = -1$ if $1/2 < t \le 1$, and given $r_n(t)$, then $r_{n+1}(t) = 1$ on the left half and -1 on the right half of each interval where $r_n(t)$ is constant.

(b) Show that the Rademacher system is an orthonormal system but *not* a complete orthonormal system.

(c) Verify that $\{r_n(t)\}_{n=1}^{\infty}$ is a sequence of mutually independent random variables on $[0,1]$, each taking value 1 with probability 1/2 and value -1 with probability 1/2. The mutual independence amounts to checking that for all integrable functions f_j we have

$$\int_0^1 \prod_{j=0}^{n} f_j(r_j(t))\, dt = \prod_{j=0}^{n} \int_0^1 f_j(r_j(t))\, dt.$$

(d) Prove *Khinchine's Inequality* (or find a proof in the literature and make sure you understand it): for all square summable sequences of scalars $\{a_n\}_{n\ge1}$ and for every p with $0 < p < \infty$, we have similarity of the L^2 and L^p norms: $\big\| \sum_{n\ge1} a_n r_n \big\|_{L^p([0,1])} \sim \big\| \sum_{n\ge1} a_n r_n \big\|_{L^2([0,1])}$.

(e) Prove a similar inequality for *lacunary* sequences of trigonometric functions: If $\{n_k\}_{k\ge1}$ is a sequence of natural numbers such that $\inf_{k\ge1}(n_{k+1}/n_k) = \lambda > 1$, then for all square summable sequences of scalars $\{a_k\}_{k\ge1}$ and for every p with $0 < p < \infty$, we have $\big\| \sum_{k\ge1} a_k e^{in_k\theta} \big\|_{L^p([0,1])} \sim \big\| \sum_{k\ge1} a_k e^{in_k\theta} \big\|_{L^2([0,1])}$.

Chapter 10

Zooming properties of wavelets

The wavelet bases of Chapter 9 have a lot of internal structure. They decompose the whole function space $L^2(\mathbb{R})$ into *subspaces*, which are nested. Each signal function f is decomposed into pieces: there is a piece of f in each subspace. These pieces, or *projections*, of f give finer and finer details of f. These ideas lead to the *multiresolution analysis* (MRA) formulation of wavelet theory, examined in this chapter.

First, we come to grips with the definition of an orthogonal multiresolution analysis (MRA) with scaling function φ (Section 10.1). Second, we discuss the two-dimensional wavelets and multiresolution analysis used in image processing, as well as the use of wavelet decompositions in the compression and denoising of images and signals (Section 10.2). We give a case study: the FBI fingerprint database. Third, we present Mallat's Theorem, which shows how to construct a wavelet from an MRA with scaling function (Section 10.3). We prove it using Fourier analysis tools from previous chapters. Associated to the scaling function are some numbers, the *low-pass filter coefficients*, which in turn define the *dilation equation* that the scaling function must satisfy. These low-pass filter coefficients are all we need to find the wavelet explicitly. Finally, we show how to identify such magical low-pass filters and how to solve the dilation equation (Section 10.4).

10.1. Multiresolution analyses (MRAs)

A *multiresolution analysis* (MRA) is an abstract, sophisticated way of formulating the idea of writing functions in terms of dilates and translates of a wavelet function. Meyer[1] and Mallat[2] devised this general framework, which provides the right setting for the construction of most wavelets. In this section, we give informal and formal definitions of multiresolution analyses, together with several examples.

10.1.1. MRA: The informal definition. The space $L^2(\mathbb{R})$ is written in terms of subspaces V_j, W_j, determined by the chosen wavelet.

Approximation spaces: The subspaces V_j are called *approximation spaces*, or *scaling spaces*. They are nested and increasing:

$$\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots .$$

The space V_0 is called the *central space*. The j^{th} approximation f_j to f is in V_j. The subspaces V_j and V_{j+1} are related by *dilation*, or *scaling*, $x \mapsto 2x$, and by *translation*, $x \mapsto x - k$.

Key property: If $f(x)$ is in V_j, then all integer translates $f(x-k)$ are in V_j and all dilates $f(2x)$ and $f(2x-k)$ are in V_{j+1}.

Detail spaces: The subspaces W_j are called *detail spaces*, or *wavelet spaces*. The j^{th} level of detail of f is in W_j, and W_j is the *difference* between V_j and V_{j+1}:

$$V_j \oplus W_j = V_{j+1}.$$

Thus, for example, we can write the subspace V_3 as

$$V_3 = V_2 \oplus W_2 = V_1 \oplus W_1 \oplus W_2 = V_0 \oplus W_0 \oplus W_1 \oplus W_2.$$

Here $\oplus$ indicates a direct sum of orthogonal subspaces: $V_j \oplus W_j$ is the set of all elements $v_j + w_j$ where $v_j \in V_j$, $w_j \in W_j$, and so $v_j \perp w_j$.

Before formally defining a multiresolution analysis, we give an example that has *most* of the required features.

[1] French mathematician Yves Meyer (born 1939) was awarded the 2010 Gauss Medal, given at the International Congress of Mathematicians, ICM 2010, Hyderabad, India, for outstanding mathematical contributions that have found significant applications outside of mathematics.

[2] Stéphane Mallat is a French engineer and mathematician.

Example 10.1 (*The Fourier Series Set-up*). This example is almost a multiresolution analysis in $L^2(\mathbb{T})$, except that it does *not* have the dilation-by-two feature (Key Property), and we have not described subspaces $\dots, V_{-2}, V_{-1}$ within V_0. For $j \geq 0$, the subspace V_j is the set of all trigonometric polynomials $\sum_{|k|\leq j} a_k e^{2\pi ikt}$ of degree at most j. The piece of $f(t)$ in V_j is $S_j f(t)$, the j^{th} partial Fourier sum of f:

$$S_j f(t) = \sum_{|k|\leq j} \widehat{f}(k) e^{2\pi ikt} \in V_j.$$

Thus $S_j f(t)$ is the *orthogonal projection* of f onto V_j, since $\widehat{f}(k) = \langle f, e^{2\pi ik\cdot}\rangle$ and the exponentials $\{e^{2\pi ikt}\}_{k\in\mathbb{Z}}$ form an orthonormal basis in $L^2(\mathbb{T})$. The energy in $S_j f(t)$ is the sum $\sum_{|k|\leq j} |\widehat{f}(k)|^2$ over low frequencies $|k| \leq j$. The energy in the difference $f(t) - S_j f(t)$ is the sum over high frequencies $|k| > j$, which goes to 0 as $j \to \infty$. The subspaces V_j fill the space $L^2(\mathbb{T})$, in the sense that $S_j f \to f$ in $L^2(\mathbb{T})$ (Chapter 5).

Now, what are the *wavelet* subspaces W_j? The subspaces W_j contain the new information $\Delta_j f(t) = S_{j+1} f(t) - S_j f(t)$, where $S_{j+1} f \in V_{j+1}$ and $S_j f \in V_j$. That is, W_j contains the *detail* in f at level j. So for individual functions,

$$S_j f(t) + \Delta_j f(t) = S_{j+1} f(t),$$

and for subspaces, $V_j \oplus W_j = V_{j+1}$. The Fourier detail $\Delta_j f(t)$ is given explicitly by

$$\Delta_j f(t) = \widehat{f}(j+1) e^{2\pi i(j+1)t} + \widehat{f}(-j-1) e^{-2\pi i(j+1)t}.$$

Thus the subspace W_j contains the terms of degree *exactly* $j+1$. These terms are orthogonal to all terms of degree at most j.

The spaces W_j are *differences* between the consecutive approximation spaces V_j and V_{j+1}. They contain details at the frequency scale 2^{-j}. The spaces V_{j+1} are *sums* of the prior detail subspaces W_n, $0 \leq n \leq j$, and the central approximation subspace V_0:

$$V_0 \oplus W_0 \oplus W_1 \oplus \cdots \oplus W_j = V_{j+1}.$$

In terms of functions, we have the telescoping sum

$$S_0 f(t) + \Delta_0 f(t) + \Delta_1 f(t) + \cdots + \Delta_j f(t) = S_{j+1} f(t). \qquad \Diamond$$

Can we modify the trigonometric polynomials in Example 10.1 to satisfy the dilation requirement? Yes. Since $t \mapsto 2t$, frequencies must *double* as we pass from V_j to V_{j+1}. Let V_j contain the frequencies up to 2^j. Let V_{j+1} contain frequencies up to $2^{j+1} = 2(2^j)$, etc. Then $\Delta_j f$ contains all the frequencies between 2^j and 2^{j+1}, for $j \geq 0$:

$$\begin{aligned} V_j \ni f_j(t) &= S_{2^j} f(t) = \sum_{|k| \leq 2^j} c_k e^{2\pi i k t}, \\ W_j \ni \Delta_j f(t) &= S_{2^{j+1}} f(t) - S_{2^j} f(t) = \sum_{2^j < |k| \leq 2^{j+1}} c_k e^{2\pi i k t}. \end{aligned}$$

We obtain a one-sided multiresolution analysis that starts at V_0. The subspaces V_j and W_j have roughly the same dimension. In the field of harmonic analysis, this set-up is known as the *Littlewood–Paley decomposition* of a Fourier series; it breaks the function into octaves instead of into single terms. (See also Section 9.5.2.)

Remark 10.2. To fit the MRA requirements exactly, we would have to go to *continuous* frequency, like this:

$$f_j(x) := \int_{|\xi| \leq 2^j} \widehat{f}(\xi)\, e^{2\pi i \xi x}\, dx, \qquad j \in \mathbb{Z}.$$

Now not only does $j \to \infty$, with the V_j spaces filling $L^2(\mathbb{R})$ as $j \to \infty$, but also $j \to -\infty$, and $f_j \to 0$ in L^2 as $j \to -\infty$. In terms of subspaces, the intersection of all the V_j subspaces is the trivial subspace $\{0\}$. This set-up leads to the *Shannon MRA* (Example 10.5). ♢

10.1.2. MRA: The formal definition. An *orthogonal multiresolution analysis (MRA)* is a collection of closed subspaces of $L^2(\mathbb{R})$ that satisfy several properties: they are nested; they have trivial intersection; they exhaust the space; the subspaces communicate via a scaling property; and finally there is a special function, the *scaling function* φ, whose integer translates form an orthonormal basis for one of the subspaces. The formal definition is hard to swallow the first time one encounters it. Concentrate on the two examples briefly described: the Haar and Shannon multiresolution analyses. Once you understand the Haar example well, you will feel more comfortable with the definition of a multiresolution analysis, which we now present and discuss without further anesthesia.

Definition 10.3. An *orthogonal multiresolution analysis* (often referred to as an orthogonal MRA) with *scaling function* φ is a collection of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ such that

(1) $V_j \subset V_{j+1}$ for all $j \in \mathbb{Z}$ (*increasing subspaces*);

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots,$$

(2) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$ (*trivial intersection*);

(3) $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$ (*completeness in* $L^2(\mathbb{R})$);

(4) $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$ (*scale invariance* $\forall j \in \mathbb{Z}$);

(5) $\varphi \in V_0$, and its integer translates, $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$, form an orthonormal basis for V_0. ♢

There is also a related sequence of *wavelet subspaces* W_j of $L^2(\mathbb{R})$. It is convenient, but not necessary, to require that $V_j \perp W_j$, for all j. These subspaces are interconnected via $V_{j+1} = V_j \oplus W_j$.

Example 10.4 (*The Haar MRA*)**.** The characteristic function $\varphi(x) = \chi_{[0,1]}(x)$ is the scaling function of the (orthogonal) *Haar MRA*. The approximation subspace V_j consists of step functions with steps on the intervals $[k2^{-j}, (k+1)2^{-j})$. Properties (2) and (3) follow from equations (9.12) and (9.13). (Check the other properties.) See Section 11.1 for more detail. ♢

Example 10.5 (*The Shannon MRA*)**.** We met the Shannon wavelet in Section 9.3. The scaling function is defined on the Fourier side by

$$\widehat{\varphi}(\xi) = \chi_{[-1/2,1/2)}(\xi).$$

The approximation subspace V_j consists of the functions f whose Fourier transforms are supported on the *window* $[-2^{j-1}, 2^{j-1})$. The detail subspace W_j consists of the functions f with Fourier transforms supported on the *double-paned window* $[-2^j, -2^{j-1}) \cup [2^{j-1}, 2^j)$. ♢

Exercise 10.6. Check that the subspaces V_j of Example 10.5 do generate an MRA. Check that the subspace W_j is orthogonal to V_j and that $V_j \oplus W_j = V_{j+1}$. That is, W_j is the orthogonal complement of V_j in V_{j+1}. (**Hint:** Look on the Fourier side.) ♢

The scaling function φ lies in the central space V_0, so $\varphi \in V_0 \subset V_1$ since the spaces are nested. Also, the scaled translates $\sqrt{2}\varphi(2x-k)$ of $\varphi(2x)$ form an orthonormal basis for V_1 (Exercise 10.8). Since $\varphi \in V_1$, we can write $\varphi(x)$ in terms of this basis with coefficients $\{h_k\}_{k\in\mathbb{Z}}$, obtaining the equation

$$\varphi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \varphi(2x-k). \tag{10.1}$$

This equation is the *dilation equation*, also known as the *scaling equation.* The coefficients h_k are the *filter coefficients* of the wavelet corresponding to the MRA. In engineering the filter coefficients are called *taps.* The dilation equation is a *two-scale equation*; it expresses a relationship between a function and dilates of the same function at a different scale. (Compare with *differential equations* which relate a function and its derivatives.)

There are three natural ways to construct or describe an MRA:

(1) by the approximating subspaces V_j;

(2) by the scaling function $\varphi(x)$;

(3) by the coefficients h_k in the dilation equation.

The Haar and Shannon MRAs sit at opposite ends of the time–frequency spectrum. There are other MRAs in between.

Example 10.7 (*Daubechies Wavelets*)**.** For each integer $N \geq 1$ there is an orthogonal MRA that generates a compactly and minimally supported wavelet, such that the length of the support is $2N$ and the filters have $2N$ taps. These *Daubechies wavelets* are denoted in MATLAB by *dbN*. Figure 10.1 shows the graphs of the Daubechies scaling and wavelet functions *db*2, *db*4, and *db*6. The wavelet *db*1 is the Haar wavelet. The low-pass filter coefficients corresponding to *db*2 are

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \quad h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \quad h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, \quad h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}. \quad \Diamond$$

A scaling function φ for an MRA completely determines the nested subspaces: the central subspace V_0 is determined by property (5), and the scaling property (4) allows us to move up and down

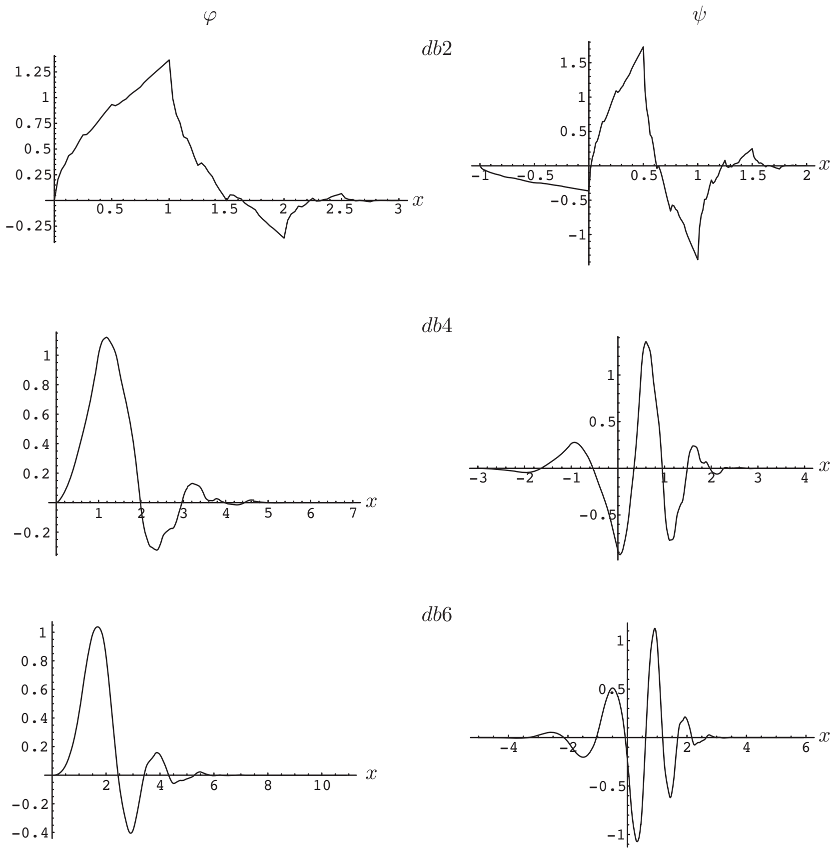


Figure 10.1. Graphs of the Daubechies scaling function φ and wavelet ψ for filter lengths 4, 8, and 12. As the filters get longer, the functions get smoother. (This figure is adapted from [**MP**, Figure 19, p. 50].)

the scale of subspaces. Once such a function φ is found, the appropriate translates and dilates of φ form an orthonormal basis for V_j. Given the scaling function φ, we denote its integer translates and dyadic dilates with integer subscripts j, k, as we did for the wavelet ψ:

$$\varphi_{j,k}(x) := 2^{j/2}\varphi(2^j x - k). \tag{10.2}$$

Exercise 10.8. Let φ be the scaling function of an orthogonal MRA. Show that the integer translates $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ of the j^{th} dilate $\varphi_{j,0}$ form an orthonormal basis for V_j. ♢

A goal is to identify scaling functions that generate orthogonal MRAs. Not all functions φ work. For instance φ must be a solution of the dilation equation for some set of filter coefficients (Section 10.4.3). However, even a solution of the dilation equation need not be the scaling function of an orthogonal MRA, as the following example shows.

Example 10.9. The hat function $\phi(x) = (1 - |x|)\chi_{\{|x|\leq 1\}}(x)$ obeys the following dilation equation (sketch the graphs!):

$$\phi(x) = \frac{1}{2}\phi(2x-1) + \phi(2x) + \frac{1}{2}\phi(2x+1).$$

However $\phi(x)$ is clearly not orthogonal to $\phi(2x)$ or to $\phi(2x \pm 1)$. ♢

Section A.2.4 is good preparation for what follows.

Definition 10.10. Given an L^2 function f, let $P_j f$ be the *orthogonal projection of f onto V_j*. Since $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of V_j, we have

$$P_j f(x) := \sum_{k\in\mathbb{Z}} \langle f, \varphi_{j,k}\rangle \varphi_{j,k}(x). \tag{10.3}$$ ♢

The function $P_j f$ is an approximation to the original function at scale 2^{-j}. More precisely, $P_j f$ is the *best approximation* to f in the subspace V_j. See Theorem 5.33, Section 9.4.4, and Theorem A.46.

Exercise 10.11. Show that $P_j(P_{j+1}f) = P_j f$ for all $f \in L^2(\mathbb{R})$. Moreover, $P_j(P_n f) = P_j f$ for all $n \geq j$. ♢

The approximation subspaces are nested, and so $P_{j+1}f$ is a better approximation to f than $P_j f$ is, or at least equally good. How do we go from the approximation $P_j f$ to the better approximation $P_{j+1}f$?

Definition 10.12. Define the *difference operator* Q_j by

$$Q_j f := P_{j+1}f - P_j f.$$ ♢

To recover $P_{j+1}f$, we add $P_j f$ to the difference $Q_j f$. We do not seem to have accomplished much, until we realize that $Q_j f$ is the orthogonal projection of f onto the orthogonal complement W_j of V_j in V_{j+1}.

Definition 10.13. Given an orthogonal MRA with approximation subspaces $\{V_j\}_{j\in\mathbb{Z}}$, denote by W_j the *orthogonal complement of* V_j *in* V_{j+1}, meaning the subspace of $L^2(\mathbb{R})$ consisting of vectors in V_{j+1} which are orthogonal to V_j:

$$W_j := \{h \in V_{j+1} : \langle h, g\rangle = 0, \text{ for all } g \in V_j\}.$$

We call W_j the *detail subspace* at scale 2^{-j}. ◊

Lemma 10.14. *The difference $Q_j f$ is the orthogonal projection of $f \in L^2(\mathbb{R})$ onto W_j.*

Lemma 10.14 is a particular instance of Lemma A.50.

By definition W_j is the orthogonal complement of the closed subspace V_j in V_{j+1}. This means that $V_j \perp W_j$, and if $f \in V_{j+1}$, there exist unique $g \in V_j$ and $h \in W_j$ such that $f = g + h$. In fact g is the orthogonal projection of f onto V_j, $g = P_j f$, and h is the orthogonal projection of f onto W_j, $h = Q_j f$. Thus for each $j \in \mathbb{Z}$,

$$V_{j+1} = V_j \oplus W_j.$$

Exercise 10.15. Show that if $j \neq k$, then $W_j \perp W_k$. Hence show that we get an orthogonal decomposition of each subspace V_j in terms of the less accurate approximation space V_n and the detail subspaces W_k at intermediate resolutions $n \leq k < j$:

$$V_j = V_n \oplus W_n \oplus W_{n-1} \oplus \cdots \oplus W_{j-2} \oplus W_{j-1}. \qquad \Diamond$$

The order of the summands does not matter and no parentheses are needed, since the direct sum is commutative and associative.

The nested subspaces $\{V_j\}_{j\in\mathbb{Z}}$ define an orthogonal MRA, and the density condition (3) holds. Therefore the detail subspaces give an orthogonal decomposition of $L^2(\mathbb{R})$:

$$L^2(\mathbb{R}) = \overline{\bigoplus_{j\in\mathbb{Z}} W_j}. \tag{10.4}$$

Here the space on the right is the closure of $\bigoplus_{j\in\mathbb{Z}} W_j$ in the L^2 norm. The equality indicates that $\bigoplus_{j\in\mathbb{Z}} W_j$ is dense in $L^2(\mathbb{R})$ with respect to the L^2 norm.

Example 10.16. The expectation and difference operators for the Haar basis (Section 9.4.4) coincide with the orthogonal projections onto the approximation and detail subspaces in the *Haar MRA*. ♢

Exercise 10.17. Show that the detail subspace W_j is a dilation of W_0; that is, show that it obeys the same scale-invariance property (4) of Definition 10.3 that the approximation subspaces V_j satisfy. ♢

We show in Section 10.3 that the scaling function φ determines a *wavelet* ψ such that $\{\psi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_0. The detail subspace W_j is a dilation of W_0; therefore the function

$$\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$$

is in W_j, and the family $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ forms an orthonormal basis for W_j. The *orthogonal projection* Q_j *onto* W_j is then given by

$$Q_j f = \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}.$$

Taking together all the scales indexed by j, the collection of functions $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis for the whole space $L^2(\mathbb{R})$. This important result is Mallat's Theorem (see [**Mall89**]). We state it here and prove it in Section 10.3.

Theorem 10.18 (Mallat's Theorem)**.** *Given an orthogonal MRA with scaling function* φ*, there is a wavelet* $\psi \in L^2(\mathbb{R})$ *such that for each* $j \in \mathbb{Z}$*, the family* $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ *is an orthonormal basis for* W_j*. Hence the family* $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ *is an orthonormal basis for* $L^2(\mathbb{R})$*.*

Example 10.19. The Haar function and the Shannon function are the wavelet ψ in Mallat's Theorem (Theorem 10.18) corresponding to the Haar and Shannon multiresolution analyses, respectively. We return to the Haar example in Section 11.1. ♢

Exercise 10.20. Given an orthogonal MRA with scaling function φ, show that if $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_0, then $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_j. Now show that the two-parameter family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis for $L^2(\mathbb{R})$. ♢

Some wavelets do not come from an MRA, but these are rare. If the wavelet has compact support, then it does come from an MRA. Ingrid Daubechies found and classified all smooth compactly supported wavelets; see [**Dau92**]. For most applications, compactly supported wavelets are desirable and sufficient. Finally, the conditions in the definition of the MRA are not independent. For full accounts of these matters and more, consult the books by Hernández and Weiss [**HW**, Chapter 2] and by Wojtaszczyk [**Woj03**, Chapter 2].

10.2. Two applications of wavelets, one case study

In this section we briefly describe two very successful applications of wavelets. First we introduce the two-dimensional wavelet transform which is of great importance in dealing with images. Second we describe in broad strokes the basics of signal compression and denoising. Finally we bring to life these generic applications with a successful real life application: the FBI fingerprint database storage, compression, denoising, and retrieval problem.

10.2.1. Two-dimensional wavelets.[3]

There is a standard procedure for constructing bases in two-dimensional space from given one-dimensional bases, namely the *tensor product*. In particular, given a wavelet basis $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ in $L^2(\mathbb{R})$, the family of tensor products

$$\psi_{j,k;i,n}(x,y) = \psi_{j,k}(x)\,\psi_{i,n}(y), \qquad j,k,i,n \in \mathbb{Z},$$

is an orthonormal basis in $L^2(\mathbb{R}^2)$. Unfortunately we have lost the multiresolution structure. Notice that we are mixing up scales in the above process, since the scaling parameters i, j need not be related.

Exercise 10.21. Show that if $\{\psi_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for a closed subspace $V \subset L^2(\mathbb{R})$, then the functions defined on $\mathbb{R}^2$ by

$$\psi_{m,n}(x,y) = \psi_m(x)\,\psi_n(y)$$

form an orthonormal basis for $V \otimes V$. Here $V \otimes V$ is the closure of the linear span of the functions $\{\psi_{m,n}\}_{m,n\in\mathbb{N}}$. ◊

[3]Most of the presentation in Subsection 10.2.1 is adapted from Section 4.3 of the first author's book [**MP**].

We use this idea at the level of the approximation spaces V_j in the orthogonal MRA. For each scale j, the family $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of V_j. Let $\mathcal{V}_j$ be the closure in $L^2(\mathbb{R}^2)$ of the linear span of the tensor products $\varphi_{j,k,n}(x,y) = \varphi_{j,k}(x)\varphi_{j,n}(y)$:

$$\mathcal{V}_j = V_j \otimes V_j := \Big\{ f(x,y) = \sum_{n,k} a_{j,n,k}\varphi_{j,k,n}(x,y) : \sum_{n,k} |a_{j,n,k}|^2 < \infty \Big\}.$$

Notice that we are no longer mixing scales. The spaces $\mathcal{V}_j$ form an MRA in $L^2(\mathbb{R}^2)$, with scaling function

$$\varphi(x,y) = \varphi(x)\varphi(y).$$

Hence the integer shifts $\{\varphi(x-k, y-n) = \varphi_{0,k,n}\}_{k,n\in\mathbb{Z}}$ form an orthonormal basis of $\mathcal{V}_0$, consecutive approximation spaces are connected via scaling by two in both variables, and the other conditions hold.

The orthogonal complement of $\mathcal{V}_j$ in $\mathcal{V}_{j+1}$ is denoted by $\mathcal{W}_j$. The rules of arithmetic are valid for direct sums and tensor products:

$$\begin{aligned} \mathcal{V}_{j+1} &= V_{j+1} \otimes V_{j+1} = (V_j \oplus W_j) \otimes (V_j \oplus W_j) \\ &= (V_j \otimes V_j) \oplus [(V_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (W_j \otimes W_j)] \\ &= \mathcal{V}_j \oplus \mathcal{W}_j. \end{aligned}$$

Thus the space $\mathcal{W}_j$ can be viewed as the direct sum of three tensor products, namely

$$\mathcal{W}_j = (W_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (V_j \otimes W_j).$$

Therefore *three wavelets* are necessary to span the detail spaces:

$$\psi^d(x,y) = \psi(x)\psi(y), \quad \psi^v(x,y) = \psi(x)\varphi(y), \quad \psi^h(x,y) = \varphi(x)\psi(y),$$

where d stands for diagonal, v for vertical, and h for horizontal. The reason for these names is that each of the subspaces somehow favors details that are oriented in those directions.

Example 10.22 (*The Two-Dimensional Haar Basis*)**.** The scaling function is the characteristic function

$$\varphi(x,y) = \chi_{[0,1]^2}(x,y) = \chi_{[0,1]}(x)\,\chi_{[0,1]}(y)$$

of the unit cube. The following pictures help us to understand the nature of the two-dimensional Haar wavelets and scaling function:

1	1
1	1

$\varphi(x, y)$

−1	1
1	−1

$\psi^d(x, y)$

1	−1
1	−1

$\psi^h(x, y)$

−1	−1
1	1

$\psi^v(x, y)$ ♢

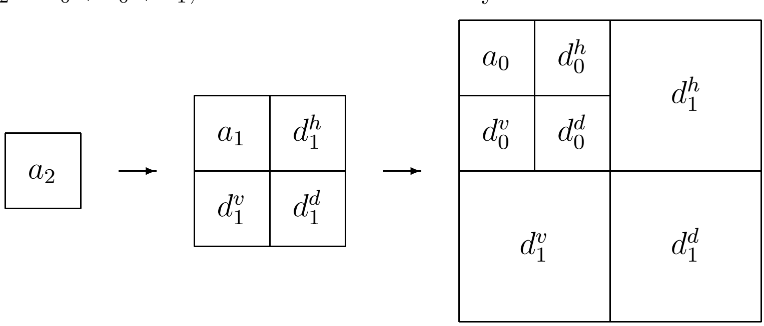
Exercise 10.23. Describe the Haar MRA in two dimensions. ♢

The same approach works in higher dimensions. There are $2^n - 1$ wavelet functions and one scaling function, where n is the dimension.

Exercise 10.24. Describe a three-dimensional orthogonal MRA. Such MRAs are useful for video compression. ♢

The construction above has the advantage that the bases are *separable.* Implementing the fast two-dimensional wavelet transform can be done easily, by successively applying the one-dimensional Fast Wavelet Transform. The disadvantage is that the analysis is very axis-dependent.

We can think of the approximation and detail images in the separable wavelet decomposition as follows. Suppose your fine resolution lives in $\mathcal{V}_2$, so your image is the approximation a_2. Decompose into a coarser approximation $a_1 \in \mathcal{V}_1$ using 1/4 of the data and the detail $d_1 = d_1^h + d_1^d + d_1^v \in \mathcal{W}_1$ which is the sum of the horizontal, diagonal, and vertical details, each carrying 1/4 of the initial data, so that $a_2 = a_1 + d_1$. Now repeat for a_1: decompose into approximation $a_0 \in \mathcal{V}_0$ and details $d_0 = d_0^h + d_0^d + d_0^v \in \mathcal{W}_0$, so that $a_1 = a_0 + d_0$ and $a_2 = a_0 + d_0 + d_1$, and so on. Schematically:



There are nonseparable MRAs, such as the twin dragon. See the project in Section 10.5.1.

10.2.2. Basics of compression and denoising. One of the main goals in signal and image processing is to be able to code the information with as few data as possible, as we saw in the example of Amanda and her mother in Chapter 1. Having fewer data points allows faster transmission and easier storage. In the presence of noise, one also wants to separate the noise from the signal (*de-noise* or *denoise* the signal), and one would like to have a basis that concentrates the signal in a few large coefficients and quarantines the noise in very small coefficients. With or without noise, the steps to follow are:

(1) Transform the data: find coefficients with respect to a given basis.

(2) Threshold the coefficients: keep the large ones and discard the small ones. Information is lost in this step, so perfect reconstruction is no longer possible.

(3) Reconstruct with the thresholded coefficients: use only the coefficients you kept in step (2). Hope that the resulting compressed signal approximates your original signal well and that in the noisy case you have successfully denoised the signal.

For instance, in the lower figure on the cover, the small localized square wave at lower right might represent a brief burst of noise, and one might discard that while keeping the large square wave at upper right.

When the basis is a wavelet basis, one simple approach is to use the projection onto an approximation space as your compressed signal, discarding all the details after a certain scale j:

$$P_j f = \sum_k \langle f, \varphi_{j,k} \rangle \varphi_{j,k}.$$

The noise is usually concentrated in the finer scales (higher frequencies!), so this approach does denoise the signal, but at the same time it removes many of the sharp features of the signal that were encoded in the finer wavelet coefficients. A more refined thresholding technique is required. The two most popular thresholding techniques are *hard thresholding* (a keep-or-toss scheme) and *soft thresholding* (the coefficients are attenuated following a linear scheme).

How to select the threshold is another issue. In the denoising case, there are some thresholding selection rules that are justified by probability theory (essentially the law of large numbers) and are used widely by statisticians. Thresholding and threshold selection principles are both encoded in MATLAB.

In traditional approximation theory there are two possible methods, *linear* approximation and *nonlinear* approximation.

Linear approximation: Select *a priori* N elements in the basis and project onto the subspace generated by those elements, regardless of the function that is being approximated. It is a linear scheme: $P_N^{\mathrm{l}} f := \sum_{n=1}^N \langle f, \psi_n\rangle \psi_n$.

Nonlinear approximation: Choose the basis elements depending on the function. For example the N basis elements could be chosen so that the coefficients are the largest in size for the particular function. This time the chosen basis elements depend on the function being approximated: $P_N^{\mathrm{nonl}} f := \sum_{n=1}^N \langle f, \psi_{n,f}\rangle \psi_{n,f}$.

The nonlinear approach has proven quite successful. For a lot more information about state-of-the-art compression and denoising techniques, see [**Mall09**, Chapters 10 and 11]. See also the project in Section 10.6.

10.2.3. Case study: The FBI Fingerprint Image Compression Standard. The *FBI Fingerprint Image Compression Standard* was developed by Tom Hopper of the FBI and Jonathan Bradley and Chris Brislawn at the Los Alamos National Laboratory. We are grateful to Chris Brislawn for permission to reproduce images and other materials from his website [**Bri02**]. That website has large, high-resolution versions of the images below and others. See also the article [**Bri95**].

When someone is fingerprinted at a police station, say in Albuquerque, New Mexico, for identification, the prints are sent to the FBI to look for matching prints. The FBI's collection of fingerprint cards was begun in 1924. By 1995, it had about 200 million cards, taking up an acre of floorspace, and about 30,000–50,000 new cards were being created per day. Each card had ten rolled prints, two unrolled prints, and two prints of all five fingers.

The FBI decided to digitize this enormous database, using 500 dots per inch with 8 bits of greyscale resolution. Then $2^8 = 256$ shades of grey are possible for each pixel, so it takes 8 bpp (bits per pixel) to store the greyscale information. A fingerprint image like the one on the left of Figure 10.2 is 768×768 pixels or 589,824 bytes. At this rate, each card corresponds to about 10 megabytes of data, so the whole database was about 2,000 terabytes ($= 2 \times 10^{12}$ bytes).

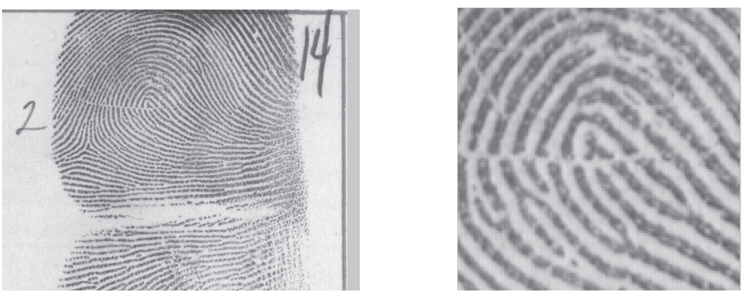


Figure 10.2. A sample fingerprint (left) and a magnified image of an area near its center (right). Reproduced from [**Bri02**], with permission.

Clearly there is a need for data compression; a factor of, say, 20:1 would help. One could try lossless methods that preserve every pixel perfectly, but in practice, lossless methods get a compression ratio of at most 2:1 for fingerprints, which is not enough. Another option for compression was the pre-1995 JPEG (Joint Photographic Experts Group) standard, using the discrete cosine transform, a version of the Discrete Fourier Transform.

The image on the right of Figure 10.2 shows a detail from the fingerprint on the left. Notice the *minutiae* such as ridge endings and bifurcations, which form permanently in childhood. From [**Bri02**]: "The graininess you see represents the individual pixels in the 500 dpi scan. The white spots in the middle of the black ridges are sweat pores, and they're admissible points of identification in court, as are the little black flesh 'islands' in the grooves between the ridges. Since these details are just a couple pixels wide, our compression method needs to preserve features right at the resolution of the scan, a VERY tough problem since most lossy algorithms work by throwing out the smallest (or highest frequency) details in the image."

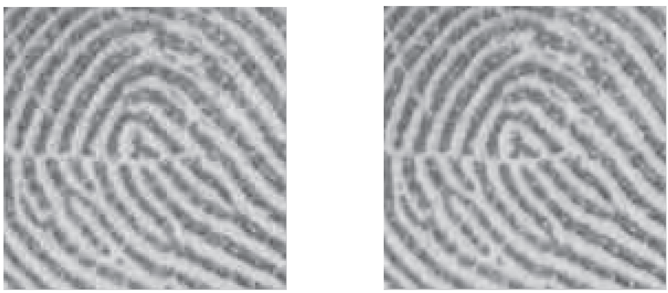


Figure 10.3. Pre-1995 JPEG compression (left) and 1995 Wavelet/Scalar Quantization compression (right) of the fingerprint image on the right of Figure 10.2, both at compression ratio 12.9:1. Reproduced from [**Bri02**], with permission.

The effect of the pre-1995 JPEG discrete cosine method on the fingerprint image above, at a compression ratio of 12.9:1 is shown on the left of Figure 10.3. The fine details have been corrupted. Also, those with good eyesight and a strong light may be able to see an artifact consisting of a regular grid of 256 small *blocks* each measuring 8 pixels by 8 pixels. This blocking artifact is much more visible in the larger image in [**Bri02**].

The Los Alamos team developed a Wavelet/Scalar Quantization (WSQ) compression method, using a specific biorthogonal wavelet with sixteen nonzero coefficients. (See Example 11.26 and the paragraphs that follow it.) The effect of this method is shown in the image on the right of Figure 10.3. The fine details are more visible than on the left, and there are no blocking artifacts. This WSQ method was adopted by the FBI in its Fingerprint Image Compression Standard. Wavelet compression methods also became part of the 2000 JPEG Standard.

On the website [**Bri02**] there is an informative visualization of the fingerprint discrete wavelet transform decomposition, showing images of sixty-four different spatial frequency subbands used to reconstruct a single fingerprint image.

10.3. From MRA to wavelets: Mallat's Theorem

In this section we show, given an orthogonal MRA with scaling function φ, how to find the corresponding wavelet whose existence is claimed in Mallat's Theorem (Theorem 10.18). The presentation is very much inspired by [**Woj03**]. The proof is not simple; however after having worked hard in previous chapters learning about Fourier series and Fourier transforms, the patient reader will be able to understand it. The impatient reader may bypass this section and move on to the next, where the more algorithmic consequences of the proof we are about to discuss will be emphasized.

The proof may be considered as an extended exercise in the use of Fourier series and transforms. We will spend a few pages setting up an appropriate time–frequency dictionary for the MRA that we are studying. First, we will give a criterion on the Fourier side that will tell us when the family of integer translates of the scaling function of the MRA forms an orthonormal family. Second, the scaling invariance paired with the nested properties of the MRA can be encoded in the important *dilation equation* which has both time and frequency incarnations. Along the way an important sequence of numbers will appear, the *filter coefficients*, which are the Fourier coefficients of a periodic function, the *low-pass filter*. Third, we will present a characterization on the Fourier side of the functions on the detail subspace W_0, in terms of the low-pass filter and the Fourier transform of the scaling function. This will be the hardest step. We will be able to write a formula on the Fourier side of a function in W_0, which we can prove is the wavelet we were looking for. Fourth, we revisit Mallat's algorithm in space and see what it says for our canonical MRA examples: Haar and Shannon.

10.3.1. Integer translates and the dilation equation. First let us find necessary and sufficient conditions on the Fourier side that guarantee that the integer translates of a square-integrable function form an orthonormal family.

Lemma 10.25. *Take $f \in L^2(\mathbb{R})$. The family $\{f_{0,k} = \tau_k f\}_{k\in\mathbb{Z}}$ of integer translates of f is orthonormal if and only if*

$$\sum_{n\in\mathbb{Z}} |\widehat{f}(\xi+n)|^2 = 1 \quad \textit{for a.e. } \xi \in \mathbb{R}.$$

Proof. First note that $\langle \tau_k f, \tau_m f\rangle = \langle \tau_{k-m} f, f\rangle$, by a change of variable (recall that $\tau_k f(x) := f(x-k)$). The orthonormality of the family of integer translates is equivalent to

$$\langle \tau_k f, f\rangle = \delta_k \quad \text{for all } k \in \mathbb{Z}.$$

Recall that the Fourier transform preserves inner products, the Fourier transform of $\tau_k f$ is the modulation of $\widehat{f}$, and the function $e^{-2\pi i k\xi}$ has period one. Therefore for all $k \in \mathbb{Z}$

$$\delta_k = \langle \widehat{\tau_k f}, \widehat{f}\rangle = \int_{\mathbb{R}} e^{-2\pi i k\xi}\, |\widehat{f}(\xi)|^2\, d\xi,$$

and using the additivity of the integral,

$$\delta_k = \sum_{n\in\mathbb{Z}} \int_n^{n+1} e^{-2\pi i k\xi} |\widehat{f}(\xi)|^2\, d\xi = \int_0^1 e^{-2\pi i k\eta} \sum_{n\in\mathbb{Z}} |\widehat{f}(\eta+n)|^2\, d\eta.$$

The last equality is obtained by performing on each integral the change of variable $\eta = \xi - n$ that maps the interval $[n, n+1)$ onto the unit interval $[0,1)$. This identity says that the periodic function of period one given by $F(\eta) = \sum_{n\in\mathbb{Z}} |\widehat{f}(\eta+n)|^2$ has k^{th} Fourier coefficient equal to the Kronecker delta δ_k. Therefore it must be equal to one almost everywhere, which is exactly what we set out to prove. □

We can now establish two important identities that hold for the scaling function φ of an orthogonal MRA.

Theorem 10.26. *Let φ be the scaling function of an orthogonal MRA. For almost every $\xi \in \mathbb{R}$,*

$$\sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi+n)|^2 = 1. \tag{10.5}$$

There exist coefficients $\{h_k\}_{k\in\mathbb{Z}}$ *such that* $\sum_{k\in\mathbb{Z}} |h_k|^2 < \infty$*:*

$$\varphi(x) = \sqrt{2} \sum_{k\in\mathbb{Z}} h_k \, \varphi(2x-k). \tag{10.6}$$

Proof. The integer translates of φ form an orthonormal family. By Lemma 10.25, for almost every $\xi \in \mathbb{R}$ equality (10.5) holds.

Also, $\varphi \in V_0 \subset V_1$, and the functions $\varphi_{1,k}(x) = \sqrt{2}\varphi(2x-k)$, for $k \in \mathbb{Z}$, form an orthonormal basis for V_1. Therefore φ must be a (possibly infinite) linear combination of the basis functions:

$$\varphi(x) = \sum_{k\in\mathbb{Z}} \langle \varphi, \varphi_{1,k} \rangle \, \varphi_{1,k}(x) = \sqrt{2} \sum_{k\in\mathbb{Z}} h_k \, \varphi(2x-k)$$

for $h_k := \langle \varphi, \varphi_{1,k} \rangle$. Furthermore, $\sum_{k\in\mathbb{Z}} |h_k|^2 = \|\varphi\|^2_{L^2(\mathbb{R})} < \infty$. □

Definition 10.27. Equation (10.6) is called a *dilation equation* or *scaling equation*. The *dilation equation on the Fourier side* reads

$$\widehat{\varphi}\,(\xi) = H(\xi/2)\, \widehat{\varphi}\,(\xi/2), \tag{10.7}$$

where $H(\xi) = (1/\sqrt{2}) \sum_{k\in\mathbb{Z}} h_k e^{-2\pi i k \xi}$ is an L^2 function of period 1, called the *low-pass filter*[4]. The coefficients $\{h_k\}_{k\in\mathbb{Z}}$ are called *filter coefficients.* ◊

For convenience we assume that the low-pass filter H is a trigonometric polynomial, so that all but finitely many of the coefficients $\{h_k\}_{k\in\mathbb{Z}}$ vanish. There are multiresolution analyses whose low-pass filters are trigonometric polynomials, for example the Haar MRA; see (11.2). In fact, for applications, these are the most useful, and they correspond to MRAs with compactly supported scaling functions; see [**Dau92**], [**Woj03**].

Exercise 10.28. Check that the scaling equation on the Fourier side is given by equation (10.7). ◊

We can now deduce a necessary property that the low-pass filter H must satisfy. This condition is known in the engineering community as the *quadrature mirror filter* (QMF) property, necessary to achieve exact reconstruction for a pair of filters.

[4]Some authors prefer the name *refinement mask*. The term low-pass filter will become clear in Section 10.3.4.

Lemma 10.29. *Given an orthogonal MRA with scaling function φ and a corresponding low-pass filter H that we assume is a trigonometric polynomial, then for every $\xi \in \mathbb{R}$,*

$$|H(\xi)|^2 + |H(\xi + 1/2)|^2 = 1.$$

Proof. Insert equation (10.7) into equation (10.5), obtaining

$$1 = \sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi + n)|^2 = \sum_{n\in\mathbb{Z}} \big|H\big((\xi+n)/2\big)\big|^2 \big|\widehat{\varphi}\left((\xi+n)/2\right)\big|^2.$$

Now separate the sum over the odd and even integers, use the fact that H has period one to factor it out from the sum, and use equation (10.5) (twice), which holds for almost every point ξ:

$$|H(\xi/2)|^2 \sum_{k\in\mathbb{Z}} |\widehat{\varphi}\,(\xi/2 + k)|^2 + |H(\xi/2 + 1/2)|^2 \sum_{k\in\mathbb{Z}} \big|\widehat{\varphi}\left((\xi+1)/2 + k\right)\big|^2$$
$$= |H(\xi/2)|^2 + |H(\xi/2 + 1/2)|^2 = 1.$$

Equality holds almost everywhere. Since H is a trigonometric polynomial, H is continuous and so equality must hold everywhere. □

10.3.2. Characterizing functions in the detail subspace W_0. We describe the functions in W_0, the orthogonal complement of V_0 in V_1, where all the subspaces correspond to an orthogonal MRA with scaling function φ and low-pass filter a trigonometric polynomial H.

For any $f \in V_1$, the same argument we used for φ shows that there must be a function of period one, $m_f(\xi) \in L^2([0,1))$, such that

$$\widehat{f}(\xi) = m_f(\xi/2)\,\widehat{\varphi}\,(\xi/2). \tag{10.8}$$

In this notation, the low-pass filter $H = m_\varphi$.

Lemma 10.30. *A function $f \in W_0$ if and only if there is a function $v(\xi)$ of period one such that*

$$\widehat{f}(\xi) = e^{\pi i \xi}\, v(\xi)\, \overline{H(\xi/2 + 1/2)}\, \widehat{\varphi}(\xi/2).$$

Proof. First, $f \in W_0$ if and only if $f \in V_1$ and $f \perp V_0$. The fact that $f \in V_1$ allows us, by observation (10.8), to reduce the problem to showing that

$$m_f(\xi) = e^{2\pi i \xi}\, \sigma(\xi)\, \overline{H(\xi + 1/2)},$$

where $\sigma(\xi)$ is some function with period 1/2. Then $v(\xi) = \sigma(\xi/2)$ will have period one.

The orthogonality $f \perp V_0$ is equivalent to $\langle f, \varphi_{0,k}\rangle = 0$ for all $k \in \mathbb{Z}$. A calculation similar to the one in the proof of Lemma 10.25 shows that

$$0 = \langle \widehat{f}, \widehat{\varphi_{0,k}}\rangle = \int_{\mathbb{R}} \widehat{f}(\xi)\, e^{2\pi ik\xi}\, \overline{\widehat{\varphi}(\xi)}\, d\xi.$$

Use equations (10.8) and (10.7) to conclude that

$$\begin{aligned} 0 &= \int_{\mathbb{R}} e^{2\pi ik\xi}\, m_f(\xi/2)\, \widehat{\varphi}\,(\xi/2)\, \overline{H(\xi/2)}\, \overline{\widehat{\varphi}\,(\xi/2)}\, d\xi \\ &= \int_{\mathbb{R}} e^{2\pi ik\xi}\, m_f(\xi/2)\, \overline{H(\xi/2)}\, |\widehat{\varphi}\,(\xi/2)|^2\, d\xi. \end{aligned}$$

At this point we use the same trick we used in Lemma 10.25: break the integral over $\mathbb{R}$ into the sum of integrals over the intervals $[n, n+1)$, change variables to the unit interval, and use the periodicity of the exponential to get

$$\begin{aligned} 0 &= \sum_{n\in\mathbb{Z}} \int_n^{n+1} e^{2\pi ik\xi}\, m_f(\xi/2)\, \overline{H(\xi/2)}\, |\widehat{\varphi}\,(\xi/2)|^2\, d\xi \\ &= \int_0^1 e^{2\pi ik\xi} \sum_{n\in\mathbb{Z}} m_f\big((\xi+n)/2\big)\, \overline{H\big((\xi+n)/2\big)}\, \big|\widehat{\varphi}\,\big((\xi+n)/2\big)\big|^2\, d\xi. \end{aligned}$$

We would also like to take advantage of the periodicity of the functions m_f and H. These are functions of period one, but we are adding half an integer. If we separate the last sum, as we did in Lemma 10.29, into the sums over the odd and even integers, we will win. In fact we obtain that for all $k \in \mathbb{Z}$

$$\begin{aligned} 0 &= \int_0^1 e^{2\pi ik\xi} \big[m_f(\xi/2)\, \overline{H(\xi/2)} \sum_{m\in\mathbb{Z}} |\widehat{\varphi}\,(\xi/2+m)|^2 \\ &\quad + m_f(\xi/2+1/2)\, \overline{H(\xi/2+1/2)} \sum_{m\in\mathbb{Z}} \big|\widehat{\varphi}\,\big((\xi+1)/2+m\big)\big|^2\big]\, d\xi \\ &= \int_0^1 e^{2\pi ik\xi} \Big[m_f(\xi/2)\, \overline{H(\xi/2)} + m_f(\xi/2+1/2)\, \overline{H(\xi/2+1/2)}\Big]\, d\xi, \end{aligned}$$

where the last equality is a consequence of applying equation (10.5) twice. This time the 1-periodic function (see Exercise 10.31)

$$F(\xi) = m_f(\xi/2)\,\overline{H(\xi/2)} + m_f(\xi/2+1/2)\,\overline{H(\xi/2+1/2)}$$

has been shown to have Fourier coefficients identically equal to zero. Hence it must be the zero function almost everywhere:

$$m_f(\xi)\,\overline{H(\xi)} + m_f(\xi+1/2)\,\overline{H(\xi+1/2)} = 0 \quad \text{a.e.} \tag{10.9}$$

This equation says that for a point ξ for which it holds, the vector $\vec{v} \in \mathbb{C}^2$ given by

$$\vec{v} = \big(m_f(\xi), m_f(\xi+1/2)\big)$$

must be orthogonal to the vector in $\vec{w} \in \mathbb{C}^2$ given by

$$\vec{w} = \big(H(\xi), H(\xi+1/2)\big).$$

The QMF property of H (see Lemma 10.29) ensures that $\vec{w}$ is not zero; in fact $|\vec{w}| = 1$. The vector space $\mathbb{C}^2$ over the complex numbers $\mathbb{C}$ is two-dimensional. Therefore the orthogonal complement of the one-dimensional subspace generated by the vector $\vec{w}$ is one-dimensional. It suffices to locate one nonzero vector $\vec{u} \in \mathbb{C}^2$ that is orthogonal to $\vec{w}$ to completely characterize all the vectors $\vec{v}$ that are orthogonal to $\vec{w}$. In fact $\vec{v} = \lambda\vec{u}$ for $\lambda \in \mathbb{C}$. The vector

$$\vec{u} = \big(-\overline{H(\xi+1/2)}\,,\overline{H(\xi)}\,\big)$$

is orthogonal to $\vec{w}$. Therefore $\vec{v} = \lambda\big(-\overline{H(\xi+1/2)}\,,\overline{H(\xi)}\,\big)$, and we conclude that the 1-periodic function m_f must satisfy

$$m_f(\xi) = -\lambda(\xi)\,\overline{H(\xi+1/2)}, \quad m_f(\xi+1/2) = \lambda(\xi)\,\overline{H(\xi)}.$$

Therefore $\lambda(\xi)$ is a function of period one such that $-\lambda(\xi+1/2) = \lambda(\xi)$. Equivalently, $\lambda(\xi) = e^{2\pi i\xi}\sigma(\xi)$, where $\sigma(\xi)$ has period $1/2$ (see Exercise 10.32). The conclusion is that $f \in W_0$ if and only if

$$m_f(\xi) = e^{2\pi i\xi}\,\sigma(\xi)\,\overline{H(\xi+1/2)},$$

where $\sigma(\xi)$ is a function with period $1/2$, as required. □

Exercise 10.31. Show that if H and G are periodic functions of period one, then the new function

$$F(\xi) = G(\xi/2)\,\overline{H(\xi/2)} + G(\xi/2+1/2)\,\overline{H(\xi/2+1/2)}$$

also has period one. ♢

Exercise 10.32. Show that $\lambda(\xi)$ is a function of period one such that $\lambda(\xi + 1/2) = -\lambda(\xi)$ if and only if $\lambda(\xi) = e^{2\pi i\xi}\sigma(\xi)$ where $\sigma(\xi)$ has period 1/2. $\Diamond$

10.3.3. Voilà, the wavelet. We are now ready to present the wavelet ψ associated to the MRA with scaling function φ and low-pass filter H, whose existence is claimed in Mallat's Theorem (Theorem 10.18), when H is a trigonometric polynomial.

Proof of Mallat's Theorem. The function ψ we are looking for is in the detail subspace W_0. Therefore, by Lemma 10.30, on the Fourier side it must satisfy the equation

$$\widehat{\psi}(\xi) = m_{\psi}(\xi/2)\,\widehat{\varphi}(\xi/2),$$

where

$$m_{\psi}(\xi) = e^{2\pi i\xi}\,\sigma(\xi)\,\overline{H(\xi + 1/2)} \tag{10.10}$$

and $\sigma(\xi)$ is a function with period 1/2. Furthermore, because we want the integer translates of ψ to form an orthonormal system, we can apply Lemma 10.25 to ψ and deduce a QMF property for the 1-periodic function $m_{\psi}(\xi)$. Namely, for almost every $\xi \in \mathbb{R}$,

$$|m_{\psi}(\xi)|^2 + |m_{\psi}(\xi + 1/2)|^2 = 1.$$

Substituting equation (10.10) into this equation implies that for almost every $\xi \in \mathbb{R}$,

$$|\sigma(\xi)|^2|H(\xi + 1/2)|^2 + |\sigma(\xi + 1/2)|^2|H(\xi)|^2 = 1.$$

But $\sigma(\xi)$ has period 1/2, and H satisfies the QMF condition. We conclude that $|\sigma(\xi)| = 1$ almost everywhere.

Choose a function of period 1/2 that has absolute value 1, for example $\sigma(\xi) \equiv 1$. Define the wavelet on the Fourier side to be

$$\widehat{\psi}(\xi) := G(\xi/2)\,\widehat{\varphi}(\xi/2),$$

where G is the 1-periodic function given by

$$G(\xi) := e^{2\pi i\xi}\,\overline{H(\xi + 1/2)}.$$

By Lemma 10.30, $\psi \in W_0$. Likewise, its integer translates are in W_0, because on the Fourier side

$$\widehat{\psi_{0,k}}(\xi) = e^{-2\pi ik\xi}\,G(\xi/2)\,\widehat{\varphi}(\xi/2),$$

and we can now use Lemma 10.30 again with $v(\xi) = e^{-2\pi ik\xi}$ a function of period one. Furthermore G satisfies a QMF property, and so the family of integer translates of ψ is an orthonormal family in W_0.

It remains to show that this family spans W_0. By Lemma 10.30, if $f \in W_0$, there is a square-integrable function $v(\xi)$ of period one such that

$$\widehat{f}(\xi) = v(\xi)\, e^{\pi i\xi}\, \overline{H(\xi/2 + 1/2)}\, \widehat{\varphi}(\xi/2) = v(\xi)\, \widehat{\psi}(\xi).$$

But $v(\xi) = \sum_{k\in\mathbb{Z}} a_k e^{-2\pi ik\xi}$, where $\sum_{k\in\mathbb{Z}} |a_k|^2 < \infty$. Inserting the trigonometric expansion of v and using the fact that modulations on the Fourier side come from translations, we obtain

$$\widehat{f}(\xi) = \sum_{k\in\mathbb{Z}} a_k e^{-2\pi ik\xi}\, \widehat{\psi}(\xi) = \sum_{k\in\mathbb{Z}} a_k \widehat{\psi_{0,k}}(\xi).$$

Taking the inverse Fourier transform, we see that

$$f(x) = \sum_{k\in\mathbb{Z}} a_k \psi_{0,k}(x).$$

That is, f belongs to the span of the integer translates of ψ. The integer translates of ψ form an orthonormal basis of W_0. By scale invariance, the functions $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis of W_j. Thus the family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis of $L^2(\mathbb{R})$, as required. $\square$

Exercise 10.33. Verify that if $\tau_k\psi$ is an orthonormal system and $\psi \in W_0$, then m_ψ satisfies the QMF property. $\diamondsuit$

10.3.4. Mallat's algorithm and examples revisited. Starting with a multiresolution analysis with scaling function ϕ, the wavelet ψ we found is an element of $W_0 \subset V_1$. Therefore it is also a superposition of the basis elements $\{\varphi_{1,k}\}_{k\in\mathbb{Z}}$ of V_1. So there are unique coefficients $\{g_k\}_{k\in\mathbb{Z}}$ such that $\sum_{k\in\mathbb{Z}} |g_k|^2 < \infty$ and

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k \varphi_{1,k}(x). \tag{10.11}$$

Definition 10.34. Given an *MRA* with associated low-pass filter coefficients $\{h_k\}_{k\in\mathbb{Z}}$, define the *high-pass filter* G, a 1-periodic function with filter coefficients $g = \{g_k\}_{k\in\mathbb{Z}}$, by

$$g_k = (-1)^{k-1}\overline{h_{1-k}}, \qquad G(\xi) = \frac{1}{\sqrt{2}}\sum_{k\in\mathbb{Z}} g_k e^{2\pi i k\xi}. \qquad \Diamond$$

With this choice of filter, the function ψ given by equation (10.11) is Mallat's wavelet, which we constructed in Section 10.3.3. It suffices to observe that with this choice, $\widehat{\psi}(\xi) = G(\xi/2)\,\widehat{\varphi}(\xi/2)$ with $G(\xi) := (1/\sqrt{2})\sum g_k e^{-2\pi i k\xi} = e^{2\pi i\xi}\,\overline{H(\xi+1/2)}$. See Exercise 10.36.

Lemma 10.35. *The high-pass filter G is itself a quadrature mirror filter (QMF):*

$$|G(\xi)|^2 + |G(\xi+1/2)|^2 = 1. \tag{10.12}$$

Furthermore,

$$H(\xi)\,\overline{G(\xi)} + H(\xi+1/2)\,\overline{G(\xi+1/2)} = 0. \tag{10.13}$$

We leave the proof of the lemma as an exercise for the reader.

The QMF condition for the low-pass filter H in Lemma 10.29 reflects the fact that for each scale j, the scaling functions $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis for V_j. Likewise, the QMF condition (10.12) for the high-pass filter G reflects the fact that for each scale j, the wavelets $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis for W_j. Furthermore, the orthogonality between W_j and V_j is reflected in condition (10.13).

The QMF condition for the 1-periodic function H together with $H(0) = 1$ (see Lemma 10.40) implies that $H(\pm 1/2) = 0$, which explains the name *low-pass filter*: low frequencies near $\xi = 0$ are kept, while high frequencies near $\xi = \pm 1/2$ are removed (filtered out). Condition (10.13), together with the knowledge that $H(0) = 1$ and $H(\pm 1/2) = 0$, implies that $G(0) = 0$ (equivalently $\sum_{k=0}^{L-1} g_k = 0$). The condition $G(0) = 0$ together with the QMF condition (10.13) for G implies that $G(\pm 1/2) = 1$. This observation explains the name *high-pass filter*: high frequencies near $\xi = \pm 1/2$ are kept, while low frequencies near $\xi = 0$ are removed.

Given a low-pass filter H associated to an MRA, we will always construct the associated high-pass filter G according to Definition 10.34.

Exercise 10.36. Consider a filter H that satisfies a QMF condition. Define a new filter $G(\xi) := e^{-2\pi i\xi}\,\overline{H(\xi+1/2)}$. Show that G also satisfies a QMF condition. Check that equation (10.13) holds. Verify that on the Fourier side $\widehat{\psi}(\xi) = G(\xi/2)\,\widehat{\varphi}(\xi/2)$ and that for each $m \in \mathbb{Z}$, $\psi_{0,m}$ is orthogonal to $\varphi_{0,k}$ for all $k \in \mathbb{Z}$. That is, $V_0 \perp W_0$. A similar calculation shows that $V_j \perp W_j$. ♢

We will see in Section 11.2 that Mallat's Theorem also provides a numerical algorithm that can be implemented very successfully, namely the *cascade algorithm*, or *Fast Wavelet Transform*.

Let us revisit the two examples of MRAs that we know.

Example 10.37 (*Haar Revisited*)**.** The characteristic function of the unit interval $\chi_{[0,1]}$ generates an orthogonal MRA, namely the Haar MRA. The nonzero low-pass filter coefficients are $h_0 = h_1 = 1/\sqrt{2}$; hence the nonzero high-pass coefficients are $g_0 = -1/\sqrt{2}$ and $g_1 = 1/\sqrt{2}$. Therefore the Haar wavelet is $\psi(t) = \varphi(2t-1) - \varphi(2t)$. The low-pass and high-pass filters are

$$H(\xi) = (1 + e^{-2\pi i\xi})/2, \qquad G(\xi) = (e^{-2\pi i\xi} - 1)/2. \qquad ♢$$

Example 10.38 (*Shannon Revisited*)**.** The Shannon scaling function is given on the Fourier side by $\widehat{\varphi}(\xi) = \chi_{[-1/2,1/2)}(\xi)$. It generates an orthogonal MRA. It follows from Exercise 10.28 that

$$H(\xi) = \chi_{[-1/2,-1/4)\cup[1/4,1/2)}(\xi).$$

Hence by Exercise 10.36,

$$G(\xi) = e^{2\pi i\xi} H(\xi + 1/2) = e^{2\pi i\xi}\chi_{[-1/4,1/4)}(\xi)$$

(recall that we are viewing $H(\xi)$ and $G(\xi)$ as periodic functions on the unit interval), and

$$\widehat{\psi}(\xi) = e^{\pi i\xi}\chi_{\{1/2<|\xi|\leq 1\}}(\xi). \qquad ♢$$

Exercise 10.39. Verify that the Haar and Shannon low-pass and high-pass filters satisfy corresponding QMF conditions and condition (10.13). ♢

Are there MRAs defined by other scaling functions φ? Yes; we have for an example the Daubechies family of wavelets. In the next section we give some concrete guidelines on how to hunt for suitable scaling functions and low-pass filters. In practice, large families of wavelets are known and are encoded in wavelet software, ready for use.

10.4. How to find suitable MRAs

Mallat's Theorem provides a mathematical algorithm for constructing the wavelet from the given MRA and the scaling function via the filter coefficients. We represent *Mallat's algorithm* schematically as

φ		H		G		ψ
scaling	$\to$	low-pass	$\to$	high-pass	$\to$	Mallat's
MRA		filter		filter		wavelet

In this section we are concerned with the problem of deciding whether a given trigonometric polynomial H is the low-pass filter of an orthogonal MRA or whether a given function φ is the scaling function of an orthogonal MRA. First, we highlight necessary conditions on a periodic function H for it to be the low-pass filter for an orthogonal MRA. These conditions are relatively easy to check. Second, we identify some easy-to-check properties that a candidate for the scaling function of an orthogonal MRA should have. Third, if a function φ is the scaling function of an MRA, then it will be a solution of the dilation equation for a given set of filter coefficients. We make a short digression over the existence and uniqueness of a solution (possibly a distribution) for a given scaling equation provided the coefficients add up to $\sqrt{2}$.

10.4.1. Easy-to-check conditions on the low-pass filter. Given an orthogonal MRA, the scaling function φ satisfies the *scaling equation* (10.6), for some set of filter coefficients $h = \{h_k\}_{k\in\mathbb{Z}}$:

$$\varphi(t) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k\varphi(2t-k).$$

The *low-pass filter* is given by $H(\xi) = (1/\sqrt{2})\sum_{k\in\mathbb{Z}} h_k e^{-2\pi ik\xi}$, a 1-periodic function of the frequency variable ξ, whose Fourier coefficients are $\widehat{H}(k) = h_{-k}/\sqrt{2}$. We assume that the low-pass filter has finite length L, meaning that the coefficients h_k are zero unless $k = 0$, $1, \ldots, L-1$. Thus H is a trigonometric polynomial:

$$H(\xi) = (1/\sqrt{2}) \sum_{k=0}^{L-1} h_k e^{-2\pi i k\xi}.$$

Such filters are called by the engineers *Finite Impulse Response* (FIR) filters. For applications it is a most desirable property, since FIR filters have compactly supported scaling functions and wavelets. See also Aside 10.50 and the project in Subsection 10.5.2.

We observed in (10.7) that on the Fourier side the scaling equation becomes

$$\widehat{\varphi}(\xi) = H(\xi/2)\,\widehat{\varphi}(\xi/2) \quad \text{for all } \xi \in \mathbb{R}.$$

Now plug in $\xi/2$ for ξ:

$$\widehat{\varphi}(\xi/2) = H(\xi/4)\,\widehat{\varphi}(\xi/4).$$

Repeating J times, we get

$$\widehat{\varphi}(\xi) = \Big(\prod_{j=0}^{J} H(\xi/2^j)\Big)\,\widehat{\varphi}(\xi/2^J). \tag{10.14}$$

If $\widehat{\varphi}$ is continuous[5] at $\xi = 0$, $\widehat{\varphi}(0) \neq 0$, then $\widehat{\varphi}(\xi/2^J) \to \widehat{\varphi}(0)$. In this case, we can let $J \to \infty$ in (10.14) and the infinite product converges to $\widehat{\varphi}(\xi)/\widehat{\varphi}(0)$:

$$\widehat{\varphi}(\xi)/\widehat{\varphi}(0) = \prod_{j=0}^{\infty} H(\xi/2^j). \tag{10.15}$$

We sum up in the following lemma.

Lemma 10.40. *If H is the low-pass filter of an MRA with scaling function φ such that $\widehat{\varphi}$ is continuous at $\xi = 0$ and $\widehat{\varphi}(0) = 1$, then:*

(i) $H(0) = 1$, *or equivalently* $\sum_{k=0}^{L-1} h_k = \sqrt{2}$.

(ii) *The following quadrature mirror filter (QMF) condition holds:*

$$|H(\xi)|^2 + |H(\xi + 1/2)|^2 = 1.$$

(iii) *The infinite product* $\prod_{j=0}^{\infty} H\left(\xi/2^j\right)$ *converges for each* ξ *to* $\varphi(\xi)$.

[5] This is a reasonable assumption; if $\varphi \in L^1(\mathbb{R})$, then its Fourier transform is continuous *everywhere* and decays to zero at $\pm\infty$ (Riemann–Lebesgue Lemma).

Proof. We observed in (10.7) that on the Fourier side the scaling equation becomes $\widehat{\varphi}(\xi) = H(\xi/2)\,\widehat{\varphi}(\xi/2)$. Setting $\xi = 0$, we conclude that $H(0) = 1$, or equivalently $\sum_{k=0}^{L-1} h_k = \sqrt{2}$.

In Lemma 10.29 we showed that the orthonormality of the integer shifts of the scaling function implies the QMF condition for H. The convergence of the infinite product was justified in the preceding paragraphs. □

Exercise 10.41. Check that the Daubechies *db*2 filter defined in Example 10.7 is a QMF and that $h_0 + h_1 + h_2 + h_3 = \sqrt{2}$. ♢

Exercise 10.42. Let $H(\xi) = (1 + e^{-2\pi i\xi})/2$ be the low-pass filter. (See Example 10.37 for the Haar scaling function $\varphi(x) = \chi_{[0,1)}(x)$.) Compute the infinite product $\prod_{j=1}^{\infty} H(\xi/2^j)$ directly for this example, and compare it with $\widehat{\varphi}(\xi)$. ♢

10.4.2. Easy-to-check conditions on scaling functions φ. We list here two useful properties of scaling functions of orthogonal MRAs, assuming that the scaling function φ is not only in $L^2(\mathbb{R})$ but also in $L^1(\mathbb{R})$. This latter condition implies, by the Riemann–Lebesgue Lemma, that $\widehat{\varphi}$ is continuous. In particular $\widehat{\varphi}$ is continuous at $\xi = 0$, which is all we will use in the proof of part (i) of Lemma 10.43. The orthonormality of the set $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ and the continuity of $\widehat{\varphi}$ at $\xi = 0$ imply that $|\widehat{\varphi}(0)| = 1$, and one usually normalizes to $\widehat{\varphi}(0) = \int \varphi(t)\,dt = 1$. This normalization happens to be useful in numerical implementations of the wavelet transform. These are easy-to-check conditions on potential scaling functions φ that we now summarize.

Lemma 10.43. *Consider an orthogonal MRA with scaling function φ. Suppose $\varphi \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$. Then*

(i) $|\widehat{\varphi}(0)| = 1$, *and* $|\widehat{\varphi}(k)| = 0$ *for all* $k \in \mathbb{Z}$ *with* $k \neq 0$, *and*

(ii) $\sum_{k\in\mathbb{Z}} \varphi(x + k) = 1$, *for a.e.* $x \in \mathbb{R}$.

Proof. Let $g \in L^2(\mathbb{R})$ be a nonzero function whose Fourier transform is supported on the interval $[-1/2, 1/2]$. Let us calculate the L^2 norm of the orthogonal projection onto V_j of this function g, namely $P_j g$. Since φ is a scaling function for an MRA, by the density property (3) in Definition 10.3 we have $\lim_{j\to\infty} \|P_j g\|_{L^2(\mathbb{R})} = \|g\|_{L^2(\mathbb{R})}$.

The functions $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis of V_j, and the Fourier transform preserves inner products. Therefore

$$\|P_j g\|_{L^2(\mathbb{R})} = \sum_{k\in\mathbb{Z}} |\langle g, \varphi_{j,k}\rangle|^2 = \sum_{k\in\mathbb{Z}} |\langle \widehat{g}, \widehat{\varphi_{j,k}}\rangle|^2.$$

Using the time–frequency dictionary, one can verify that

$$\widehat{\varphi_{j,k}}(\xi) = 2^{-j/2} 2^{-2\pi i k 2^{-j}\xi} \widehat{\varphi}(2^{-j}\xi). \tag{10.16}$$

Substituting this expression into each inner product in the sum and using the hypothesis on the support of $\widehat{g}$, we find that for $j \geq 0$,

$$\langle \widehat{g}, \widehat{\varphi_{j,k}}\rangle = \int_{-2^{j-1}}^{2^{j-1}} \widehat{g}\,(\xi)\, 2^{-j/2} 2^{2\pi i k 2^{-j}\xi} \overline{\widehat{\varphi}(2^{-j}\xi)}\, d\xi.$$

The functions $\{E_k^j(\xi) := 2^{-j/2} 2^{-2\pi i k 2^{-j}\xi}\}_{k\in\mathbb{Z}}$ form the trigonometric basis corresponding to the interval $[-2^{j-1}, 2^{j-1})$ of length 2^j. The integral on the left-hand side is therefore the inner product of the function $\mathcal{G}(\xi) := \widehat{g}\,(\xi)\, \overline{\widehat{\varphi}(2^{-j}\xi)}$ with the elements of this orthonormal basis. In all, we conclude that

$$\|P_j g\|^2_{L^2(\mathbb{R})} = \sum_{k\in\mathbb{Z}} |\langle \mathcal{G}, E_k^j\rangle|^2 = \|\mathcal{G}\|^2_{L^2([-2^{j-1}, 2^{j-1}))}.$$

Replacing $\mathcal{G}$ by its expression in terms of $\widehat{g}$ and $\widehat{\varphi}$ and remembering that the support of $\widehat{g}$ is $[-1/2, 1/2]$, we obtain

$$\|P_j g\|^2_{L^2(\mathbb{R})} = \int_{-2^{j-1}}^{2^{j-1}} |\widehat{g}\,(\xi)\, \overline{\widehat{\varphi}(2^{-j}\xi)}|^2\, d\xi = \int_{-1/2}^{1/2} |\widehat{g}\,(\xi)|^2 |\widehat{\varphi}(2^{-j}\xi)|^2\, d\xi.$$

Now take the limit as $j \to \infty$. The left-hand side converges to $\|g\|^2_{L^2(\mathbb{R})}$, and the right-hand side converges to

$$\int_{-1/2}^{1/2} |\widehat{g}\,(\xi)|^2 |\widehat{\varphi}(0)|^2\, d\xi = \|\widehat{g}\|^2_{L^2(\mathbb{R})} |\widehat{\varphi}(0)|^2,$$

because $\widehat{\varphi}$ is continuous at $\xi = 0$, and the interchange of limit and integral is legal by the Lebesgue Dominated Convergence Theorem (Theorem A.59). The two limits must coincide; therefore

$$\|g\|^2_{L^2(\mathbb{R})} = \|\widehat{g}\|^2_{L^2(\mathbb{R})} |\widehat{\varphi}(0)|^2.$$

By Plancherel's Identity, $\|g\|_{L^2(\mathbb{R})} = \|\widehat{g}\|_{L^2(\mathbb{R})}$, and we conclude that $\widehat{\varphi}(0) = 1$.

The scaling function φ satisfies the identity (10.5) for all $\xi \in \mathbb{R}$, because $\varphi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Evaluating the identity at $\xi = 0$, we see that

$$\sum_{k\in\mathbb{Z}} |\widehat{\varphi}(k)|^2 = 1.$$

We have just proved that $\widehat{\varphi}(0) = 1$. Then necessarily $\widehat{\varphi}(k) = 0$ for all $k \neq 0$. This ends the proof of part (i).

Part (ii) is a direct application of the Poisson Summation Formula (Theorem 8.37) to φ, in light of the knowledge we have just acquired about the values of $\widehat{\varphi}(k)$. □

Exercise 10.44. Verify that the Fourier transform of $\varphi_{j,k}$ is given by formula (10.16). ♢

10.4.3. Solution of the dilation equation. Given an orthogonal MRA, the bridge between the filter coefficients and the scaling function φ is the *dilation equation* (10.6), for some finite set of filter coefficients $h = \{h_k\}_{k=0}^{L-1}$: $\varphi(x) = \sqrt{2}\sum_{k=0}^{L-1} h_k\varphi(2x - k)$.

It is easy to check that this equation is satisfied for the Haar filter coefficients $h_0 = h_1 = 1/\sqrt{2}$ (all other coefficients vanish) and the Haar scaling function $\varphi(x) = \chi_{[0,1)}(x)$. It is, however, very surprising that there are any other solutions at all. We show that for each finite set of filter coefficients that sum to $\sqrt{2}$, there is a unique solution of the scaling equation. The filter coefficients for the hat function in Example 10.9 satisfy this condition.

Theorem 10.45. *Every dilation equation*

$$\varphi(x) = \sqrt{2}\sum_{k=0}^{L-1} h_k\, \varphi(2x - k) \quad \textit{where} \quad \sum_{k=0}^{L-1} h_k = \sqrt{2}$$

has a solution $\varphi(x)$. The solution is unique and compactly supported. This solution may be a function or a distribution.

We will give a *partial proof* only. It relies on two assumptions and a big theorem (Paley–Wiener). Our theorem is essentially one proved in Daubechies' book [**Dau92**, pp. 174–177]. We outline the main ideas.

Assume without proof: (1) If φ does exist, it has mean value 1. That is, $\int \varphi(x)\,dx = 1$. We saw that it was reasonable to assume the mean is nonzero so as to be able to represent functions of nonzero mean. Once the mean is nonzero, we may as well normalize it to be one. (2) If φ does exist and has a Fourier transform, then $\widehat{\varphi}(\xi)$ is continuous at $\xi = 0$.

Proof of Theorem 10.45. We start with a *formal* (unjustified symbol-pushing) argument which leads to the same infinite product formula for $\widehat{\varphi}(\xi)$, the Fourier transform of $\varphi(x)$, that we encountered in formula (10.15). (There we had an MRA with scaling function φ, so we knew that under assumptions (1) and (2) the product converged to $\varphi(x)$.) We must show that this formula is well-defined and that we can take its inverse Fourier transform to get the function $\varphi(x)$. Then we must show that this $\varphi(x)$ does satisfy the dilation equation (yes, it does, by taking the Fourier transform).

We begin with $\varphi(x) = \sqrt{2}\sum_{k=0}^{L-1} h_k\varphi(2x-k)$, the known dilation equation. Take the Fourier transform to get

$$\begin{aligned}\widehat{\varphi}(\xi) &= \int_{-\infty}^{\infty}\left(\sqrt{2}\sum_{k=0}^{L-1} h_k\varphi(2x-k)\right)e^{-2\pi i\xi x}\,dx\\ &= \sqrt{2}\sum_{k=0}^{L-1} h_k\int_{-\infty}^{\infty}\varphi(2x-k)e^{-2\pi i\xi x}\,dx.\end{aligned}$$

Make the substitution $u = 2x - k$ to get

$$\begin{aligned}\widehat{\varphi}(\xi) &= \sqrt{2}\sum_{k=0}^{L-1} h_k\int_{-\infty}^{\infty}\varphi(u)e^{-2\pi i\xi u/2}e^{-i\xi k/2}\,\frac{du}{2}\\ &= \frac{1}{\sqrt{2}}\sum_{k=0}^{L-1}\big(h_k e^{-2\pi i\xi k/2}\int_{-\infty}^{\infty}\varphi(u)e^{-2\pi i\xi u/2}\,du\big)\\ &= \frac{1}{\sqrt{2}}\big(\sum_{k=0}^{L-1} h_k e^{-2\pi i\xi k/2}\big)\widehat{\varphi}(\xi/2) = m_0(\xi/2)\,\widehat{\varphi}(\xi/2).\end{aligned}$$

The function $m_0(\xi) := (1/\sqrt{2})\sum_{k=0}^{L-1} h_k e^{-i\xi k}$ is the low-pass filter corresponding to the sequence $h_0, h_1, \ldots, h_{L-1}$. It is a *trigonometric* *polynomial.* We will later need the result, which holds by hypothesis:

$$m_0(0) = \frac{1}{\sqrt{2}} \sum_{k=0}^{L-1} h_k e^{-i0k} = \frac{1}{\sqrt{2}} \sum_{k=0}^{L-1} h_k = \frac{1}{\sqrt{2}}(\sqrt{2}) = 1.$$

Summing up, we have that the Fourier transform of a solution to the dilation equation must satisfy $\widehat{\varphi}(\xi) = m_0(\xi/2)\, \widehat{\varphi}(\xi/2)$. Now plug in $\xi/2$ for ξ: $\widehat{\varphi}(\xi/2) = m_0(\xi/4)\, \widehat{\varphi}(\xi/4)$. Repeating and using the fact that $\widehat{\varphi}(0) = \int \varphi(x)\, dx = 1$ (see below), we get

$$\widehat{\varphi}(\xi) = \Big(\prod_{j=1}^{J} m_0(\xi/2^j) \Big)\, \widehat{\varphi}(\xi/2^j).$$

Now let $J \to \infty$ and conclude that

$$\widehat{\varphi}(\xi) = \Big(\prod_{j=1}^{\infty} m_0(\xi/2^j) \Big)\, \widehat{\varphi}(0) = \prod_{j=1}^{\infty} m_0(\xi/2^j).$$

We assumed $\widehat{\varphi}(\xi)$ continuous at $\xi = 0$, so $\widehat{\varphi}(\xi/2^J) \to \widehat{\varphi}(0)$ as $J \to \infty$. We also assumed $\int \varphi(x)\, dx = 1$. An important fact about the Fourier transform is that $\widehat{\varphi}(0) = \int_{-\infty}^{\infty} \varphi(x) e^{-i0x}\, dx = \int_{-\infty}^{\infty} \varphi(x)\, dx$, which is the integral of $\varphi(x)$. So we have $\widehat{\varphi}(0) = 1$, as required.

The formula above is our *guess* for the Fourier transform of $\varphi(x)$: an infinite product of dilations of the symbol $m_0(\xi)$. Does our infinite product *converge*? We are asking whether

$$\prod_{j=1}^{\infty} m_0(\xi/2^j) := \lim_{J\to\infty} \prod_{j=1}^{J} m_0(\xi/2^j)$$

exists for fixed ξ. This infinite product is a limit of partial products, if it exists, just as the sum of an infinite series is the limit of its partial sums, if it exists. We show that in fact the product *converges absolutely*. Just as for sums, this implies that the product converges.

We will need the fact that $m_0(0) = 1$, which we showed earlier. Also, define $C := \max_\xi |m_0'(\xi)|$. This quantity C is finite, since $m_0(\xi)$ is a finite trigonometric polynomial. For each ξ,

$$\begin{aligned} |m_0(\xi)| &= |m_0(\xi) + 1 - m_0(0)| \le 1 + |m_0(\xi) - m_0(0)| \\ &\le 1 + (\max_\xi |m_0'(\xi)|)|\xi - 0| = 1 + C|\xi| \le e^{C|\xi|}, \end{aligned}$$

where in the second inequality we have used the mean value theorem and in the last inequality we have used the Taylor series for the exponential function. Therefore

$$\begin{aligned}|\widehat{\varphi}(\xi)| &= |m_0(\xi/2)||m_0(\xi/4)||m_0(\xi/8)|\cdots \leq e^{C|\xi|/2}e^{C|\xi|/4}e^{C|\xi|/8}\cdots \\ &= \exp\left(C|\xi|\left[2^{-1}+4^{-1}+8^{-1}+\cdots\right]\right) = \exp(C|\xi|) < \infty.\end{aligned}$$

For each $|\xi|$, the absolute value of the infinite product is finite. Therefore the product converges. (It converges absolutely, and in fact uniformly on compact sets.)

Existence: The existence of a distribution $\varphi(x)$ with Fourier transform $\prod_{j=1}^{\infty} m_0(\xi/2^j)$ follows from the Paley–Wiener Theorem. For more details see [**Dau92**, p. 176]. One can conclude that the product $\prod_{j=1}^{\infty} m_0(\xi/2^j)$ *is* the Fourier transform of some function or distribution $\varphi(x)$, supported on $[0, N]$ (see Aside 10.49) and satisfying the dilation equation.

Uniqueness: Take *any* compactly supported distribution or function $f(x)$ satisfying the dilation equation and also having a Fourier transform, with $\widehat{f}(0) = \int f(x)\,dx = 1$. Then

$$\widehat{f}(\xi) = m_0(\xi/2)\widehat{f}(\xi/2) = \cdots = \Big(\prod_{j=1}^{J} m_0(\xi/2^j)\Big)\,\widehat{f}\left(\xi/2^J\right).$$

Take the limit as $J \to \infty$ and use the fact that $\widehat{f}(0) = 1$ to get

$$\widehat{f}(\xi) = \prod_{j=1}^{\infty} m_0(\xi/2^j) = \widehat{\varphi}(\xi).$$

Since the Fourier transform is one-to-one, $f = \varphi$. □

Remark 10.46 (*Zeros at* $\xi = 1/2$*, Vanishing Moments, Order of Approximation*)**.** In general, the faster $|\widehat{\varphi}(\xi)|$ decays to zero as $|\xi| \to \infty$, the better behaved the inverse Fourier transform $\varphi(x)$ is. If the decay "at ∞" of $|\widehat{\varphi}(\xi)|$ is fast enough, $\varphi(x)$ will be a *function*, not just a distribution. If a certain *Condition E* in [**SN**] holds (see p. 310 of this book), then $\varphi(x)$ is a *function* and $\varphi \in L^2(\mathbb{R})$. (If the cascade algorithm converges in L^2, it will have a limit in L^2 since $L^2(\mathbb{R})$ is complete.)

A common requirement on the low-pass filter of a scaling function φ is that $H(\xi)$ *has a zero at* $\xi = 1/2$. This condition reads $0 = H(1/2) = (1/\sqrt{2})\sum_{k=0}^{L-1} h_k e^{\pi i k}$. We conclude that the following *sum rule* must hold:

$$(1/\sqrt{2})\big(h_0 - h_1 + h_2 - \dots (-1)^{L-1} h_{L-1}\big) = 0.$$

This requirement ensures that $\widehat{\varphi}$ has good decay as $|\xi| \to \infty$. First, $\widehat{\varphi}(1) = \widehat{\varphi}(2) = \widehat{\varphi}(4) = \cdots$. (Why?)

(1) $H(0) = (1/\sqrt{2})\sum_{k=0}^{L-1} h_k = 1$.

(2) $H(\xi)$ is 1-periodic since each $e^{-2\pi i k(\xi+1)} = e^{-2\pi i k\xi}(e^{-2\pi i})^k = e^{-2\pi i k\xi}$. Therefore, $1 = H(0) = H(1) = H(2) = H(3) = \cdots$.

(3) Therefore, $\widehat{\varphi}(1) = H(1)\,\widehat{\varphi}(1) = \widehat{\varphi}(2) = H(2)\,\widehat{\varphi}(2) = \widehat{\varphi}(4)\cdots$.

If $|\widehat{\varphi}(\xi)| \to 0$ as $|\xi| \to \infty$ (decay at ∞) and $\widehat{\varphi}(1) = \widehat{\varphi}(2) = \widehat{\varphi}(4) = \cdots = \widehat{\varphi}(2^j) = \cdots$, then we must have all these terms equal to 0. Recall that $\widehat{\varphi}(1) = H(1/2)\,\widehat{\varphi}(1/2)$. So we can make $\widehat{\varphi}(1) = 0$ by requiring $H(1/2) = 0$. Having a zero at $\xi = 1/2$ implies that $H(\xi) = \big[(1 + e^{-i\pi\xi})/2\big]Q(\xi)$, where Q is a trigonometric polynomial.

Exercise 10.47. Verify that if H has p zeros at the point $\xi = 1/2$, that is, $H(\xi) = \big[(1 + e^{-i\pi\xi})/2\big]^p Q(\xi)$, where $Q(\xi)$ is a trigonometric polynomial, then the following p *sum rule* holds: $\sum_{k=0}^{L-1}(-1)^k k^j h_k = 0$, for all $j = 0, 1, \dots, p-1$. **Hint:** Take derivatives of $H(\xi)$ and evaluate them at $\xi = 1/2$. ◊

The more zeros H has at $\xi = 1/2$, the faster $\widehat{\varphi}$ will decay as $|\xi| \to \infty$ and the smoother φ will be. Moreover, if H has p zeros at $\xi = 1/2$, then the wavelet ψ will have p *vanishing moments*, that is,

$$\int_{\mathbb{R}} x^k \psi(x)\,dx = 0, \qquad k = 0, 1, \dots, p.$$

It is argued in [**SN**, Theorem 7.4] that the polynomials 1, x, x^2, ..., x^p belong to the central subspace V_0, and consequently W_0 would be orthogonal to them, which would explain the vanishing moments for $\psi \in W_0$. However the polynomials are not functions in $L^2(\mathbb{R})$, so one should take this argument with a grain of salt (see [**SN**, Note on p. 228]). Vanishing moments for the wavelet are a desirable property for applications. ◊

Exercise 10.48. Show that if the wavelet ψ has p vanishing moments and f has p derivatives, then the wavelet coefficients $\langle f, \psi_{j,k}\rangle$ decay like 2^{-jp}. **Hint:** Write f as its p^{th} Taylor polynomial + remainder. ♢

Aside 10.49. Here is the usual criterion for convergence of an infinite product. Rewrite the product $\prod_{j=1}^{\infty} b_j$ as $\prod_{j=1}^{\infty}(1+a_j)$. The individual terms b_j must tend to 1, so the terms a_j tend to 0. Now, as shown in [**Ahl**, p. 192],

$$\prod_{j=1}^{\infty}(1+a_j) \text{ converges } \Leftrightarrow \sum_{j=1}^{\infty} a_j \text{ converges.}$$

We have shown that $|b_j| = |m_0(\xi/2^j)| \leq 1 + C|\xi|/2^j = 1 + a_j$ and $\sum_{j=1}^{\infty} a_j = \sum_{j=1}^{\infty} C|\xi|/2^j = C|\xi| < \infty$. Therefore the infinite product also converges. ♢

Aside 10.50. If we know only that $\varphi(x)$ is supported in *some* finite interval $[a,b]$, we can use the dilation equation once more to show that this interval *must be* $[0, L-1]$. For suppose the support of $\varphi(x)$ is $[a,b]$. Dilate by $1/2$ and translate by k to get $\operatorname{supp}(\varphi(2x)) = [a/2, b/2]$ and $\operatorname{supp}(\varphi(2x-k)) = [(a+k)/2, (b+k)/2]$. Thus the right-hand side of the dilation equation is supported between $a/2$ (when $k=0$) and $(b+L-1)/2$ (when $k = L-1$). So $[a,b] = [a/2, (b+L-1)/2]$. Thus $a = 0$ and $b = L-1$, and $\varphi(x)$ is supported on $[0, L-1]$.

Exercise 10.51. Show that if the scaling function φ of an MRA is supported on the interval $[0, L-1]$, then Mallat's wavelet is supported on the interval $[-(L-2)/2, L/2]$. ♢

We have just shown that "FIR ⇒ compact support". In other words, if the wavelet has *finite impulse response* (that is, only finitely many filter coefficients are nonzero), then the wavelet and scaling functions are compactly supported. Is the converse true? If the wavelet and scaling functions are compactly supported, are there only finitely many taps (coefficients)? **Answer:** No! An example was given by V. Strela at the 1999 AMS/MAA Joint Meetings, Wavelets special session. See the project in Subsection 10.5.2. ♢

10.4.4. Summary. To summarize, if a given filter H is the low-pass filter of an orthogonal MRA, then H must satisfy both the QMF condition and the normalization condition $H(0) = 1$. For finite filters H, the QMF condition and the normalization condition are sufficient to guarantee the existence of a solution φ to the scaling equation. Then φ can be calculated via the infinite product (10.15). Example 11.23 shows how to produce good approximations for both φ and the Mallat wavelet ψ.

For infinite filters an extra decay assumption is necessary. However, it is not sufficient to guarantee the orthonormality of the integer shifts of φ. But if for example the inequality $\inf_{|\xi|\leq 1/4} |H(\xi)| > 0$ also holds, then $\{\varphi_{0,k}\}$ is an orthonormal set in $L^2(\mathbb{R})$. For more details, see for example [**Fra**, Chapter 5] or [**Woj03**].

The existence of a solution for the scaling equation can be expressed in the language of fixed-point theory. Given a low-pass filter H, define a transformation T by $T\varphi(t) := \sqrt{2}\sum_k h_k \varphi(2t-k)$. If T has a fixed point, that is, if there is a $\varphi \in L^2(\mathbb{R})$ such that $T\varphi = \varphi$, then the fixed point is a solution to the dilation equation. This is the idea behind the cascade algorithm that we describe in Chapter 11.

10.5. Projects: Twin dragon; infinite mask

We invite you to investigate some unusual wavelets.

10.5.1. Project: The twin dragon. We mentioned that there are *nonseparable* two-dimensional MRAs, meaning two-dimensional MRAs that are not tensor products of one-dimensional MRAs (Section 10.2.1). In this project we explore an analogue of the Haar basis. The scaling function is the characteristic function of a certain two-dimensional set.

(a) The set can be rather complicated, for example a self-similar set with fractal boundary, such as the so-called *twin dragon*. Search on the Internet or elsewhere for an explanation of what they are and why the characteristic function of a twin dragon is the scaling function of a two-dimensional MRA. Design a program to generate the set, based on the ideas of multiresolution.

(b) More recently, simpler sets have been found whose characteristic functions are the scaling function of a two-dimensional MRA. Write a report about this approach. A place to start is [**KRWW**].

10.5.2. Project: Wavelets with infinite mask. We say a function $\varphi(x)$ is *refinable* if it satisfies a dilation equation, that is, if $\varphi(x) = \sqrt{2}\sum_k h_k\varphi(2x-k)$. The sequence of filter coefficients $\{h_k\}_{k\in\mathbb{Z}}$ is called the *mask* of the function. Can a compactly supported refinable function have an *infinite* mask?

The answer, "Yes", was announced in 1999 by Vasily Strela and was proved as part of a much more general result in joint work with Gilbert Strang and Ding-Xuan Zhou [**SSZ**]. Another example appeared in [**Zho**]. Letting $z = e^{2\pi i\xi}$ be a point on the unit circle, rewrite the *symbol*

$$m_0(\xi) = (1/\sqrt{2})\sum_k h_k e^{-2\pi ik\xi}$$

of a filter as $m_0(z) = (1/\sqrt{2})\sum_k h_k z^{-k}$.

With this notation, consider the filter associated with

$$m_0(z) = \frac{1}{2} + \frac{1}{4}z^{-1} + \frac{1}{8}z^{-2} + \sum_{k=0}^{\infty}(-1/2)^{k+1}\,\frac{3}{8}\,z^{-(k+3)}.$$

(a) Prove that the function $\varphi(x)$ that is supported on $[0,2]$ (!) takes the value 2 on $[0,1)$ and the value 1 on $[1,2]$ and satisfies the dilation equation with the (infinite!) mask from the $m_0(z)$ above.

(b) Does this filter satisfy the quadrature mirror filter condition?

(c) Is there a corresponding wavelet? If so, find an analytic description or use the cascade algorithm to generate a picture of the wavelet.

10.6. Project: Linear and nonlinear approximations

In Section 10.2.2 we discussed linear approximation and nonlinear approximation. The *linear approximation error* $\varepsilon_l[N]$, or *mean square error*, for the N^{th} Fourier partial sum of f is defined by

$$\varepsilon_l[N] := \|f - S_N f\|^2_{L^2([0,1))} = \sum_{|n|>N} |\widehat{f}(n)|^2.$$

Now consider the same coefficients, arranged in decreasing order. We call the resulting sum the *nonlinear approximation error* $\varepsilon_{\text{nonl}}[N]$ for the N^{th} Fourier partial sum of f:

$$\varepsilon_{\text{nonl}}[N] := \sum_{k\in\mathbb{N}, |n_k|>N, |\widehat{f}(n_k)|\geq|\widehat{f}(n_{k+1})|} |\widehat{f}(n_k)|^2 \leq \varepsilon_{\text{l}}[N].$$

(a) The smoother a function is, the faster its Fourier coefficients shrink as $n \to \infty$. By Theorem 3.14, if $f \in C^k(\mathbb{T})$, then $|\widehat{f}(n)| \leq Cn^{-k}$ for some constant C. Therefore the Fourier linear approximation error has a similar rate of decay. Show that for $f \in C^k(\mathbb{T})$,

$$\varepsilon_{\text{l}}[N] \leq \big(2C^2/(2k-1)\big)\, N^{-(2k-1)}.$$

In Section 3.2.3 we relate smoothness to the uniform rate of convergence on $\mathbb{T}$ of $f - S_N f$. Here you are asked to study the square of the L^2 rate of convergence. More precise estimates, obtained by introducing an L^2 *translation error*, have applications to the absolute convergence of Fourier series; see [**Pin**, Section 1.3.3]. A converse theorem goes by the name of *Bernstein's Theorem*[6]; see [**Pin**, Section 1.5.5].

(b) Consider a function that has a discontinuity, for instance $\chi_{[0,2/3)}(x)$ on $(0,1)$. Show that the linear approximation error and the nonlinear approximation error in the Fourier basis are both of order N^{-1}:

$$\varepsilon_{\text{l}}[N] \sim N^{-1} \quad \text{and} \quad \varepsilon_{\text{nonl}}[N] \sim N^{-1}.$$

(c) Replacing the Fourier basis by the Haar basis on $[0,1]$, write definitions for $\varepsilon_{\text{l}}(N)$, the *linear approximation error in the Haar basis* for the N^{th} Haar partial sum of a function f, and for $\varepsilon_{\text{nonl}}(N)$, the *nonlinear approximation error in the Haar basis* for the N^{th} Haar partial sum of f. Be aware that there are 2^j Haar functions corresponding to intervals of length 2^{-j}. For example, the first fifteen Fourier basis functions correspond to eight different frequencies (in pairs $e^{\pm iN\theta}$ when $N \geq 1$), while the first fifteen Haar basis functions correspond to only four frequencies. Also, to obtain a basis of $L^2([0,1]$, one has to include the characteristic function of the interval $[0,1]$ with the Haar functions.

(d) Show that for the $L^2(0,1)$ function $\chi_{[0,2/3)}(x)$, the linear approximation error in the Haar basis is at least of order N^{-1}: there is a

[6]Named after Russian mathematician Sergei Natanovich Bernstein (1880–1968).

constant C such that

$$\varepsilon_l(N) \geq CN^{-1} \qquad \text{for all } N \in \mathbb{N}.$$

(e) The decay of the linear approximation error, in the Fourier or wavelet basis, for any discontinuous function with $n \geq 1$ vanishing moments can be shown to obey a similar lower bound to that in part (d). However, for the discontinuous function $\chi_{[0,2/3)}$, the *nonlinear* approximation error in the Haar basis is much smaller than the linear one. Show that for $\chi_{[0,2/3)}$ on $(0,1)$,

$$\varepsilon_{\text{nonl}}(N) = C2^{-N}.$$

(f) For some discontinuous functions, wavelet bases achieve a decay in the nonlinear approximation error as rapid as the decay in the linear approximation error for continuous functions. Investigate the literature to make this statement precise. Start with [**Mall09**, Theorem 9.12]. There is a wealth of information related to this project in [**Mall09**, Chapter 9].

Chapter 11

Calculating with wavelets

In this chapter we explore how to compute with wavelets.

First, we recall the *zooming* properties of the Haar basis, how they are mathematically encoded in the *Haar multiresolution analysis*, and we describe with a cartoon example the existence of a fast algorithm to actually compute with the Haar functions (Section 11.1). Second, we discuss the cascade algorithm to search for a solution of the scaling equation (Section 11.2). Third, we show how to implement the wavelet transform via filter banks in a way analogous to what we did for the Haar functions (Section 11.3). The key to the algorithm is a sequence of numbers (the filter coefficients) that determine the low-pass filter, the scaling function, and the wavelet; this is the famous Fast Wavelet Transform. Fourth, we present informally some competing attributes we would like the wavelets to have and a catalog of existing selected wavelets listing their properties (Section 11.4). We end with a number of projects inviting the reader to explore some of the most intriguing applications of wavelets (Section 11.5).

11.1. The Haar multiresolution analysis

Before considering how to construct wavelets from scaling functions associated to an orthogonal MRA, we revisit the Haar multiresolution analysis.

The scaling function φ for the Haar MRA is $\chi_{[0,1)}$, the characteristic function of the unit interval, $\varphi(x) = 1$ for $0 \leq x < 1$ and $\varphi(x) = 0$ elsewhere. The subspace V_0 is the closure in $L^2(\mathbb{R})$ of the linear span of the integer translates of the Haar scaling function φ,

$$V_0 := \overline{\text{span}(\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}})}.$$

It consists of piecewise constant functions f with jumps only at the integers and such that the sequence of coefficients lies in $\ell_2(\mathbb{Z})$, and therefore the function f lies in $L^2(\mathbb{R})$. More precisely,

$$V_0 = \Big\{ f = \sum_{k\in\mathbb{Z}} a_k \varphi_{0,k} : \sum_{k\in\mathbb{Z}} |a_k|^2 < \infty \Big\}.$$

The functions $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis of V_0.

Similarly, the subspace

$$V_j := \overline{\text{span}(\{\varphi_{j,k}\}_{k\in\mathbb{Z}})}$$

is the closed subspace consisting of piecewise constant functions in $L^2(\mathbb{R})$ with jumps only at the integer multiples of 2^{-j}, so that

$$V_j = \Big\{ f = \sum_{k\in\mathbb{Z}} a_k \varphi_{j,k} : \sum_{k\in\mathbb{Z}} |a_k|^2 < \infty \Big\}.$$

Likewise, the functions $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis of V_j.

The orthogonal projection $P_j f$ onto V_j is the piecewise constant function with jumps at the integer multiples of 2^{-j}, whose value on the interval $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ is given by the integral average of f over the interval $I_{j,k}$. To go from $P_j f$ to $P_{j+1} f$, we add the difference $Q_j f$ (the expectation operators P_j and the difference operators Q_j were defined in Section 9.4.4). We showed in Lemma 9.35 that $Q_j f$ coincides with

$$Q_j f = \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k},$$

where ψ is the Haar wavelet h, defined by $\psi(x) = -1$ for $0 \leq x < 1/2$, $\psi(x) = 1$ for $1/2 \leq x < 1$, and $\psi(x) = 0$ elsewhere. Hence Q_j is the orthogonal projection onto W_j, the closure of the linear span of $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$. The subspace

$$W_j = \overline{\text{span}(\{\psi_{j,k}\}_{k\in\mathbb{Z}})}$$

consists of piecewise constant functions in $L^2(\mathbb{R})$ with jumps only at integer multiples of $2^{-(j+1)}$ and average value zero between integer multiples of 2^{-j}.

We can view the averages $P_j f$ at resolution j as successive approximations to the original signal $f \in L^2(\mathbb{R})$. These approximations are the orthogonal projections $P_j f$ onto the approximation spaces V_j. The *details*, necessary to move from level j to the next level $(j+1)$, are encoded in the Haar coefficients at level j, more precisely in the orthogonal projections $Q_j f$ onto the detail subspaces W_j. Starting at a low resolution level, we can obtain better and better resolution by adding the details at the subsequent levels. As $j \to \infty$, the resolution is increased. The steps get smaller (length 2^{-j}), and the approximation converges to f in L^2 norm (this is the content of equation (9.13) in Theorem 9.36) and a.e. (the Lebesgue Differentiation Theorem). Clearly the subspaces are nested, that is, $V_j \subset V_{j+1}$, and their intersection is the trivial subspace containing only the zero function (this is equation (9.12) in Theorem 9.36). We have shown that the Haar scaling function generates an orthogonal MRA.

We give an example of how to decompose a function into its projections onto the Haar subspaces. Our presentation closely follows [**MP**, Section 3.2], but uses a different initial vector.

We select a coarsest scale V_n and a finest scale V_m, $n < m$. We truncate the doubly infinite chain of nested approximation subspaces

$$V_n \subset V_{n+1} \cdots \subset V_{m-1} \subset V_m$$

and obtain

$$V_m = V_n \oplus W_n \oplus W_{n+1} \oplus \cdots \oplus W_{m-2} \oplus W_{m-1}. \tag{11.1}$$

In the example we choose $n = -3$ and $m = 0$, so the coarser scale corresponds to intervals of length $2^3 = 8$, and the finest to intervals of length one. We will go through the decomposition process in the text using vectors. It is also enlightening to look at the graphical version in Figure 11.1.

We begin with a vector of $8 = 2^3$ *samples* of a function, which we assume to be the average value of the function on eight intervals of length 1, so that our function is supported on the interval $[0, 8]$. For

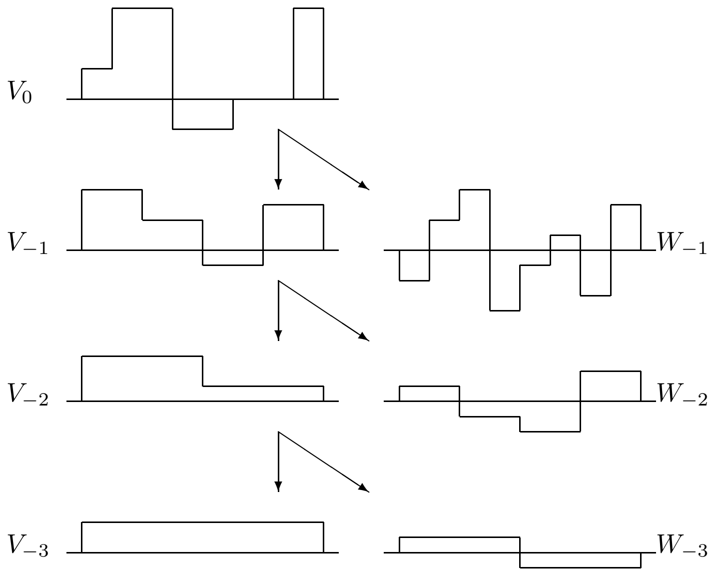


Figure 11.1. A wavelet decomposition of a subspace: $V_0 = V_{-3} \oplus W_{-3} \oplus W_{-2} \oplus W_{-1}$. This figure is adapted from Figure 14 of [**MP**, p. 35], which also graces the cover of that book.

our example, we choose the vector

$$v_0 = [2, 6, 6, -2, -2, 0, 0, 6]$$

to represent our function in V_0. The convention we are using throughout the example is that the vector $v = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7] \in \mathbb{C}^8$ represents the step function $f(x) = a_j$ for $j \le x < j+1$ where $j = 0, 1, \ldots, 7$ and $f(x) = 0$ otherwise.

Exercise 11.1. Use the above convention to describe the scaling functions in V_0, V_{-1}, V_{-2}, V_{-3} supported on $[0, 8)$ and the Haar functions in W_{-1}, W_{-2}, W_{-3} supported on $[0, 8)$. For example,

$$\varphi_{0,3} = [0, 0, 0, 1, 0, 0, 0, 0], \quad \varphi_{-1,1} = [0, 0, 1/\sqrt{2}, 1/\sqrt{2}, 0, 0, 0, 0]. \quad \Diamond$$

Exercise 11.2. Verify that if f, g are functions in V_0 supported on $[0, 8)$, described according to our convention by the vectors $v, w \in \mathbb{C}^8$, then the $L^2(\mathbb{R})$ inner product $\langle f, g\rangle_{L^2(\mathbb{R})}$ coincides with $v\cdot w$, the inner product in $\mathbb{C}^8$. $\Diamond$

To construct the projection of v_0 onto V_{-1}, we average pairs of values, obtaining

$$v_{-1} = [4, 4, 2, 2, -1, -1, 3, 3].$$

The difference $v_0 - v_{-1} = w_{-1}$ is in W_{-1}, so we have

$$w_{-1} = [-2, 2, 4, -4, -1, 1, -3, 3].$$

Repeating this process, we obtain the projections of v_0 onto V_{-2}, W_{-2}, V_{-3}, and W_{-3}:

$$\begin{aligned} v_{-2} &= [3, 3, 3, 3, 1, 1, 1, 1], \\ w_{-2} &= [1, 1, -1, -1, -2, -2, 2, 2], \\ v_{-3} &= [2, 2, 2, 2, 2, 2, 2, 2], \quad \text{and} \\ w_{-3} &= [1, 1, 1, 1, -1, -1, -1, -1]. \end{aligned}$$

To compute the coefficients of the expansion (10.3), we must compute the inner product $\langle f, \varphi_{j,k}\rangle$ for the function (10.2). By the results of Exercises 11.1 and 11.2 in terms of our vectors, we have for example

$$\langle f, \varphi_{0,3}\rangle = \langle [2, 6, 6, -2, -2, 0, 0, 6], [0, 0, 0, 1, 0, 0, 0, 0]\rangle = -2$$

and

$$\langle f, \varphi_{-1,1}\rangle = \langle [2, 6, 6, -2, -2, 0, 0, 6], [0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, 0]\rangle = \frac{4}{\sqrt{2}}.$$

The Haar scaling function φ satisfies the *dilation equation*

$$\varphi(t) = \varphi(2t) + \varphi(2t - 1). \tag{11.2}$$

Therefore $\varphi_{j,k} = (\varphi_{j+1,2k} + \varphi_{j+1,2k+1})/\sqrt{2}$, and so

$$\langle f, \varphi_{j,k}\rangle = (1/\sqrt{2})(\langle f, \varphi_{j+1,2k}\rangle + \langle f, \varphi_{j+1,2k+1}\rangle).$$

Thus we can also compute

$$\langle f, \varphi_{-1,1}\rangle = (1/\sqrt{2})\left(\langle f, \varphi_{0,2}\rangle + \langle f, \varphi_{0,3}\rangle\right) = 4/\sqrt{2}.$$

Similarly, the Haar wavelet satisfies the *two-scale difference equation*

$$\psi(t) = \varphi(2t) - \varphi(2t - 1), \tag{11.3}$$

and thus we can recursively compute

$$\langle f, \psi_{j,k}\rangle = (1/\sqrt{2})(\langle f, \varphi_{j+1,2k}\rangle - \langle f, \varphi_{j+1,2k+1}\rangle).$$

Exercise 11.3. Verify that the two-scale recurrence equation (11.2) and the two-scale difference equation (11.3) hold for the Haar scaling function and the Haar wavelet. ◊

The Haar coefficients $\langle f, \psi_{j,k}\rangle$ for fixed j are called the *differences*, or *details*, of f at scale j and are denoted $d_{j,k}$. The coefficients $\langle f, \varphi_{j,k}\rangle$ for fixed j are called the *averages* of f at scale j and are denoted $a_{j,k}$. Evaluating the whole set of Haar coefficients $d_{j,k}$ and averages $a_{j,k}$ requires $2(N-1)$ additions and $2N$ multiplications, where N is the number of data points. The discrete wavelet transform can be performed using a similar *cascade algorithm* with complexity N, the *Fast Wavelet Transform* (FWT); see Section 11.3. This is the same algorithm as the Fast Haar Transform we discussed in Chapter 6 in the language of matrices. Let us remark that an arbitrary change of basis in N-dimensional space requires multiplication by an $N \times N$ matrix; hence *a priori* one requires N^2 multiplications. The attentive reader may remember that the Haar matrix had $N \log N$ nonzero entries, which was already as good as the FFT without any magical algorithm. However we can do even better with a magical algorithm, the FWT. The FWT is a significant improvement over the brute force matrix multiplication of an already sparse matrix, and therefore it has important consequences for applications.

11.2. The cascade algorithm

In this section we describe the cascade algorithm and illustrate how it can be used to obtain approximations for the scaling function.

Given the filter coefficients $h_0, \ldots, h_N$ such that $\sum_k h_k = \sqrt{2}$, we already have an explicit formula for the scaling function. But our formula involves an infinite product on the Fourier side, so it is not ideal for actually computing the scaling function. The *cascade algorithm*, below, is an iterative procedure for computing successive approximations to the scaling function. It arises from iteration of the dilation equation and is the counterpart of the repeated products on the Fourier side.

The dilation equation is $\varphi(x) = \sqrt{2}\sum_k h_k\varphi(2x-k)$.

Here are some examples of filter coefficients.

Example 11.4. The Haar filter coefficients:

$$\boxed{h_0 = \frac{1}{\sqrt{2}}, \qquad h_1 = \frac{1}{\sqrt{2}}.}$$

See Figure 11.2 on the left. ◊

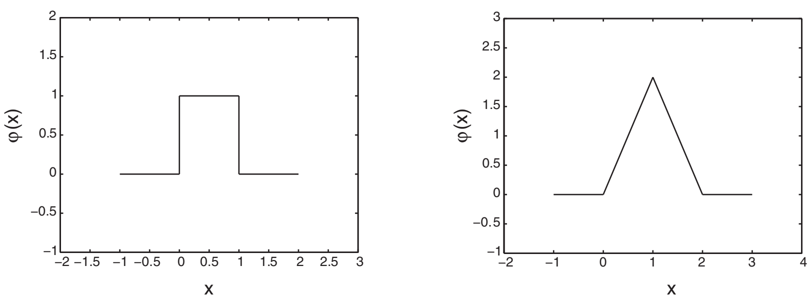


Figure 11.2. The Haar scaling function (left) and the hat function (right).

Example 11.5. The hat function filter coefficients:

$$\boxed{h_0 = \frac{1}{2\sqrt{2}}, \qquad h_1 = \frac{1}{\sqrt{2}}, \quad h_2 = \frac{1}{2\sqrt{2}}.}$$

The corresponding dilation equation is

$$\varphi(x) = \frac{1}{2}\varphi(2x) + \varphi(2x-1) + \frac{1}{2}\varphi(2x-2).$$

See Figure 11.2 on the right. ◊

Example 11.6. The Daubechies 4-tap filter coefficients:

$$\boxed{h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \quad h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \quad h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, \quad h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}.}$$ ◊

Given the filter coefficients h_k, we can think of the solution $\varphi(x)$ as a *fixed point* of an iterative scheme. A sequence of functions $\varphi^{(i)}(x)$ is generated by iterating the sum on the right-hand side of the dilation equation:

$$\varphi^{(i+1)}(x) := \sqrt{2}\sum_k h_k \varphi^{(i)}(2x-k). \tag{11.4}$$

If $\varphi^{(i)}(x)$ converges to a limit function $\varphi(x)$, then this $\varphi(x)$ automatically satisfies the dilation equation. That is, $\varphi(x)$ is a fixed point of the iterative scheme. On the Fourier transform side, the Fourier transform of equation (11.4) is

$$\widehat{\varphi}^{(i+1)}(\xi) := m_0(\xi/2)\widehat{\varphi}^{(i)}(\xi/2).$$

Remark 11.7. Iterative algorithms do not always converge. What conditions must we impose on the filter coefficients to guarantee that the cascade algorithm converges? It turns out that the following condition, named Condition E (E for eigenvectors) by Gilbert Strang[1], is sufficient (see [**SN**, pp. 239–240]).

Condition E: All eigenvalues of the *transition matrix* $\mathrm{T} = (\downarrow 2)HH^T$, where $H_{ij} = \sqrt{2}h_{i-j}$, have $|\lambda| < 1$, except for a simple eigenvalue at $\lambda_1 = 1$.

Here $(\downarrow 2)\varphi[n] = \varphi[2n]$; this is *downsampling.* The cascade algorithm is really an iteration of the *filter matrix* $(\downarrow 2)H$ with ij^{th} entry $H_{ij} := \frac{1}{2}h_{i-j}$. It is an infinite iteration, producing an infinite product. In practice we might stop at some finite stage. See Section 11.3. $\Diamond$

Remark 11.8. Which initial functions (*seeds*) $\varphi^{(0)}(x)$ will actually lead to a solution $\varphi(x)$? (Compare with Newton's method, say for a quadratic such as $f(z) = z^2 + z + 1 = 0$. Let L be the perpendicular bisector of the line segment between the two (complex) roots $\alpha_\pm = (-1/2) \pm i(\sqrt{3}/2)$ of the quadratic. If the initial seed z_0 for Newton's method[2] lies closer to one root than to the other, then Newton's method will converge to that closer root. However, if the seed lies on the line L of equal distance from the two roots, then Newton's method

[1] American mathematician William Gilbert Strang (born 1934).

[2] Named after the English physicist, mathematician, astronomer, natural philosopher, alchemist, and theologian Sir Isaac Newton (1643–1727).

does not converge. The case of the cubic is even more interesting: see [**CG**, p. 29]). ◊

(1) The cascade normally begins with $\varphi^{(0)}(x) = \chi_{[0,1)}(x)$, the box function (corresponds to an *impulse*).

(2) Here is an exact condition on $\varphi^{(0)}(x)$ (*assuming* the filter coefficients satisfy Condition E): If for all t the sum of the integer translates of $\varphi^{(0)}$ is 1, namely, $\sum_k \varphi^{(0)}(x-k) \equiv 1$, then the cascade algorithm converges.

We give several examples to show how the cascade algorithm works on various filter coefficients and seed functions.

Example 11.9. Filter coefficients

$$\boxed{h_0 = \frac{4}{3\sqrt{2}}, \quad h_1 = \frac{2}{3\sqrt{2}}.}$$

The seed $\varphi^{(0)}(x)$ is the box function. First, $\sum_{k\in\mathbb{Z}} h_k = \sqrt{2}$. Iterating the filter by plugging $\varphi^{(0)}(x)$ into the right-hand side of the dilation equation, we get

$$\varphi^{(1)}(x) = (4/3)\varphi^{(0)}(2x) + (2/3)\varphi^{(0)}(2x-1).$$

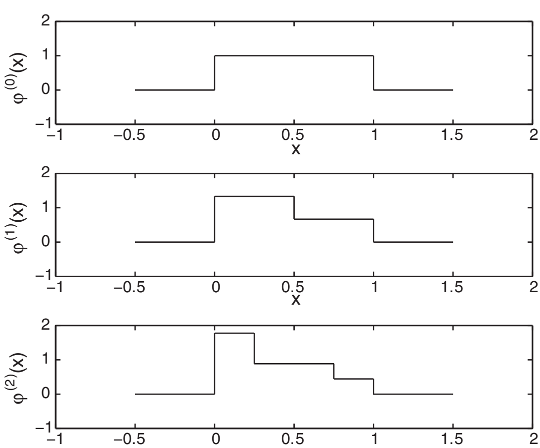


Figure 11.3. Three steps of a cascade with a bad filter.

Note that the area under $\varphi^{(i)}(x)$ is preserved. (We want $\int \varphi = \int \varphi^{(i)} = 1$ for all i.) Filter again by plugging in $\varphi^{(1)}(x)$:

$$\varphi^{(2)}(x) = (4/3)\varphi^{(1)}(2x) + (2/3)\varphi^{(1)}(2x-1).$$

Height of 1^{st} quarter box : $(\frac{4}{3})^2 = \frac{16}{9}$.
Height of 2^{nd} quarter box : $(\frac{4}{3})(\frac{2}{3}) = \frac{8}{9}$.
Height of 3^{rd} quarter box : $(\frac{2}{3})(\frac{4}{3}) = \frac{8}{9}$.
Height of 4^{th} quarter box : $(\frac{2}{3})^2 = \frac{4}{9}$.

We want the functions $\varphi^{(i)}$ to converge to a limit function. But $x = 0$ is a bad point: $\varphi^{(i)}(0) = (4/3)^i \to \infty$. These $\varphi^{(i)}$ perhaps converge in some weak sense, but *not* pointwise. See Figure 11.3, which shows three steps of a cascade with a bad filter. These coefficients h_k *don't* satisfy Condition E (check!). ♢

Example 11.10. This is the same as Example 11.9, but on the z-side. Recall that

$$m_0(z) = (1/\sqrt{2})\sum_k h_k z^{-k} \quad \text{and} \quad m_0(\xi) = (1/\sqrt{2})\sum_k h_k e^{-2\pi i k\xi}.$$

(We are using m_0 to denote both a function of ξ and a function of $z = e^{2\pi i\xi}$.) Hence, if

$$\boxed{h_0 = \frac{4}{3\sqrt{2}}, \quad h_1 = \frac{2}{3\sqrt{2}},}$$

then the symbol is $m_0(z) = (1/2)\left((4/3) + (2/3)z^{-1}\right)$.

The heights of the boxes are as follows (from Example 11.9):

$$\begin{aligned}
\varphi^{(1)}\colon\ \tfrac{4}{3},\ \tfrac{2}{3} \quad &\longleftrightarrow \quad \tfrac{4}{3} + \tfrac{2}{3}z^{-1} &&= 2m_0(z),\\
\varphi^{(2)}\colon\ \tfrac{16}{9},\ \tfrac{8}{9},\ \tfrac{8}{9},\ \tfrac{4}{9} \quad &\longleftrightarrow \quad \tfrac{16}{9} + \tfrac{8}{9}z^{-1} + \tfrac{8}{9}z^{-2} + \tfrac{4}{9}z^{-3} &&= 2^2 m_0(z^2)m_0(z).
\end{aligned}$$

It is important to note that we did *not* square $m_0(z)$. Instead, we multiplied it by $m_0(z^2)$. The same pattern gives us $m_0^{(3)}(z) = m_0(z^4)m_0(z^2)m_0(z)$. After i steps, we obtain

$$m_0^{(i)}(z) = \prod_{j=0}^{i-1} m_0(z^{2^j}).$$

This expression is reminiscent of the infinite product we had before:

$$\widehat{\varphi}(\xi) = \lim_{J\to\infty} \prod_{j=1}^{J} m_0(\xi/2^j). \qquad \Diamond$$

Example 11.11. The usual averaging filter:

$$\boxed{h_0 = \frac{1}{\sqrt{2}}, \quad h_1 = \frac{1}{\sqrt{2}}}$$

for the Haar wavelet. As before, $\varphi^{(0)}(x)$ is the box function. Then with the usual filter step we get

$$\varphi^{(1)}(x) = (1)\varphi^{(0)}(2x) + (1)\varphi^{(0)}(2x-1) = \varphi^{(0)}(x).$$

The algorithm converges immediately, because we started iterating *at a fixed point* of the iteration, as Figure 11.4 illustrates.

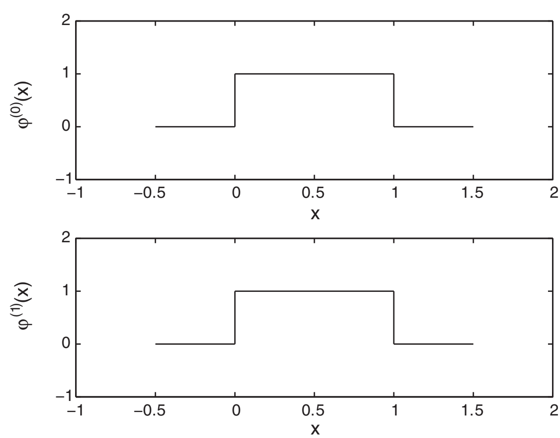


Figure 11.4. Two steps of a cascade with the Haar filter.

On the z-side: The symbol $m_0(z) = (1/2) + (1/2)z^{-1}$. Notice that m_0 is the Haar low-pass filter H. The next step is

$$\begin{aligned} m_0^{(2)}(z) &= m_0(z^2)m_0(z) = \big((1/2) + (1/2)z^{-2}\big)\big((1/2) + (1/2)z^{-1}\big) \\ &= (1/4) + (1/4)z^{-1} + (1/4)z^{-2} + (1/4)z^{-3}. \end{aligned}$$

The heights of the boxes are 2^2 times these coefficients, so they all equal 1. At the i^{th} step we have 2^i boxes, each of height 1. So we end up with the *box function* again. ♢

Example 11.12. Filter given by

$$\boxed{h_0 = \frac{1}{2\sqrt{2}}, \quad h_1 = \frac{1}{\sqrt{2}}, \quad h_2 = \frac{1}{\sqrt{2}}.}$$

Beginning with $\varphi^{(0)}(x)$ as the usual box function, we obtain *three half-boxes.* Filter:

$$\varphi^{(1)}(x) = (1/2)\varphi^{(0)}(2x) + \varphi^{(0)}(2x-1) + (1/2)\varphi^{(0)}(2x-2).$$

As we keep filtering, we see (Figure 11.5) that $\varphi^{(i)}$ converges towards the *hat function.* Surprise, surprise .... ♢

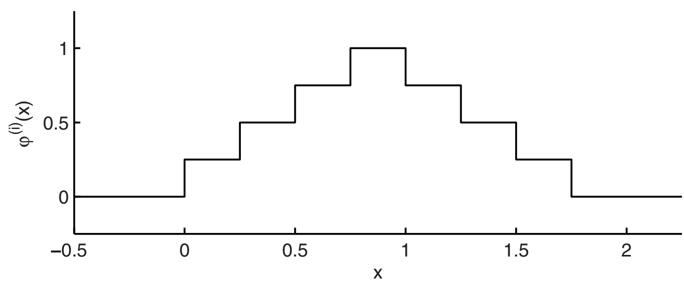


Figure 11.5. A cascade where half-boxes lead to a hat.

11.3. Filter banks, Fast Wavelet Transform

In this section we give a glimpse of how we would implement the wavelet transform once an MRA is at our disposal, in a way analogous to the implementation for the Haar functions. A *fast algorithm*, similar to the one described for the Haar basis in Section 11.1, can be implemented using *filter banks* to provide a Fast Wavelet Transform. First we need to define *convolution of sequences.*

Definition 11.13. Let $a = \{a_m\}_{m\in\mathbb{Z}}$ and $b = \{b_m\}_{m\in\mathbb{Z}}$ be two sequences indexed by the integers. Their *convolution* $a * b$ is the new sequence whose ℓ^{th} term is given by

$$a * b(\ell) = \sum_{m\in\mathbb{Z}} a_{\ell-m} b_m. \qquad \lozenge$$

If one of the sequences has only finitely many nonzero terms, the convolution is a well-defined sequence, since the apparently infinite sums reduce to finite sums.

We have at our disposal the low-pass filter H of length L, with nonzero coefficients $\{h_j\}_{j=0}^{L-1}$, corresponding to an orthogonal MRA with scaling function φ.

Definition 11.14. Given a function $f \in L^2(\mathbb{R})$, the *approximation coefficient* $a_{j,k}$ and the *detail coefficient* $d_{j,k}$ are the inner products of f with the scaling function $\varphi_{j,k}$ and with the wavelet $\psi_{j,k}$, respectively. In other words,

$$a_{j,k} := \langle f, \varphi_{j,k}\rangle, \qquad d_{j,k} := \langle f, \psi_{j,k}\rangle. \qquad \lozenge$$

We would like to calculate the approximation and detail coefficients efficiently. Later we would like to use them to reconstruct efficiently the orthogonal projections onto the the approximation subspaces V_j and the detail subspaces W_j.

Theorem 11.15. *Suppose we have the approximation coefficients (or* scaled samples*)* $\{a_{J,k}\}_{k=0}^{N-1}$ *at scale* J*, where* $2^J = N$*, of a function* f *defined on the interval* $[0,1]$*. Then the coarser approximation and detail coefficients* $a_{j,k}$ *and* $d_{j,k}$ *for scales* $0 \le j < J$*, where* $k = 0, 1, \ldots, 2^j - 1$*, can be calculated in order* LN *operations. Here* L *is the length of the filter, and* N *is the number of samples, in other words, the number of coefficients in the finest approximation scale.*

Proof. Let a_j denote the sequence $\{a_{j,k}\}_{k=0}^{2^j-1}$ and let d_j denote the sequence $\{d_{j,k}\}_{k=0}^{2^j-1}$. Notice that we want to calculate a total of $2(1+2+2^2+\cdots+2^{J-1}) \sim 2^{J+1} = 2N$ coefficients. We will check that calculating each coefficient $a_{j,\ell}$ and $d_{j,\ell}$ requires at most L multiplications when the finer scale approximation coefficients $\{a_{j+1,k}\}_{k=0}^{2^{J+1}-1}$

are given. The complexity of the algorithm will therefore be of order LN as claimed.

In order to see why this is possible, let us consider the simpler case of calculating a_0 and d_0 given a_1. The scaling equation connects $\varphi_{0,\ell}$ to $\{\varphi_{1,m}\}_{m\in\mathbb{Z}}$:

$$\begin{aligned}\varphi_{0,\ell}(x) &= \varphi(x-\ell) = \sum_{k\in\mathbb{Z}} h_k \varphi_{1,k}(x-\ell)\\ &= \sum_{k\in\mathbb{Z}} h_k\sqrt{2}\varphi(2x-(2\ell+k))\\ &= \sum_{m\in\mathbb{Z}} h_{m-2\ell}\varphi_{1,m} = \sum_{m\in\mathbb{Z}} \overline{\tilde{h}_{2\ell-m}}\varphi_{1,m},\end{aligned}$$

where the *conjugate flip of* h_k is $\tilde{h}_k := \overline{h_{-k}}$ and in the third line we made the change of summation index $m = 2\ell + k$ (so $k = m - 2\ell$).

We can now compute $a_{0,\ell}$ in terms of a_1:

$$\begin{aligned}a_{0,\ell} &= \langle f, \varphi_{0,\ell}\rangle = \sum_{m\in\mathbb{Z}} \tilde{h}_{2\ell-m}\langle f, \varphi_{1,m}\rangle\\ &= \sum_{m\in\mathbb{Z}} \tilde{h}_{2\ell-m}a_{1,m} = \tilde{h} * a_1(2\ell),\end{aligned}$$

where $\widetilde{h} = \{\widetilde{h}_k\}_{k\in\mathbb{Z}}$ is the conjugate flip of the low-pass filter h. The sequence $\widetilde{h}$ has the same length L as the sequence h. Therefore the convolution above requires only L multiplications.

Similarly, for the detail coefficients $d_{0,\ell}$, the two-scale equation connects $\psi_{0,\ell}$ to $\{\varphi_{1,m}\}_{m\in\mathbb{Z}}$. The only difference is that the filter coefficients are $\{g_k\}_{k\in\mathbb{Z}}$ instead of $\{h_k\}_{k\in\mathbb{Z}}$. We obtain

$$\psi_{0,\ell}(x) = \sum_{m\in\mathbb{Z}} \overline{\tilde{g}_{2\ell-m}}\varphi_{1,m},$$

and thus $d_{0,\ell} = \sum_{m\in\mathbb{Z}} \tilde{g}_{2\ell-m}a_{1,m} = \tilde{g} * a_1(2\ell)$. Here $\widetilde{g} = \{\widetilde{g}_k\}_{k\in\mathbb{Z}}$ is the conjugate flip of the high-pass filter g, which also has length L.

One can verify that for all $j \in \mathbb{Z}$, the scaling equation

$$\varphi_{j,\ell} = \sum_{m\in\mathbb{Z}} h_{m-2\ell}\varphi_{j+1,m} \tag{11.5}$$

and difference equation

$$\psi_{j,\ell} = \sum_{m\in\mathbb{Z}} g_{m-2\ell}\varphi_{j+1,m} \tag{11.6}$$

holds.

As a consequence of these equations, for all $j \in \mathbb{Z}$ we obtain that

$$\begin{aligned} a_{j,\ell} &= \sum_{n\in\mathbb{Z}} \tilde{h}_{2\ell-n}\, a_{j+1,n} = \widetilde{h} * a_{j+1}(2\ell), \\ d_{j,\ell} &= \sum_{n\in\mathbb{Z}} \tilde{g}_{2\ell-n}\, a_{j+1,n} = \widetilde{g} * a_{j+1}(2\ell). \end{aligned}$$

We have expressed each detail and approximation coefficient at level j as a convolution of the approximation coefficients at level $j+1$ and the conjugate flips of the filters h and g, respectively. Since these filters have length L, calculating each approximation and detail coefficient requires L multiplications. □

Exercise 11.16. Verify the scaling and difference equations (11.5) and (11.6). ♢

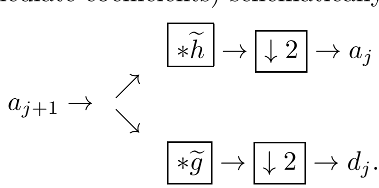
We can compute the approximation and detail coefficients at a rougher scale (j) by convolving the approximation coefficients at the finer scale ($j+1$) with the low- and high-pass filters $\widetilde{h}$ and $\widetilde{g}$ and *downsampling* by a factor of two, as follows.

Definition 11.17. The *downsampling operator* takes an N-vector (or a sequence) and maps it into a vector half as long (or another sequence) by discarding the odd-numbered entries:

$$Ds(n) := s(2n).$$

The downsampling operator is denoted by the symbol $\downarrow 2$. ♢

In electrical engineering terms, we have just described the *analysis phase* of a *subband filtering scheme*. We can represent the *analysis phase* (how to calculate coefficients) schematically as



Another useful operation is *upsampling*, which is the right inverse of downsampling.

Definition 11.18. The *upsampling operator* takes an N-vector (or a sequence) and maps it to a vector twice as long (or another sequence), by intertwining zeros:

$$Us(n) = \begin{cases} s(n/2), & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

The upsampling operator is denoted by the symbol $\uparrow 2$. ◊

Exercise 11.19. Compute the Fourier transform for the upsampling and downsampling operators in finite-dimensional space. ◊

Exercise 11.20. Verify that given a vector $s \in \mathbb{C}^N$, then $DUs = s$, but UDs is not always s. For which vectors s is $UDs = s$? Now answer the same question for sequences. ◊

Theorem 11.21. *The reconstruction of the samples at level $j+1$ from the averages and details at the previous level j is also an order L algorithm, where L is the length of the filters. If we wish to reconstruct, from averages at a coarser level M and all details from the coarser levels $M \le j < M+J$, all the averages at the finer scales $M < j \le M+J$, the algorithm will be of order LN where $N = 2^J$ is the number of samples which is also the number of total coefficients given.*

Proof. We show how to get from the coarse approximation and detail coefficients a_j and d_j to the approximation coefficients at the finer scale a_{j+1}. The calculation is based on the equation $P_{j+1}f = P_jf + Q_jf$, which is equivalent to

$$\sum_{m\in\mathbb{Z}} a_{j+1,m}\varphi_{j+1,m} = \sum_{\ell\in\mathbb{Z}} a_{j,\ell}\varphi_{j,\ell} + \sum_{\ell\in\mathbb{Z}} d_{j,\ell}\psi_{j,\ell}. \tag{11.7}$$

We have formulas that express $\varphi_{j,\ell}$ and $\psi_{j,\ell}$ in terms of the functions $\{\varphi_{j+1,m}\}_{m\in\mathbb{Z}}$; see Exercise 11.16. Inserting those formulas in the right-hand side (RHS) of equality (11.7) and collecting all terms that are multiples of $\varphi_{j+1,m}$, we get

$$\begin{aligned}\text{RHS} &= \sum_{\ell\in\mathbb{Z}} a_{j,\ell} \sum_{m\in\mathbb{Z}} h_{m-2\ell}\varphi_{j+1,m} + \sum_{\ell\in\mathbb{Z}} d_{j,\ell} \sum_{m\in\mathbb{Z}} g_{m-2\ell}\varphi_{j+1,m} \\ &= \sum_{m\in\mathbb{Z}} \Big[\sum_{\ell\in\mathbb{Z}} h_{m-2\ell}a_{j,\ell} + g_{m-2\ell}d_{j,\ell}\Big]\varphi_{j+1,m}.\end{aligned}$$

This calculation, together with equation (11.7), implies that

$$a_{j+1,m} = \sum_{\ell\in\mathbb{Z}} h_{m-2\ell}a_{j,\ell} + g_{m-2\ell}d_{j,\ell}.$$

Rewriting in terms of convolutions and upsamplings, we obtain

$$a_{j+1,m} = h * Ua_j(m) + g * Ud_j(m), \tag{11.8}$$

where we first upsample the approximation and detail coefficients to restore the right dimensions, then convolve with the filters, and finally add the outcomes. □

Exercise 11.22. Verify equation (11.8). ♢

The *synthesis phase* (how to reconstruct from the coefficients) of the *subband filtering scheme* can be represented schematically as

$$\begin{array}{ccccccc} a_j \to & \boxed{\uparrow 2} & \to & \boxed{*h} & & & \\ & & & & \searrow & & \\ & & & & & \oplus \to a_{j+1}. & \\ & & & & \nearrow & & \\ d_j \to & \boxed{\uparrow 2} & \to & \boxed{*g} & & & \end{array}$$

The process of computing the coefficients can be represented by the following tree, or *pyramid scheme* or *algorithm*:

$$\begin{array}{ccccccccc} a_j & \to & a_{j-1} & \to & a_{j-2} & \to & a_{j-3} & \cdots \\ & \searrow & & \searrow & & \searrow & & \\ & & d_{j-1} & & d_{j-2} & & d_{j-3} & \cdots. \end{array}$$

The reconstruction of the signal can be represented with another pyramid scheme as

$$\begin{array}{ccccccccc} a_j & \to & a_{j+1} & \to & a_{j+2} & \to & a_{j+3} & \cdots \\ & \nearrow & & \nearrow & & \nearrow & & \\ d_j & & d_{j+1} & & d_{j+2} & & \cdots . & \end{array}$$

Note that once the low-pass filter H is chosen, everything else—high-pass filter G, scaling function φ, wavelet ψ, and MRA—is completely determined. In practice one never computes the values of φ and ψ. All the manipulations are performed with the filters g and h, even if they involve calculating quantities associated to φ or ψ, like moments or derivatives. However, to help our understanding, we might want to produce pictures of the wavelet and scaling functions from the filter coefficients.

Example 11.23. (Drawn from Section 3.4.3 in [**MP**].) The cascade algorithm can be used to produce very good approximations for both ψ and φ, and this is how one usually obtains pictures of the wavelets and the scaling functions. For the scaling function φ, it suffices to observe that $a_{1,k} = \langle \varphi, \varphi_{1,k} \rangle = h_k$ and $d_{j,k} = \langle \varphi, \psi_{j,k} \rangle = 0$ for all $j \geq 1$ (the first because of the scaling equation, the second because $V_0 \subset V_j \perp W_j$ for all $j \geq 1$). That is what we need to initialize and iterate as many times as we wish (say n times) the synthesis phase of the filter bank:

$$h \to \boxed{\uparrow 2} \to \boxed{\; *h \;} \to \cdots \to \boxed{\uparrow 2} \to \boxed{\; *h \;} \to \{\langle \varphi, \varphi_{n+1,k} \rangle\}.$$

The output after n iterations is the set of approximation coefficients at scale $j = n + 1$. After multiplying by a scaling factor, one can make precise the statement that

$$\varphi(k2^{-j}) \sim 2^{-j/2} \langle \varphi, \varphi_{j,k} \rangle.$$

A graph can now be plotted (at least for real-valued filters). We illustrated the power of the cascade algorithm in Section 11.2, although the perspective there was a bit different from here.

We proceed similarly for the wavelet ψ. This time $a_{1,k} = \langle \psi, \varphi_{1,k} \rangle = g_k$ and $d_{j,k} = \langle \psi, \psi_{j,k} \rangle = 0$ for all $j \geq 1$. Now the cascade algorithm produces the approximation coefficients at scale j after $n = j - 1$ iterations:

$$g \to \boxed{\uparrow 2} \to \boxed{\;*h\;} \to \cdots \to \boxed{\uparrow 2} \to \boxed{\;*h\;} \to \{\langle \psi, \varphi_{j,k}\rangle\}.$$

This time

$$\psi(k2^{-j}) \sim 2^{-j/2}\langle \psi, \varphi_{j,k}\rangle,$$

and we can plot a reasonable approximation of ψ from these approximated samples. ♢

Exercise 11.24. Use the algorithm described in Example 11.23 to create pictures of the scaling and the wavelet functions corresponding to the Daubechies wavelets *db*2 after four iterations and after ten iterations. The coefficients for *db*2 are listed in Example 10.7. ♢

Altogether, we have a *perfect reconstruction filter bank*:

$$\begin{array}{ccccc}
 & & \boxed{*\widetilde{h}} \to \boxed{\downarrow 2} \to a_{j-1} \to \boxed{\uparrow 2} \to \boxed{*h} & & \\
a_j \to & \begin{matrix}\nearrow \\ \searrow\end{matrix} & & \begin{matrix}\searrow \\ \nearrow\end{matrix} & \oplus \to a_j. \\
 & & \boxed{*\widetilde{g}} \to \boxed{\downarrow 2} \to d_{j-1} \to \boxed{\uparrow 2} \to \boxed{*g} & &
\end{array}$$

Exercise 11.25 (*Human Filter Banks*)**.** Write the above filter bank on paper on a long table. Use the filter coefficients from the Daubechies *db*2 wavelets. Assign each convolution with a filter to a different person. Arrange the humans around the table next to their filters, and hand out pens. Choose a signal vector and feed it in to the filter bank. Can you analyse the signal and reconstruct it perfectly? ♢

The filter bank need not always involve the same filters in the analysis and synthesis phases. When it does, as in our case, we have an *orthogonal filter bank*. One can also obtain perfect reconstruction in more general cases.

Example 11.26 (*Biorthogonal Filter Bank*)**.** Replace the perfect reconstruction filter bank $\widetilde{h}$ by a filter $\widetilde{h^*}$, and $\widetilde{g}$ by $\widetilde{g^*}$. Then we obtain a *biorthogonal filter bank* with dual filters h, h^*, g, g^*. (Consider reviewing Section 6.2 on dual bases in the finite-dimensional setting.)

In the case of a biorthogonal filter bank, there is an associated *biorthogonal MRA* with *dual scaling functions* φ, φ^*. Under certain conditions we can find biorthogonal wavelets ψ, ψ^* that generate a *biorthogonal wavelet basis* $\{\psi_{j,k}\}$ and *dual basis* $\{\psi^*_{j,k}\}$, so that

$$f(x) = \sum_{j,k} \langle f, \psi^*_{j,k} \rangle \psi_{j,k} = \sum_{j,k} \langle f, \psi_{j,k} \rangle \psi^*_{j,k}.$$

The following substitute for Plancherel's Identity holds (*Riesz basis property*[3]): For all $f \in L^2(\mathbb{R})$,

$$\sum_{j,k} |\langle f, \psi^*_{j,k} \rangle|^2 \approx \|f\|_2^2 \approx \sum_{j,k} |\langle f, \psi_{j,k} \rangle|^2,$$

where $A \approx B$ means that there exist constants c, $C > 0$ such that $cA \leq B \leq CA$. The relative size of the similarity constants c, C becomes important for numerical calculations; it is related to the *condition number* of a matrix.

As in the case of orthogonal filters, one can find necessary and sufficient conditions on the dual filters to guarantee perfect reconstruction. Such conditions parallel the work done in Section 10.3.4. See [**MP**]. $\Diamond$

A general filter bank is a sequence of convolutions and other simple operations such as upsampling and downsampling. The study of such banks is an entire subject in engineering called *multirate signal analysis*, or *subband coding*. The term *filter* is used to denote a convolution operator because such an operator can cut out various frequencies if the corresponding Fourier multiplier vanishes (or is very small) at those frequencies. See Strang and Nguyen's book [**SN**].

Filter banks can be implemented quickly because of the Fast Fourier Transform. Remember that circular convolution becomes multiplication by a diagonal matrix on the Fourier side (needs only N products, whereas ordinary matrix multiplication requires N^2 products), and to go back and forth from the Fourier to the time domain, we can use the Fast Fourier Transform. So the total number of operations is of order $N \log_2 N$; see Chapter 6.

The fact that there is a large library of wavelet filter banks has made wavelets a fundamental tool for practitioners. Ingrid Daubechies

[3] Named after the Hungarian mathematician Frigyes Riesz (1880–1956).

is responsible for making a large number of filters available to engineers. She has received many honors and awards in her career. In particular she received the 2011 AMS Leroy P. Steele Prize for Seminal Contributions to Research for her paper [**Dau88**]. In her response (*Notices of the AMS*, Vol. 58, No. 4 (April 2011)), she says, "I would like to thank *Communications in Pure and Applied Mathematics*, where the paper appeared, for accepting those long tables of coefficients—its impact in engineering would not have been the same without the tables, at that time a standard feature of papers on filter reconstruction in signal analysis." In 1994 Ingrid Daubechies became the first female full professor of mathematics at Princeton. In 2010 she was elected as the first female President of the International Mathematical Union.

There are commercial and free software programs dealing with these applications. The main commercial one is the MATLAB *Wavelet Toolbox*. *Wavelab* is a free software program based on MATLAB. It was developed at Stanford. *Lastwave* is a toolbox with subroutines written in C, with "a friendly shell and display environment", according to Mallat. It was developed at the École Polytechnique. There is much more on the Internet, and you need only search on *wavelets* to see the enormous amount of information, codes, and so on that are available online.

11.4. A wavelet library

In this section we first describe the most common properties one would like wavelets to have. Second, we list some popular wavelets and highlight their properties.

11.4.1. Design features. Most of the applications of wavelets exploit their ability to approximate functions as efficiently as possible, that is, with as few coefficients as possible. For different applications one wishes the wavelets to have various other properties. Some of them are in competition with each other, so it is up to the user to decide which ones are most important for his or her problem. The most popular conditions are orthogonality, compact support, vanishing moments, symmetry, and smoothness. A similar list of competing attributes can be found in [**MP**, Section 3.4.2]. For a more extended discussion on these issues see [**Mall09**, Section 7.2.1].

Orthogonality. Orthogonality allows for straightforward calculation of the coefficients (via inner products with the basis elements). It guarantees that energy is preserved. Sometimes orthogonality is replaced by the weaker condition of *biorthogonality*. In this case, there is an auxiliary set of dual functions that is used to compute the coefficients by taking inner products; also the energy is almost preserved. See Example 11.26.

Compact support. We have stressed that compact support of the scaling function is important for numerical purposes. In most cases it corresponds to a finite low-pass filter, which can be implemented easily. If the low-pass filter is supported on $[n, m]$, then so is φ, and then the wavelet ψ has support of the same length $m - n$.

With regard to detecting point singularities, it is clear that if the signal f has a singularity at a point t_0 in the support of $\psi_{j,n}$, then the corresponding coefficient could be large. If ψ has support of length l, then at each scale j there are l wavelets interacting with the singularity (that is, whose support contains t_0). The shorter the support, the fewer wavelets will interact with the singularity.

Smoothness. The regularity of the wavelet affects the error introduced by thresholding, or quantizing, the wavelet coefficients. If an error ε is added to the coefficient $\langle f, \psi_{j,k}\rangle$, then an error of the form $\varepsilon\psi_{j,k}$ is added to the reconstruction. Smooth errors are often less *visible* or *audible*. Often better quality images are obtained when the wavelets are smooth. However, the smoother the wavelet, the longer the support. There is no orthogonal wavelet that is C^∞ and that has exponential decay. Therefore there is no hope of finding an orthogonal wavelet that is infinitely differentiable and that has compact support.

Vanishing moments. A function ψ *has* p *vanishing moments* if the integral of $x^k\psi(x)$ vanishes for all $k = 0, 1, \ldots, p-1$. Such a wavelet ψ is orthogonal to all polynomials of degree $p - 1$. If the function to be analyzed is k times differentiable, then it can be approximated well by a Taylor polynomial of degree k. If $k < p$, then the wavelets are orthogonal to that Taylor polynomial, and the coefficients are small.

If ψ has p vanishing moments, then the polynomials of degree $p-1$ are reproduced by the scaling functions.

If an orthogonal wavelet ψ has p vanishing moments, then its support is of length at least $2p-1$; see [**Mall09**, Theorem 7.9, p. 294]. (Daubechies wavelets have minimum support length for a given number of vanishing moments.) So there is a trade-off between the length of the support and the number of vanishing moments. If the function has few singularities and is smooth between singularities, then we might as well take advantage of the vanishing moments. If there are many singularities, we might prefer to use wavelets with shorter supports.

Symmetry. It is not possible to construct compactly supported symmetric orthogonal wavelets, except for the Haar wavelets. However, symmetry is often useful for image and signal analysis. Symmetry can be obtained at the expense of one of the other properties. If we give up orthogonality, then there are compactly supported, smooth, and symmetric biorthogonal wavelets. *Multiwavelets*[4] can be constructed to be orthogonal, smooth, compactly supported, and symmetric. Some wavelets have been designed to be nearly symmetric (Daubechies symmlets, for example).

11.4.2. A catalog of wavelets. We list the main wavelets and indicate their properties. This list was compiled from the first author's book [**MP**, mostly Sections 3.4.3 and 3.5], where you can also find references for many other friends, relatives, and mutations of wavelets.

Haar wavelet. We have studied these wavelets in detail. They are perfectly localized in time, less localized in frequency. They are discontinuous and symmetric. They have the shortest possible support and only one vanishing moment; hence they are not well adapted to approximating smooth functions.

Shannon wavelet. This wavelet does not have compact support, but it is infinitely differentiable. It is band-limited, but its Fourier transform is discontinuous, and hence $\psi(t)$ decays like $1/|t|$ at infinity.

[4]Multiwavelets have more than one scaling function and more than one wavelet function, whose dilates and translates form a basis in the MRA.

The Fourier transform $\widehat{\psi}(\xi)$ of the Shannon wavelet is zero in a neighborhood of $\xi = 0$, so all its derivatives are zero at $\xi = 0$, and by the time–frequency dictionary, ψ has infinitely many vanishing moments.

Meyer wavelet. This wavelet is a symmetric band-limited function whose Fourier transform is smooth; hence $\psi(t)$ has fast decay at infinity. The scaling function is also band-limited. Hence both ψ and φ are infinitely differentiable. The wavelet ψ has an infinite number of vanishing moments. (This wavelet was found by Strömberg in 1983 and went unnoticed for several years; see [**Woj03**] and [**Mall09**, p. 289].)

Battle–Lemarié spline wavelets[5]**.** These wavelets are *polynomial splines* of degree m, that is, C^{m-1} functions that are piecewise polynomials of degree at most m with nodes on integer multiples of a number $a > 0$. The wavelet ψ has $m + 1$ vanishing moments. They do not have compact support, but they have exponential decay. They are $m-1$ times continuously differentiable. For m odd, ψ is symmetric around 1/2. For m even, ψ is antisymmetric around 1/2. The linear spline wavelet is the Franklin wavelet. See [**Woj03**, Section 3.3] and [**Mall09**, p. 291].

Daubechies compactly supported wavelets. They have compact support, which is moreover of the minimum possible length for any given number of vanishing moments. More precisely, if ψ has p vanishing moments, then the wavelet ψ is supported on an interval of length $2p - 1$. In terms of taps, ψ has $2p$ taps. For large p, the functions φ and ψ are uniformly Lipschitz-α of order $\alpha \sim 0.2p$. They are asymmetric. When $p = 1$, we recover the Haar wavelet. See [**Mall09**, Section 7.3.2].

Daubechies symmlets. They have p vanishing moments. They have compact support, of the minimum possible length, namely $2p-1$. They have $2p$ taps. They are as symmetric as possible. See [**Mall09**, p. 294].

Coiflets[6]**.** The coiflet ψ has p vanishing moments and the corresponding scaling function φ has $p - 1$ vanishing moments (from the

[5]Named after the American mathematical physicist Guy Battle and the French mathematician Pierre-Gilles Lemarié-Rieusset.

[6]Named after the American mathematician Ronald Raphael Coifman (born 1941).

second to the p^{th} moment, but not the first since $\int \varphi = 1$). The minimum possible length of the support of such a ψ is $3p - 1$, and the coiflet has support of length $3p-1$. The vanishing moments of φ allow the approximation coefficients of a smooth function f to be close to the signal samples.

Mexican hat. This wavelet has a closed formula involving second derivatives of the Gaussian: $\psi(t) = C(1 - t^2)e^{-t^2/2}$, where the constant is chosen so that $\|\psi\|_2 = 1$. It does not come from an MRA, and it is not orthogonal. It is appropriate for the continuous wavelet transform. It has exponential decay but not compact support. See [**Mall09**, p. 180].

Morlet wavelet. The wavelet is given by the closed formula $\psi(t) = Ce^{-t^2/2}\cos(5t)$. It does not come from an MRA, and it is not orthogonal. It is appropriate for the continuous wavelet transform. It has exponential decay but not compact support.

Spline biorthogonal wavelets. These wavelets are compactly supported, symmetric, and continuous. There are two positive integer parameters N, N^*. The parameter N^* determines the scaling function φ^*, which is a spline of order $[N^*/2]$. The other scaling function and both wavelets depend on both parameters. The function ψ^* is a compactly supported piecewise polynomial of order $N^* - 1$, which is C^{N^*-2} at the knots and whose support gets larger as N increases. The wavelet ψ has support increasing with N and in addition has vanishing moments. Their regularity can differ notably. The filter coefficients are dyadic rationals, which makes them very attractive for numerical purposes. The functions ψ are known explicitly. The dual filters are very unequal in length, which could be a nuisance when doing image analysis, for example. See [**Mall09**, p. 271].

Wavelet packets. To perform the wavelet transform, we iterated at the level of the low-pass filter (approximation). In principle it is an arbitrary choice; one could iterate at the high-pass filter level or any desirable combination. The full dyadic tree gives an overabundance of information. It corresponds to the *wavelet packets*. Each finite wavelet packet has the information for reconstruction in many different bases, including the wavelet basis. There are fast algorithms to search for the *best basis*. The *Haar packet* includes the Haar basis and the *Walsh basis*[7]. A good source of information on this topic is [**Wic**].

The wavelet packet and *cosine packet* create large libraries of orthogonal bases, all of which have fast algorithms. The Fourier basis, the wavelet bases, and Gabor-like bases are examples in this time–frequency library. All these wavelets are encoded in MATLAB. We encourage the reader to review them and their properties online.

11.5. Project: Wavelets in action

Here are some applications of wavelets to a variety of subjects. Start your search simply by typing each item into your favorite search engine. There is also a wealth of information on the Wavelet Digest website `www.wavelet.org` and on Amara Graps's webpage: `http://www.amara.com/current/wavelet.html`.

(1) Solving partial differential equations: Galerkin's method.

(2) Analysis of Brownian motion.

(3) X-ray source detection: pixons and wavelets.

(4) Wavelets in medical imaging: tomography, electrocardiograms, identifying tumors. There is some background at `http://www.cs.brown.edu/stc/resea/director/research_dp3.html`.

(5) Audio applications (or *sound fun*) such as distinguishing Cremona violins, whale calls, wavelet records, listening for defects, tonebursts, and wavelets in electroacoustic music can be found at `http://www.amara.com/current/wavesoundfun.html`.

(6) Seismography. You could start your search at the Wavelet Seismic Inversion Lab at Colorado School of Mines: `http://timna.mines.edu/~zmeng/waveletlab/waveletlab.html`.

(7) Wavelets for artist identification: `http://www.pacm.princeton.edu/~ingrid/VG_swirling_movie/`.

(8) What are curvelets and diffusion wavelets? What are their uses?

[7]Named after the American mathematician Joseph Leonard Walsh (1895–1973).

Chapter 12

The Hilbert transform

This chapter is about a very important operator in harmonic analysis, the *Hilbert transform*[1] This operator was introduced by David Hilbert in 1905 in order to solve a special case of the Riemann–Hilbert problem[2] for holomorphic functions, in complex analysis. The Hilbert transform is a prototypical example for two general classes of operators: *Fourier multipliers* and *singular integral operators*, which arise in many areas of mathematics, including Fourier analysis, complex analysis, partial differential equations, and stochastic processes. Nowadays the Hilbert transform has applications in signal processing and chemistry as well as in mathematics; see the projects in Section 12.8 and Section 12.9.

We define the Hilbert transform in three equivalent ways: first via its action in the frequency domain as a Fourier multiplier (Section 12.1), second via its action in the time domain as a singular integral (Section 12.2), and third via its action in the *Haar domain* as an average of Haar shift operators (Section 12.3), taking advantage of its invariance properties. We present the mapping properties of the Hilbert transform and, in particular, what it does to functions in $L^p(\mathbb{R})$ (Section 12.4). We revisit boundedness properties in L^2 and

[1]Named after the same Hilbert as Hilbert spaces (Subsection 2.1.2).

[2]The Riemann–Hilbert problem: Given f defined on $\mathbb{R}$, find holomorphic functions F^+ and F^- defined on the upper and lower half-planes, respectively, such that $f = F^+ - F^-$. Then the Hilbert transform is given by $Hf = (1/i)(F^+ - F^-)$.

show how translation and dilation invariance properties imply that an operator is bounded from L^p into L^q only when $p = q$.

The upshot is that the Hilbert transform maps $L^p(\mathbb{R})$ onto itself when $1 < p < \infty$ but not when $p = 1$ or $p = \infty$. We briefly sketch the standard classical proof via the Marcinkiewicz Interpolation Theorem (which we will not state) and the original proof by M. Riesz via the Riesz–Thorin Interpolation Theorem, and we present the modern nonstandard argument to deduce the L^p boundedness of the Hilbert transform from the L^p boundedness of the Haar shift operators, bypassing interpolation theory (Section 12.4). We compare weak boundedness in L^1 with boundedness in L^1 for both the Hilbert transform and the maximal function and present the useful *Layer Cake Formula* for the L^p norm of a function in terms of the length of its level sets E_λ (Section 12.5).

We state the Riesz–Thorin Interpolation Theorem in Section 12.6. This powerful result leads to short proofs of some of the most beautiful and useful inequalities in analysis (Section 12.6.2). We have used these inequalities throughout the book: the Hausdorff–Young Inequality for the Fourier transform, Hölder's Inequality and a generalization, Young's Inequality for convolutions, Minkowski's Inequality (the Triangle Inequality for L^p), and Minkowski's Integral Inequality.

We close by outlining the connections of the Hilbert transform with complex analysis and with Fourier theory (Section 12.7), returning in particular to our book's theme of Fourier analysis.

Excellent accounts of the Hilbert transform and its generalizations can be found in [**SW71**, Chapter 6], [**Graf08**, Chapter 4], [**Sad**, Chapter 6], [**Duo**, Chapter 3], and [**Tor**, Chapter V].

12.1. In the frequency domain: A Fourier multiplier

In this section we describe the Hilbert transform by its action on the Fourier domain. There the Hilbert transform acts by multiplying the part of $\widehat{f}$ with positive frequencies ($\widehat{f}(\xi)$ for $\xi > 0$) by $-i$ and the part of $\widehat{f}$ with negative frequencies ($\widehat{f}(\xi)$ for $\xi < 0$) by i. Then the new function $\widehat{f}(\xi)+i\widehat{Hf}(\xi)$ effectively has no negative frequencies and twice as much as f does of the positive frequencies. This property may have been why Hilbert introduced the transform that now bears his name, in relation to the Riemann–Hilbert problem.

Definition 12.1. The *Hilbert transform* H is defined on the Fourier side by the formula

$$(Hf)^\wedge(\xi) := -i\,\operatorname{sgn}(\xi)\widehat{f}(\xi), \tag{12.1}$$

for f in $L^2(\mathbb{R})$. Here the *signum function* sgn is defined by $\operatorname{sgn}(\xi) := 1$ for $\xi > 0$, $\operatorname{sgn}(\xi) := -1$ for $\xi < 0$, and $\operatorname{sgn}(0) := 0$. ♢

Exercise 12.2. Show that $\widehat{f}(\xi) + i\widehat{Hf}(\xi)$ is equal to $2\widehat{f}(\xi)$ for $\xi > 0$ and zero for $\xi < 0$. ♢

Since the Fourier transform is a bijection on $L^2(\mathbb{R})$, formula (12.1) defines an operator on $L^2(\mathbb{R})$: the Hilbert transform of $f \in L^2(\mathbb{R})$ at ξ is the L^2 function whose Fourier transform is $-i\operatorname{sgn}(\xi)$ times the Fourier transform of f at ξ. Further, the Hilbert transform is an *isometry* on $L^2(\mathbb{R})$: it preserves L^2 norms, as we see by using Plancherel's Identity twice:

$$\|Hf\|_2 = \|(Hf)^\wedge\|_2 = \|-i\,\operatorname{sgn}(\cdot)\widehat{f}(\cdot)\|_2 = \|\widehat{f}\|_2 = \|f\|_2.$$

Definition 12.1 also implies that $H^2 = -I$, where $-I$ denotes the negative of the identity operator, since for all $\xi \neq 0$ we have

$$(H^2 f)^\wedge(\xi) = -i\,\operatorname{sgn}(\xi)(Hf)^\wedge(\xi) = (-i\,\operatorname{sgn}(\xi))^2\widehat{f}(\xi) = -\widehat{f}(\xi).$$

Exercise 12.3 ($\mathcal{F}^4$ *Is the Identity Operator*)**.** Let $\mathcal{F}$ denote the Fourier transform: $\mathcal{F}f := \widehat{f}$. Show that $\mathcal{F}^4$ is the identity operator on those spaces where the Fourier transform is a bijection, namely the Schwartz class $\mathcal{S}(\mathbb{R})$, the Lebesgue space $L^2(\mathbb{R})$, and the class $\mathcal{S}'(\mathbb{R})$ of tempered distributions. ♢

Exercise 12.4 (*Commutativity, Take 1*)**.** Show that the Hilbert transform commutes with translations and dilations and that it anticommutes with reflections, as follows.

(i) Let τ_h denote the *translation operator*: $\tau_h f(x) := f(x - h)$ for $h \in \mathbb{R}$. Then $\tau_h H = H\tau_h$.

(ii) Let δ_a denote the *dilation operator*: $\delta_a f(x) := f(ax)$ for $a > 0$. Then $\delta_a H = H\delta_a$.

(iii) Let $\widetilde{f}(x) := f(-x)$ denote the *reflection of f across the origin*. Then $\widetilde{Hf} = -H(\widetilde{f})$. ◊

The Hilbert transform is an example of a *Fourier multiplier*.

Definition 12.5. A *Fourier multiplier* is an operator T that is defined on the Fourier side, at least for smooth functions f, by multiplication of the Fourier transform $\widehat{f}$ of f by a function m:

$$(Tf)^\wedge(\xi) := m(\xi)\widehat{f}(\xi). \tag{12.2}$$

The function $m(\xi)$ is called the *symbol* of the operator. ◊

The symbol m_H of the Hilbert transform is the bounded function

$$m_H(\xi) = -i\ \mathrm{sgn}(\xi).$$

Exercise 12.6. Verify that the partial Fourier integral operator $S_R f$ (see the project in Section 7.8) and the Cesàro means $\sigma_R f$ on $\mathbb{R}$ (equation (8.3)) are Fourier multipliers, by finding their symbols. ◊

Exercise 12.7. Is the Fourier transform a Fourier multiplier? ◊

The results of Exercise 12.6 are no accident. Both S_R and σ_R are convolution operators: $Tf = K * f$. The time–frequency dictionary tells us that on the Fourier side $(Tf)^\wedge(\xi) = \widehat{K}(\xi)\widehat{f}(\xi)$. So as long as the function K is nice enough, the convolution operator will be a Fourier multiplier with multiplier $m(\xi) := \widehat{K}(\xi)$. Exercise 12.4(i) can be generalized as follows.

Exercise 12.8 (*Commutativity, Take 2*)**.** Let T_m be a Fourier multiplier with bounded symbol m. Show that T_m commutes with translations. For the converse, see Remark 12.23. ◊

How can we define the *periodic Hilbert transform* $H_P : L^2(\mathbb{T}) \to L^2(\mathbb{T})$? Given what we have learned so far, it is natural to define $H_P f$ on the Fourier side by requiring its Fourier coefficients to be $\pm i$ times the Fourier coefficients of f.

Definition 12.9. The *periodic Hilbert transform*, or *conjugate function*, is defined on the Fourier side for all $n \in \mathbb{Z}$ by

$$(H_P f)^\wedge(n) = -i\ \mathrm{sgn}(n)\widehat{f}(n). \quad ◊$$

This definition corresponds to the second periodization method introduced in Section 8.5.1 for the Poisson Summation Formula.

Exercise 12.10. For $s(t) := \sin(2\pi\theta)$ and $c(\theta) := \cos(2\pi\theta)$ on $[0,1]$, show that $H_P c(\theta) = s(\theta)$ and $H_P s(\theta) = -c(\theta)$. ♢

Exercise 12.11. Show that for $f(\theta) = \sum_{|n|\leq N} c_n e^{2\pi i n\theta}$, we have $H_P f(\theta) = -i\sum_{n=1}^{N} c_n e^{2\pi i n\theta} + i\sum_{n=-N}^{-1} c_n e^{2\pi i n\theta}$. ♢

12.2. In the time domain: A singular integral

We now develop the direct definition of the Hilbert transform in the time domain, given by formula (12.3) below.

Using the notation we have just introduced and the properties of the Fourier transform, we see that the Hilbert transform Hf of a function f in $L^2(\mathbb{R})$ is given by

$$Hf = (\widehat{Hf})^{\vee} = (m_H \widehat{f}\,)^{\vee}.$$

Our experience with the time–frequency dictionary suggests that in the time domain, at least formally, the Hilbert transform should be given by convolution with a kernel k_H:

$$Hf = (m_H \widehat{f}\,)^{\vee} = k_H * f,$$

where $k_H(x) = (m_H)^{\vee}(x)$. Here we have used the fact that the Fourier transform converts convolution into multiplication. The question is: what is the inverse Fourier transform of the symbol? For a bounded symbol such as m_H, the inverse Fourier transform can be calculated in the sense of distributions, and in fact k_H turns out to be a distribution and not a function. This heuristic works in general: all bounded Fourier multipliers correspond to convolution operators with distributional kernels [**Graf08**, Section 2.5].

One can compute the inverse Fourier transform of $(Hf)^{\wedge}$ explicitly, at least when f is a Schwartz function. It is given formally by the following principal value integral:

$$Hf(x) := \text{p.v.}\frac{1}{\pi}\int \frac{f(y)}{x-y}\,dy = \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|x-y|>\varepsilon}\frac{f(y)}{x-y}\,dy. \tag{12.3}$$

Note that $Hf = k_H * f(x)$ is given by convolution with the principal value distribution $k_H(x) := \text{p.v.}\,1/(\pi x)$, introduced in the project in Section 8.7.

We use formula (12.3) as the *definition* of the Hilbert transform of functions f in $L^2(\mathbb{R})$, not just of Schwartz functions.

Definition 12.12. The *Hilbert transform H* is defined by

$$Hf(x) := \lim_{\varepsilon\to 0} \frac{1}{\pi} \int_{|x-y|>\varepsilon} \frac{f(y)}{x-y}\,dy \quad \text{for } f \in L^2(\mathbb{R}). \qquad \Diamond$$

We justify this broader definition as follows. Consider the well-defined and nonsingular kernel

$$k_{\varepsilon,R}(y) := (1/\pi y)\chi_{\{y\in\mathbb{R}:\varepsilon<|y|<R\}}(y),$$

which lies in $L^1(\mathbb{R})\cap L^2(\mathbb{R})$ for $0<\varepsilon<R<\infty$. This function $k_{\varepsilon,R}(y)$ is a *truncated*, or *cut-off*, version of the kernel k_H; the values of k_H have been set to zero for $|y| \le \varepsilon$ and for $|y| \ge R$. Note that $(K_{\varepsilon,R})^\wedge(0) = 0$, since $K_{\varepsilon,R}$ is an odd function. We calculate the Fourier transform of $k_{\varepsilon,R}$ for $\xi \neq 0$:

$$\begin{aligned}
(k_{\varepsilon,R})^\wedge(\xi) &= \int_{\varepsilon<|y|<R} \frac{1}{\pi y} e^{-2\pi i y\xi}\,dy \\
&= \int_\varepsilon^R \frac{1}{\pi y} e^{-2\pi i y\xi}\,dy - \int_\varepsilon^R \frac{1}{\pi y} e^{2\pi i y\xi}\,dy \\
&= -2i \int_\varepsilon^R \frac{\sin(2\pi y\xi)}{\pi y}\,dy = -\frac{2i}{\pi}\int_{2\pi\varepsilon\xi}^{2\pi R\xi} \frac{\sin t}{t}\,dt \\
&= -i\,\operatorname{sgn}(\xi)\left[\frac{2}{\pi}\int_{2\pi\varepsilon|\xi|}^{2\pi R|\xi|} \frac{\sin t}{t}\,dt\right].
\end{aligned}$$

Notice that for each real $\xi \neq 0$,

$$\frac{2}{\pi}\int_{2\pi\varepsilon|\xi|}^{2\pi R|\xi|} \frac{\sin t}{t}\,dt \to \frac{2}{\pi}\int_0^\infty \frac{\sin t}{t}\,dt = 1$$

as $\varepsilon \to 0$ and $R \to \infty$. Therefore the truncated Hilbert transforms $H_{\varepsilon,R}$ on $L^2(\mathbb{R})$, given by the equivalent formulas

$$H_{\varepsilon,R}f = k_{\varepsilon,R} * f, \qquad (H_{\varepsilon,R}f)^\wedge(\xi) = (k_{\varepsilon,R})^\wedge(\xi)\widehat{f}(\xi),$$

can be seen to be uniformly bounded (in R and ε) for all f in $L^2(\mathbb{R})$, by Plancherel's Identity. Furthermore, in the limit as $R\to\infty$ and $\varepsilon \to 0$, the symbols $(k_{\varepsilon,R})^\wedge(\xi)$ converge pointwise to the symbol $m_H(\xi) = -i\,\mathrm{sgn}(\xi)$ of the Hilbert transform. Also, for f in $L^2(\mathbb{R})$, we have $H_{\varepsilon,R}f \to Hf$ where H is given by Definition 12.12. By the continuity of the Fourier transform, it follows that

$$(H_{\varepsilon,R}f)^\wedge \to m_H\widehat{f}, \quad \text{and so} \quad (Hf)^\wedge = m_H\widehat{f}.$$

Thus the two definitions of the Hilbert transform of an L^2 function, via its actions in the frequency domain and in the time domain, agree.

Recall that by definition $\int_{\mathbb{R}} := \lim_{R\to\infty}\int_{-R}^{R}$, so the symmetry at infinity is built into the definition of the improper integral. The symmetry around zero in the truncation provides the correct cancellation necessary to carry out the calculation above.

See Exercise 12.55 for a complex analysis proof of formula (12.3).

Exercise 12.13. Show that the Hilbert transform of the characteristic function of the interval $I = [a,b]$ is given by $H\chi_I(x) = (1/\pi)\log\big(|x-a|/|x-b|\big)$. ♢

Figure 12.1 on page 349 shows the graph of $H\chi_I$ for $I = [1,3]$.

A similar argument applies to the periodic Hilbert transform H_P defined on the Fourier side in Section 12.1. For a periodic, real-valued, continuously differentiable function f on $\mathbb{T}$ we can calculate the periodic Hilbert transform H_Pf in the time domain via the formula

$$H_Pf(\theta) = \text{p.v.}\frac{1}{\pi}\int_0^1 f(t)\cot\big(\pi(\theta - t)\big)\,dt. \tag{12.4}$$

The singularity on the diagonal $\theta = t$ is comparable to that of the Hilbert transform.

Exercise 12.14. Use the integral formula (12.4) to verify that the periodic Hilbert transform maps cosines into sines. ♢

The Hilbert transform is the prototypical example of a *Calderón–Zygmund singular integral operator*[3]. These operators are given by integration against a kernel $K(x,y)$: $Tf(x) = \text{p.v.}\int_{\mathbb{R}} K(x,y)f(y)\,dy$.

Near the diagonal $y = x$, Calderón–Zygmund kernels $K(x,y)$ grow like the Hilbert kernel $k_H(x-y)$, and therefore the integral must

[3]Named after the same Calderón as the Calderón reproducing formula in Section 9.3, and the Polish-born American mathematician Antoni Zygmund (1900–1992).

be understood in a principal-value sense. The cancellation encoded for the Hilbert transform in the symmetry of p.v. $1/x$ is replaced by an appropriate estimate for a Calderón–Zygmund operator on the gradient of the kernel, or on differences in the values of the kernel at certain pairs of points in $\mathbb{R}^2$.

Aside 12.15. A one-dimensional *standard Calderón–Zygmund kernel* $K(x, y)$ is a function on $\mathbb{R}^2\backslash\{(0, 0)\}$ such that $|K(x, y)| \leq C/|x - y|$, and there exists $\delta > 0$ such that $|K(x, y) - K(x, z)| \leq C|y - z|^\delta/|x - y|^{1+\delta}$ if $|x-y| > 2|y-z|$ and $|K(x, y)-K(w, y)| \leq C|x-w|^\delta/|x-y|^{1+\delta}$ if $|x-y| > 2|x-w|$. These are excellent references on singular integral operators: [**Duo**], [**Graf08**], [**Graf09**], [**Sad**], [**Ste70**], and [**Tor**]. $\Diamond$

Exercise 12.16. Show that the kernel $K_H(x, y) := k_H(x - y)$ of the Hilbert transform is a standard kernel. $\Diamond$

12.3. In the Haar domain: An average of Haar shifts

We first recall the dyadic intervals and the associated Haar basis and introduce the *random dyadic grids*. We recall some important properties of the Haar basis, shared with wavelet bases, such as being an unconditional system in each L^p space (Section 9.5.2). We then describe Petermichl's averaging method and give some intuition for why it works. This section is based on the survey article [**Per12**].

12.3.1. Dyadic intervals and random dyadic grids. The *standard dyadic grid* $\mathcal{D}$ is the collection of intervals in $\mathbb{R}$ of the form $[k2^{-j}, (k + 1)2^{-j})$, for all integers k, $j \in \mathbb{Z}$. They are organized by generations: $\mathcal{D} = \bigcup_{j\in\mathbb{Z}} \mathcal{D}$, where $I \in \mathcal{D}$ if and only if $|I| = 2^{-j}$. They satisfy the trichotomy, or nestedness, property: if I, $J \in \mathcal{D}$, then exactly one of the three conditions $I \cap J = \emptyset$, $I \subseteq J$, or $J \subset I$ holds. If $I \in \mathcal{D}_j$ then there is a unique interval $\widetilde{I} \in \mathcal{D}_{j-1}$, the parent of I, such that $I \subset \widetilde{I}$, and there are two intervals I_l, $I_r \in \mathcal{D}_{j+1}$, the left and right children of I, such that $I = I_l \cup I_r$. The special point 0 does not lie in the interior of any interval $I \in \mathcal{D}$, and if $0 \in I$, then 0 is the left endpoint of I. See Section 9.4.

A *dyadic grid* in $\mathbb{R}$ is a collection of intervals, organized in generations, each generation being a partition of $\mathbb{R}$, that have the trichotomy, unique-parent, and two-children-per-interval properties. For example, translates and rescalings of the standard dyadic grid are dyadic grids. We can obtain more general dyadic grids by abolishing the special rôle of the point 0, which we do by an ingenious translation of certain generations.

The following parametrization captures all dyadic grids. Consider a *scaling parameter* or *dilation parameter* r with $1 \leq r < 2$ and a *random parameter* β with $\beta = \{\beta_i\}_{i\in\mathbb{Z}}$, $\beta_i \in \{0,1\}$. Let $x_j := \sum_{i<-j} \beta_i 2^i$. Define

$$\mathcal{D}_j^\beta := x_j + \mathcal{D}_j \qquad \text{and} \qquad \mathcal{D}_j^{r,\beta} := r\mathcal{D}_j^\beta.$$

The family of intervals $\mathcal{D}^{r,\beta} := \bigcup_{j\in\mathbb{Z}} \mathcal{D}_j^{r,\beta}$ is a *random dyadic grid.*

Once we have identified an interval in a dyadic grid, its descendants are completely determined: simply subdivide. However there are two possible choices for the parent, since the original interval can be the left or right half of the parent. The parameter β captures all these possibilities. For the standard dyadic grid, zero is never an interior point of a dyadic interval, and it is always the left endpoint of the dyadic intervals it belongs to. If we translate $\mathcal{D}$ by a fixed number, it will simply shift zero by that amount, and the translated point will still have this singular property. These translated grids correspond, in our parametrization, to parameters β such that β_j is constant for all sufficiently large j. Those β that do not eventually become constant eliminate the presence of a singular point such as zero in the standard grid; in fact most grids are of this type (Exercise 12.18).

Exercise 12.17. Show that if $1 \leq r < 2$ and $\beta \in \{0,1\}^{\mathbb{Z}}$, then the collection of random grids indexed by r and β includes every possible dyadic grid on $\mathbb{R}$. ♢

Exercise 12.18. What is the probability that a sequence β becomes eventually constant? This question is reminiscent of counting the rational numbers and the real numbers. ♢

Random dyadic grids were introduced by Nazarov, Treil, and Volberg[4] in 1997 in their study of Calderón–Zygmund singular integrals

[4]Russian mathematicians, active in the US, Fedor Nazarov (born 1967), Sergei Treil (born 1961), and Alexander Volberg (born 1966).

on nonhomogeneous spaces [**NTV97**] and are used by Hytönen[5] in his beautiful 2010 representation theorem for all Calderón–Zygmund operators in terms of averages of Haar shift operators [**Hyt08, Hyt12**]. This is an active area of research mathematics today.

One advantage of the parametrization above is that there is a natural probability space (Ω, P) associated to the parameters r and β, and averaging here means calculating the expectation $\mathbb{E}f = \int_\Omega f\,dP$ in this probability space.

12.3.2. Haar basis revisited. The Haar function h_I associated to an interval I is defined by $h_I(x) := (1/\sqrt{|I|})\big(\chi_{I_r}(x) - \chi_{I_l}(x)\big)$, where $\chi_I(x) = 1$ if $x \in I$ and zero otherwise. Note that $\|h_I\|_2 = 1$ and $\int h_I = 0$. We showed in Section 9.4 that the family $\{h_I\}_{I\in\mathcal{D}}$ is an orthonormal basis for $L^2(\mathbb{R})$.

In 1910, Alfred Haar introduced the Haar basis for $L^2([0,1])$ and showed that the Haar expansion of each continuous function converges uniformly [**Haa**], unlike the expansions in the trigonometric (Fourier) basis [**Duo**], [**Graf08**], [**Ste70**]. Furthermore, $\{h_I\}_{I\in\mathcal{D}}$ is an unconditional basis for $L^p(\mathbb{R})$.

Definition 12.19. A basis for $L^p(\mathbb{R})$ is called *unconditional* if changes in the signs of the coefficients of a function create a new function in $L^p(\mathbb{R})$ whose L^p norm is bounded by a constant times the L^p norm of the original function. (See Theorem A.35(iii).) $\Diamond$

In terms of the *martingale transform* T_σ (Section 9.5.2) given by

$$T_\sigma f(x) = \sum_{I\in\mathcal{D}} \sigma_I \langle f, h_I\rangle h_I, \qquad \text{where} \quad \sigma_I = \pm 1, \tag{12.5}$$

unconditionality of the Haar basis in $L^p(\mathbb{R})$ is equivalent to the martingale transform being bounded on $L^p(\mathbb{R})$ with norm independent of the choice of signs: $\sup_\sigma \|T_\sigma f\|_p \leq C_p\|f\|_p$. This boundedness is well known [**Graf08**], [**Ste70**]. Burkholder[6] found the optimal constant $C_p = p^* - 1$, where $p^* = \max\{p-1, 1/(p-1)\}$ [**Burkh**].

The trigonometric system $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ does not form an unconditional basis for $L^p([0,1))$, except when $p = 2$ [**Woj91**], [**Zyg59**].

---

[5] Finnish mathematician Tuomas Hytönen (born 1980).

[6] US mathematician Donald L. Burkholder (born 1927), known for his contributions to probablity theory.

Historically, the Haar basis is the first example of a wavelet basis, meaning a basis $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ that is given by appropriately translating and dilating a fixed *wavelet function* ψ: the basis functions are $\psi_{j,k}(x) = 2^{-j/2}\psi(2^j x + k)$. The Haar functions are translates and dyadic dilates of the fixed function $h(x) := \chi_{[0,1/2)}(x) - \chi_{[1/2,1)}(x)$. The unconditionality properties are shared by a large class of wavelet bases (Section 9.5.2).

12.3.3. Petermichl's dyadic shift operator. Petermichl's shift operator associated to the standard dyadic grid $\mathcal{D}$ is denoted by the Russian letter Ш, pronounced "sha", and is defined for $f \in L^2(\mathbb{R})$ by

$$Шf(x) := \sum_{I\in\mathcal{D}} \langle f, h_I\rangle H_I(x), \quad \text{where} \quad H_I = (1/\sqrt{2})(h_{I_r} - h_{I_l}).$$

Petermichl's shift operator is an isometry in $L^2(\mathbb{R})$, meaning that it preserves the L^2 norm: $\|Шf\|_2 = \|f\|_2$ (Section 9.5.2).

Notice that if $I \in \mathcal{D}$, then $Шh_I(x) = H_I(x)$. What is $ШH_I$? The periodic Hilbert transform H_p maps sines into negative cosines. Sketch the graphs of h_I and H_I, and you will see them as a localized squared-off sine and negative cosine. Thus the shift operator Ш may be a good dyadic model for the Hilbert transform. More evidence comes from the way Ш interacts with translations, dilations, and reflections.

Denote by $Ш_{r,\beta}$ Petermichl's shift operator associated to the dyadic grid $\mathcal{D}^{r,\beta}$. The individual shift operators $Ш_{r,\beta}$ do not commute with translations and dilations, nor do they anticommute with reflections. However, the following symmetries for the family of shift operators $\{Ш_{r,\beta}\}_{(r,\beta)\in\Omega}$ hold.

(i) Translation ($h \in \mathbb{R}$): $\tau_h(Ш_{r,\beta}f) = Ш_{r,\tau_{-(h/r)}\beta}(\tau_h f)$.

(ii) Dilation ($a > 0$): $\delta_a(Ш_{r,\beta}f) = Ш_{\delta_a}(r,\beta)(\delta_a f)$.

(iii) Reflection: $(Ш_{r,\beta}f)^{\sim} = Ш_{r,\widetilde{\beta}}(\widetilde{f})$, where $\widetilde{\beta}_i := 1 - \beta_i$.

In item (i) $\tau_h\beta$ denotes a new random parameter in $\{0,1\}^{\mathbb{Z}}$. In item (ii) $\delta_a(r,\beta)$ denotes a new pair $(r',\beta') \in [1,2) \times \{0,1\}^{\mathbb{Z}}$ of dilation and random parameters that index the random dyadic grids and the sha operator. In Exercise 12.20 we ask you to find these new parameters so that properties (i) and (ii) actually hold.

Note that the operations of translation, dilation, and reflection not only pass to the argument f but also affect the translation and/or dilation parameters (r, β).

Exercise 12.20. Make sense of these invariance properties. For (i), consider the cases $h \geq 0$ and $h < 0$ in order to define $\tau_{-(h/r)}\beta$ properly in terms of the binary expansion of h/r when $h \geq 0$ and of h^*/r when $h < 0$, where $h^* = 2^k + h > 0$ and $2^{k-1} \leq |h| < 2^k$. For (ii), consider the cases $2^j \leq ar < 2^{j+1}$ for each $j \in \mathbb{Z}$ and for $a > 0$ to define $\delta_a(r, \beta)$. ◊

The Hilbert transform commutes with translations and dilations and anticommutes with reflections. Moreover these invariance properties characterize the Hilbert transform up to a multiplicative constant (see Exercise 12.24).

The key point in Petermichl's proof is that although the individual dyadic shift operators do not have the symmetries that characterize the Hilbert transform, one can show that the average over all dyadic grids does have these symmetries, yielding the following result.

Theorem 12.21 (Petermichl [**Pet**], [**Hyt08**]). *The average of the dilated and translated dyadic shift operators is a nonzero multiple of the Hilbert transform: there is a constant $c \neq 0$ such that*

$$\mathbb{E}Ш_{r,\beta} = \int_\Omega Ш_{r,\beta} dP(r, \beta) = cH.$$

A similar approach works for the *Beurling–Ahlfors transform*[7] [**DV**] and the *Riesz transforms* [**PTV**]. Sufficiently smooth one-dimensional Calderón–Zygmund convolution operators are averages of Haar shift operators of bounded complexity [**Vag**]. Finally Hytönen's representation theorem (which builds on the recent work of many researchers) for all Calderón–Zygmund singular integral operators as averages of Haar shift operators of arbitrary complexity is the jewel in the crown [**Hyt12**]. Here, a shift operator that pushes Haar coefficients up by m generations and down by n generations is said

[7] Named after the Swedish mathematician Arne Carl-August Beurling (1905–1986) and the Finnish mathematician and Fields Medalist Lars Valerian Ahlfors (1907–1996).

to have *complexity* (m, n). For example, Petermichl's Ш is a Haar shift operator of complexity $(0, 1)$, while the martingale transform is a Haar shift operator of complexity $(0, 0)$.

12.4. Boundedness on L^p of the Hilbert transform

In this section we discuss the boundedness properties of the Hilbert transform on $L^p(\mathbb{R})$, for $1 < p < \infty$. We revisit boundedness on $L^2(\mathbb{R})$ for the Hilbert transform and for Fourier multipliers with bounded symbols. We show how translation and dilation invariance properties restrict the values of p and q for which the Hilbert transform can be bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$ to the case $p = q$. We state Riesz's[8] Theorem: the Hilbert transform is bounded on $L^p(\mathbb{R})$ for all $1 < p < \infty$. We sketch three proofs: first, Riesz's original 1927 proof relying on the *Riesz–Thorin Interpolation Theorem*[9] (stated in Section 12.6); second, the standard proof relying on the *Marcinkiewicz Interpolation Theorem*[10] and the *Calderón–Zygmund decomposition*; third, the twenty-first-century proof, relying on the representation of the Hilbert transform as an average of Haar shift operators (see the project in Section 12.9.4) and Minkowski's Integral Inequality (Section 12.6).

12.4.1. Boundedness on L^2 revisited. The Hilbert transform is an isometry on $L^2(\mathbb{R})$: $\|Hf\|_2 = \|f\|_2$. In particular, there is a positive constant C such that for all $f \in L^2(\mathbb{R})$,

$$\|Hf\|_2 \le C\|f\|_2.$$

In other words, the operator H is bounded on $L^2(\mathbb{R})$.

[8]Named after the Hungarian mathematician Marcel Riesz (1886–1969), younger brother of mathematician Frigyes Riesz.

[9]Named after Marcel Riesz and the Swedish mathematician G. Olof Thorin (1912–2004), who was Riesz's student.

[10]Named after Polish mathematician Józef Marcinkiewicz (1910–1940).

The Fourier multiplier T_m with symbol $m \in L^\infty(\mathbb{R})$ is a bounded operator on $L^2(\mathbb{R})$. Indeed, using Plancherel's Identity twice,

$$\|T_m f\|_2 = \|\widehat{T_m f}\|_2 = \Big(\int_{\mathbb{R}} |m(\xi)\widehat{f}(\xi)|^2\, d\xi\Big)^{1/2}$$
$$\leq \|m\|_\infty \|\widehat{f}\|_2 = \|m\|_\infty \|f\|_2.$$

The converse is also true: given a Fourier multiplier T_m bounded on $L^2(\mathbb{R})$, its symbol m must be in $L^\infty(\mathbb{R})$. The proof depends on the properties of essentially bounded functions; see [**Graf08**, Section 2.5] or [**Duo**, Chapter 3, Section 5]. Furthermore, the operator norm[11] from $L^2(\mathbb{R})$ into itself of the Fourier multiplier T_m is the L^∞ norm of its symbol $m \in L^\infty(\mathbb{R})$. The corresponding result in the periodic setting is proved in Lemma 12.58.

Exercise 12.22 (*The Norms of an Operator and of Its Symbol*)**.** Verify that if $m \in L^\infty(\mathbb{R})$, then $\|T_m\|_{L^2 \to L^2} = \|m\|_\infty$. ♢

Remark 12.23. Every bounded linear operator T on $L^2(\mathbb{R})$ that commutes with translations is given by convolution with a tempered distribution whose Fourier transform is a bounded function. It follows that such an operator is a Fourier multiplier with bounded multiplier. (See [**SS05**, Problem 5, p. 260], [**Graf08**, Section 2.5].) ♢

Exercise 12.24 (*Commutativity, Take 3*)**.** Let T be a bounded operator on $L^2(\mathbb{R})$ that commutes with translations and dilations. Show that if T commutes with reflections: $(Tf)^\sim = T(\widetilde{f})$ for all $f \in L^2(\mathbb{R})$, then T is a constant multiple of the identity operator: $T = cI$ for some $c \in \mathbb{R}$. Show that if T anticommutes with reflections: $(Tf)^\sim = -T(\widetilde{f})$ for all $f \in L^2(\mathbb{R})$, then T is a constant multiple of the Hilbert transform: $T = cH$ for some $c \in \mathbb{R}$ [**Ste70**, Chapter 2]. ♢

12.4.2. Boundedness from L^p into L^q. Is the Hilbert transform bounded on $L^p(\mathbb{R})$ for $p \neq 2$? Or more generally, is it bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$? We first define what it means for an operator to be bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$.

[11] If T is a bounded operator in $L^2(\mathbb{R})$, its *operator norm* $\|T\|_{L^2 \to L^2} := \inf\{C > 0 : \|Tf\|_2 \leq C\|f\|_2 \text{ for all } f \in L^2(\mathbb{R})\}$.

Definition 12.25. For p and q with $1 \le p \le \infty$ and $1 \le q \le \infty$, we say that an *operator T is bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$* if the operator $T : L^p(\mathbb{R}) \to L^q(\mathbb{R})$ and if there is a constant C such that

$$\|Tf\|_q \le C\|f\|_p \qquad \text{for all } f \in L^p(\mathbb{R}). \qquad \Diamond$$

Exercise 12.26 (*For Linear Operators Boundedness Is Equivalent to Continuity at Zero*)**.** Let T be a linear operator that maps $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$. Show that T is a bounded operator if and only if T is continuous at $f \equiv 0$. In other words, a linear operator T is bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$ if and only if $\|Tf_n\|_q \to 0$ for all sequences $\{f_n\}_{n\in\mathbb{N}} \subset L^p(\mathbb{R})$ such that $\|f_n\|_p \to 0$ as $n \to \infty$. $\Diamond$

The boundedness properties of the Fourier transform on $L^p(\mathbb{R})$ are summarized in Table 8.3. In particular, the Fourier transform is a bounded mapping from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$, provided that $1 \le p \le 2$ and that p, q are dual exponents: $1/p + 1/q = 1$. Furthermore, the *Hausdorff–Young Inequality*[12] holds for all $f \in L^p(\mathbb{R})$:

$$\|\widehat{f}\|_q \le \|f\|_p \quad \text{for } \ 1/p + 1/q = 1, \quad 1 \le p \le 2.$$

In other words, for all such p and q the L^q norm of the Fourier transform of f is bounded by the L^p norm of f. See Corollary 12.42.

What about the boundedness properties of the Hilbert transform? If a dilation invariant linear operator is bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$, then q must be equal to p (Exercise 12.27). The Hilbert transform is a dilation invariant linear operator. Thus it can only be bounded from $L^p(\mathbb{R})$ into itself and not from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$ when $p \ne q$.

Exercise 12.27 (*Dilation Invariant Operators from L^p to L^q*)**.** For $a > 0$ and f in $L^p(\mathbb{R})$, let $\delta_a f(x) := f(ax)$. Show that $\|\delta_a f\|_p = a^{-1/p}\|f\|_p$. Let T be a dilation invariant linear operator that is bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$. Check that for all $a > 0$, $\|Tf\|_q \le a^{1/q-1/p}C\|f\|_p$, where $C = \|T\|_{L^p(\mathbb{R})\to L^q(\mathbb{R})}$. Conclude that $p = q$. $\Diamond$

It is also known that if a translation invariant linear operator T is bounded from $L^p(\mathbb{R})$ to $L^q(\mathbb{R})$, then p cannot be greater than q. See Exercise 12.28 and [**Duo**, Section 6.7, p. 66].

---

[12] Named after the German mathematician Felix Hausdorff (1858–1942) and the English mathematician William Henry Young (1863–1942).

Exercise 12.28 (*Translation Invariant Operators from L^p to L^q*). For f in $L^p(\mathbb{R})$, show that $\lim_{h\to\infty} \|f + \tau_h f\|_p = 2^{1/p}\|f\|_p$. Let T be a translation invariant linear operator that is bounded from L^p into L^q. Check that $\|Tf + \tau_h Tf\|_q \leq C\|f + \tau_h f\|_p$, where $C = \|T\|_{L^p\to L^q}$. Let $h \to \infty$ to obtain $\|Tf\|_q \leq 2^{1/p-1/q}C\|f\|_p$. Conclude that $p \leq q$. **Hint:** First consider functions f of compact support. ♢

12.4.3. The Riesz Theorem. Since the Hilbert transform is a linear operator that is both dilation invariant and translation invariant, it could only be bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$ if $p = q$. But *is* it bounded on $L^p(\mathbb{R})$? The answer is *yes* for $1 < p < \infty$ and *no* for $p = 1$ or ∞; see Exercise 12.31.

On the time domain, the Hilbert transform is given by convolution with the distributional kernel $k_H(x) = \text{p.v. } 1/(\pi x)$, which is not an integrable function. If the kernel were an integrable function $k \in L^1(\mathbb{R})$, then the *convolution operator* T given by convolution against the kernel k,

$$Tf(x) := k * f(x) = \int_{\mathbb{R}} k(x-y)f(y)\,dy,$$

would automatically be bounded on $L^p(\mathbb{R})$ for all p with $1 \leq p \leq \infty$, by *Young's Inequality*[13]: $\|f * k\|_p \leq \|k\|_1\|f\|_p$. (We prove Young's Inequality as Corollary 12.52.) Unfortunately, *the Hilbert transform kernel k_H is not integrable* for two reasons: the integral of k_H on any interval containing zero is infinite, and the integral of k_H near infinity is also infinite.

Even so, the Hilbert transform is bounded on $L^p(\mathbb{R})$ for all p with $1 < p < \infty$. (Although it is not bounded on either $L^1(\mathbb{R})$ or $L^\infty(\mathbb{R})$, it satisfies a weaker boundedness property on each of these two spaces, as we will see below for $p = 1$.) We state the precise boundedness result for $1 < p < \infty$, and we sketch three proofs: the classical proof, the original proof, and the modern dyadic proof.

Theorem 12.29 (Riesz's Theorem, 1927). *The Hilbert transform is a bounded operator on $L^p(\mathbb{R})$ for all p with $1 < p < \infty$: for each such p there is a constant $C_p > 0$ such that for all $f \in L^p(\mathbb{R})$,*

$$\|Hf\|_p \leq C_p\|f\|_p.$$

[13] Also named after William Henry Young.

The best constant C_p in Riesz's Theorem is $C_p = \tan(\pi/(2p))$ when $1 < p \leq 2$ (found by Pichorides[14] in 1972) and $C_p = \cot(\pi/(2p))$ when $2 < p < \infty$ (found by Grafakos[15] in 2004).

The classical proof. The classical argument uses the facts that the Hilbert transform is bounded on $L^2(\mathbb{R})$ and *weakly bounded* on $L^1(\mathbb{R})$, followed by an *interpolation argument* to capture the range $1 < p < 2$ and then a *duality argument* for the range $p > 2$. □

Interpolation theorems tell us that if a linear operator is bounded or weakly bounded at each of two endpoints, then it is automatically bounded at intermediate points. The interpolation result used above is the *Marcinkiewicz Interpolation Theorem*, which says that if an operator is weakly bounded on L^p and on L^q, then it is bounded on L^r for all r between p and q. This is a powerful *convexity* result about bounded or weakly bounded operators on L^p spaces. See for example the books [**Graf08**, Chapter 1], [**Sad**, Chapter 4], [**Duo**, Chapter 2], and [**Tor**, Chapter IV]. We explain in Section 12.5 what it means to be weakly bounded on L^1.

Riesz's proof. When M. Riesz proved the boundedness on $L^p(\mathbb{T})$ of the periodic Hilbert transform in 1927, the Marcinkiewicz Interpolation Theorem had not yet appeared, and the the weak boundedness properties of the periodic Hilbert transform were not yet known. Using a complex-variables argument, M. Riesz showed that the periodic Hilbert transform is bounded on $L^p(\mathbb{T})$ for even numbers $p = 2k$, $k \geq 1$. Then by Riesz–Thorin interpolation (here the endpoint initial results needed were boundedness on $L^{2k}(\mathbb{T})$ and on $L^{2(k+1)}(\mathbb{T})$), which was already known, he obtained boundedness on $L^p(\mathbb{T})$ for p in the range $2 \leq p < \infty$. Finally, duality takes care of the range $1 < p < 2$. See [**Pin**, Section 3.3.1], [**Graf08**, Section 4.1.3]. □

The modern dyadic proof. Estimates for H follow from uniform estimates for Petermichl's shift operators introduced in Section 9.5.2 and again in Section 12.3.3.

---

[14] Cretan mathematician Stylianos Pichorides (1940–1992).

[15] Greek mathematician Loukas Grafakos (born 1962).

The case $p = 2$ follows from orthonormality of the Haar basis. First rewrite Petermichl's shift operator as follows (see Exercise 9.60), where $\widetilde{I}$ is the parent of I in the dyadic grid $\mathcal{D}^{r,\beta}$:

$$Ш_{r,\beta} f = \sum_{I \in \mathcal{D}_{r,\beta}} \operatorname{sgn}(I, \widetilde{I}) \langle f, h_{\widetilde{I}} \rangle h_I / \sqrt{2}.$$

Here $\operatorname{sgn}(I, \widetilde{I}) := -1$ if I is the left child of $\widetilde{I}$ and 1 if I is the right child. We can now use Plancherel's Identity to compute the L^2 norm. Noticing that each parent has two children, we see that

$$\|Ш_{r,\beta} f\|_2^2 = \sum_{I \in \mathcal{D}_{r,\beta}} |\langle f, h_{\widetilde{I}} \rangle|^2 / 2 = \|f\|_2^2.$$

Minkowski's Integral Inequality (Lemma 12.49) for measures then shows that $\|\mathbb{E} Ш_{r,\beta} f\|_2 \leq \mathbb{E} \|Ш_{r,\beta} f\|_2 \leq \|f\|_2$.

For the case $p \neq 2$, first check that

$$\sup_{r,\beta} \|Ш_{r,\beta} f\|_p \leq C_p \|f\|_p. \tag{12.6}$$

Then use Minkowski's Integral Inequality for measures (which holds for all $p \geq 1$, not just $p = 2$) to get $\|\mathbb{E} Ш_{r,\beta} f\|_p \leq \mathbb{E} \|Ш_{r,\beta} f\|_p \leq \|f\|_p$.

Inequality (12.6) follows from the boundedness of the dyadic square function in $L^p(\mathbb{R})$; see Section 9.5.2. □

12.5. Weak boundedness on $L^1(\mathbb{R})$

We introduce the idea of an operator being *weakly bounded* on $L^1(\mathbb{R})$, using some concepts from measure theory. Bounded operators are weakly bounded. However some operators, such as the Hilbert transform and the *Hardy–Littlewood maximal function*, are weakly bounded but not bounded; we discuss both these examples. We also establish the *Layer Cake Representation* for the L^p norm of a function in terms of the length, or *measure*, of its level sets.

12.5.1. Bounded vs. weakly bounded on $L^1(\mathbb{R})$. We begin with the definition of weak boundedness. The notation $m(E)$ stands for the *measure of a set* E; it is explained below.

Definition 12.30. An operator T is *weakly bounded on* $L^1(\mathbb{R})$ if there is a constant $C \geq 1$ such that for all $f \in L^1(\mathbb{R})$,

$$m(\{x \in \mathbb{R} : |Tf(x)| > \lambda\}) \leq C\|f\|_1/\lambda. \tag{12.7}$$

In other words, the size of the set where $|Tf|$ is bigger than λ is inversely proportional to λ, with a constant depending linearly on the L^1 norm of f. We often say T is of *weak-type* $(1, 1)$. ◊

For $f \in L^1(\mathbb{R})$ and $\lambda > 0$, define the λ *level set* $E_\lambda(f)$ of f to be the set of points $x \in \mathbb{R}$ where the absolute value $|f(x)|$ is larger than λ: $E_\lambda(f) := \{x \in \mathbb{R} : |f(x)| > \lambda\} \subset \mathbb{R}$.

If f belongs to $L^1(\mathbb{R})$, then the level sets $E_\lambda(f)$ are *measurable sets*. This technical term means that we can assign a *length*, or *measure*, to each such set. The measure can be thought of as a map from the measurable subsets of $\mathbb{R}$ to nonnegative numbers, which is additive for finite collections of disjoint sets and has other convenient properties. We choose to use *Lebesgue measure*, which has the pleasing property that the Lebesgue measure $m([a, b])$ of each interval $[a, b]$ coincides with the ordinary length $|b-a|$ of $[a, b]$. The symbol $m(A)$ is read as *the measure of the set* A. Also, $m(A) = \int_A 1\, dx = \int_{\mathbb{R}} \chi_A(x)\, dx$, so that the Lebesgue measure $m(A)$ of a measurable set A is equal to the number given by the Lebesgue integral over $\mathbb{R}$ of the characteristic function χ_A. To delve into the many technical details of the beautiful subject of measure theory, see **[SS05]**, **[Bar66]**, **[Fol]**, or **[Roy]**.

Returning to boundedness, suppose T is a bounded linear operator from $L^1(\mathbb{R})$ to itself: there is a constant $C > 0$ such that for all $f \in L^1(\mathbb{R})$, $\|Tf\|_1 := \int_{\mathbb{R}} |Tf(x)|\, dx \leq C\|f\|_1$. Then, using several properties of Lebesgue integration[16] and recalling that

[16] The properties of the integral that we are using are: (i) if $f \geq 0$ and $A \subset B$, then $\int_A f \leq \int_B f$; (ii) if $g \leq f$ on A, then $\int_A g \leq \int_A f$; (iii) if $\lambda \in \mathbb{R}$, then $\int_A \lambda f = \lambda \int_A f$; (iv) $\int_A 1\, dx = \int \chi_A(x)\, dx$. See also the Appendix and Chapter 2 for some comments on the Lebesgue integral.

$E_\lambda(Tf) = \{x \in \mathbb{R} : |Tf(x)| > \lambda\} \subset \mathbb{R}$, we find that

$$\|Tf\|_1 = \int_{\mathbb{R}} |Tf(x)|\,dx \geq \int_{E_\lambda(Tf)} |Tf(x)|\,dx \geq \int_{E_\lambda(Tf)} \lambda\,dx$$
$$= \lambda \int_{E_\lambda(Tf)} 1\,dx = \lambda \int_{\mathbb{R}} \chi_{E_\lambda(Tf)}(x)\,dx = \lambda\, m(E_\lambda(Tf)).$$

Rearranging, we see that inequality (12.7) holds, and so T is weakly bounded on $L^1(\mathbb{R})$.

Thus *bounded on* $L^1(\mathbb{R})$ implies *weakly bounded on* $L^1(\mathbb{R})$. The converse is false. Although the Hilbert transform is weakly bounded, it is not bounded on $L^1(\mathbb{R})$, as the following example shows.

Exercise 12.31. Show that the function $H\chi_{[0,1]}$ is neither integrable nor bounded, despite the fact that the characteristic function $\chi_{[0,1]}$ is both. **Hint:** Use the formula for $H\chi_{[0,1]}$ in Exercise 12.13. ♢

12.5.2. The Hilbert transform is weakly bounded on $L^1(\mathbb{R})$.

We can now state precisely the boundedness result on $L^1(\mathbb{R})$ that holds for the Hilbert transform.

Theorem 12.32 (Kolmogorov's Inequality[17], 1927)**.** *The Hilbert transform is weakly bounded on* $L^1(\mathbb{R})$ *but is not bounded on* $L^1(\mathbb{R})$.

Exercise 12.31 shows that H cannot be bounded on $L^1(\mathbb{R})$, nor on $L^\infty(\mathbb{R})$. However, H is weakly bounded on $L^1(\mathbb{R})$. The proof uses the *Calderón–Zygmund decomposition*, which is stated and discussed in most harmonic analysis books; see for example [**Duo**].

The Hilbert transform maps bounded functions ($f \in L^\infty(\mathbb{R})$) into the larger space $\mathrm{BMO}(\mathbb{R})$ of functions of *bounded mean oscillation*, which contains certain unbounded functions such as $\log|x|$ as well as all bounded functions. See [**Gar**, Chapter VI].

Operators in the important class of *Calderón–Zygmund singular integral operators* also map $L^\infty(\mathbb{R})$ to $\mathrm{BMO}(\mathbb{R})$ (see the project in Section 12.9.3). The departure point of the *Calderón–Zygmund theory* is an *a priori* L^2 estimate; everything else unfolds from there. Having methods other than Fourier analysis to prove an L^2 estimate

[17]Named after the Russian mathematician Andrey Nikolaevich Kolmogorov (1903–1987).

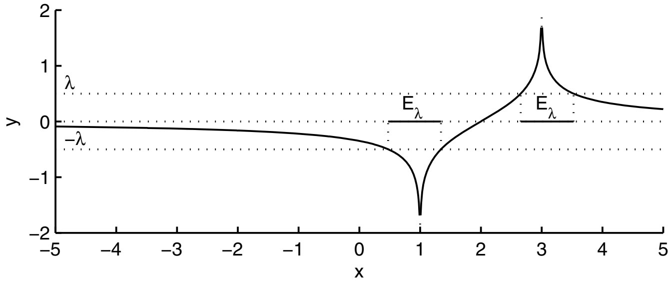


Figure 12.1. Graph of the Hilbert transform $H\chi_{[1,3]}(x) = (1/\pi)\log|(x-1)/(x-3)|$ of the characteristic function of the interval $[1,3]$, showing the λ-level set $E_\lambda = \{x \in \mathbb{R} : |H\chi_{[1,3]}(x)| \geq \lambda\}$ for $\lambda = 0.5$.

becomes important in other contexts such as *singular integral operators of nonconvolution type*; *analysis on curves or on domains other than* $\mathbb{R}^n$; and *analysis on weighted* L^2 *spaces* (where the underlying measure is not Lebesgue measure). One such method is *Cotlar's Lemma*[18] (see the project in Section 12.9.1). Others include *Schur's Test*[19] and $T(1)$ *theorems*. See [**Per01**, Section 2], [**Duo**, Chapter 9], and [**Graf09**, Chapter 8]. Dyadic methods, as in the modern dyadic proof of Riesz's Theorem, have become quite prominent recently.

Exercise 12.33. The Hilbert transform of $\chi_{[0,1]}$ is $H\chi_{[0,1]}(x) = \pi^{-1}\log(|x|\,|x-1|^{-1})$. (See Figure 12.1 for the graph of $H\chi_{[1,3]}$.) Find $C > 0$ such that $m\{x \in \mathbb{R} : |H\chi_{[0,1]}(x)| \geq \lambda\} \leq C/\lambda$, for all $\lambda > 0$. This fact alone does not prove that H is weakly bounded on $L^1(\mathbb{R})$, since for that we need the above inequality to hold with a uniform constant C for *all* $f \in L^1(\mathbb{R})$, but it is at least comforting.◊

Exercise 12.34. We investigate the measure of the level sets of the Hilbert transforms of characteristic functions of more complicated sets. Let $A = \bigcup_{j=1}^n [a_j, b_j]$ be a finite disjoint union of finite disjoint closed intervals, where $a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n$. The measure of A is $|A| := \sum_{j=1}^n |b_j - a_j|$. Show that the measure of the

[18] Named after Ukranian-born, Argentinian-Venezuelan by adoption, mathematician Mischa Cotlar (1911–2007).

[19] Named after the German mathematician Issai Schur (1875–1941).

level set $\{x \in \mathbb{R} : |H(\chi_A)(x)| > \lambda\}$ is exactly $4|A|/(e^{\pi\lambda} - e^{-\pi\lambda})$ (see **[SW59]** and **[Zyg71]**). ♢

12.5.3. The maximal function is weakly bounded on L^1. The *Hardy–Littlewood maximal function* Mf of a locally integrable function f is given by

$$Mf(x) = \sup_{x\in I} \frac{1}{|I|} \int_I |f(t)|\, dt.$$

See Aside 9.46. The maximal operator M is bounded on $L^\infty(\mathbb{R})$ and is weakly bounded on $L^1(\mathbb{R})$ but not bounded on $L^1(\mathbb{R})$.

Exercise 12.35. Verify that the maximal operator is bounded on $L^\infty(\mathbb{R})$. Specifically, show that $|Mf(x)| \leq \|f\|_\infty$, for all $f \in L^\infty(\mathbb{R})$ and each $x \in \mathbb{R}$, and hence $\|Mf\|_\infty \leq \|f\|_\infty$. ♢

Exercise 12.36. Calculate $M\chi_{[0,1]}$ explicitly and verify that it is not a function in $L^1(\mathbb{R})$. ♢

Theorem 12.37 (Hardy–Littlewood Maximal Theorem, 1930)**.** *The maximal operator is weakly bounded on $L^1(\mathbb{R})$, meaning that there is a constant C such that for each $f \in L^1(\mathbb{R})$ and for each $\lambda > 0$,*

$$m(\{x \in \mathbb{R} : |Mf(x)| > \lambda\}) \leq C\|f\|_1/\lambda.$$

The project in Section 12.9.2 outlines several proofs of Theorem 12.37. We can now prove the boundedness of the Hardy–Littlewood maximal operator on each L^p space, for $1 < p < \infty$, in one stroke.

Corollary 12.38. *The maximal operator M is bounded on $L^p(\mathbb{R})$, for each p with $1 < p < \infty$.*

Proof. By the Hardy–Littlewood Maximal Theorem (Theorem 12.37), M is weakly bounded on $L^1(\mathbb{R})$. Also, M is bounded on $L^\infty(\mathbb{R})$ (Exercise 12.35). The Marcinkiewicz Interpolation Theorem then implies that M is bounded on $L^p(\mathbb{R})$ for each p with $1 < p < \infty$. □

We turn to the useful Layer Cake Representation. The *measure*, or length, of the level sets introduced in the discussion above gives information about the size of f, but not about the behavior of f

near any given point. However, it does give enough information to compute the L^p norm of f.

Theorem 12.39. *For $f \in L^p(\mathbb{R})$ and $1 \leq p < \infty$ we have*

$$\|f\|_p^p = p \int_0^\infty \lambda^{p-1} m\{x \in \mathbb{R} : |f(x)| > \lambda\}\, d\lambda. \tag{12.8}$$

Formula (12.8) is the key to the proof of the Marcinkiewicz Interpolation Theorem, since it relates the λ level sets to the L^p norm. See Exercise 12.40 and [**LL**, Section 1.13] for more general versions of the *Layer Cake Representation* in Theorem 12.39.

Proof of Theorem 12.39. We have

$$\begin{aligned} p \int_0^\infty \lambda^{p-1} m\{x \in \mathbb{R} : |f(x)| > \lambda\}\, d\lambda &= \int_0^\infty \int_{\{x:|f(x)|>\lambda\}} p\lambda^{p-1}\, dx\, d\lambda \\ &= \int_{\mathbb{R}} \int_0^{|f(x)|} p\lambda^{p-1}\, d\lambda\, dx = \int_{\mathbb{R}} |f(x)|^p\, dx = \|f\|_p^p. \end{aligned}$$

We have used Fubini's Theorem (Theorem A.58) to interchange the order of integration. See Figure 12.2. □

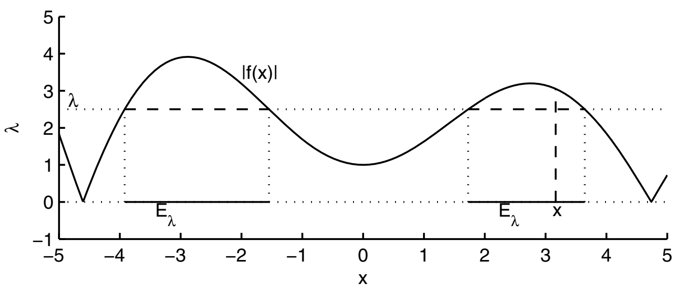


Figure 12.2. The two-dimensional region of integration $-\infty < x < \infty$, $0 \leq \lambda \leq |f(x)|$ in the upper half-plane, for Fubini's Theorem in the proof of Theorem 12.39. Here $E_\lambda = E_\lambda(f) := \{x \in \mathbb{R} : |f(x)| > \lambda\}$, and we have used $f(x) = x \sin(0.72x) \exp(-x/20) + 1$. The dashed lines are those one draws in a multivariable calculus class to explain interchanging the order of integration.

Exercise 12.40 (*Layer Cake Representation*). Show that for each continuously differentiable increasing function ϕ on $[0,\infty)$, we have $\int_{\mathbb{R}} \phi(|f|)\,dx = \int_0^\infty \phi'(\lambda) d_f(\lambda)\,d\lambda$. Theorem 12.39 corresponds to the choice $\phi(\lambda) = \lambda^p$. ♢

12.6. Interpolation and a festival of inequalities

In this section we state the very useful Riesz–Thorin Interpolation Theorem and deduce from it several widely used inequalities from analysis (Section 12.6.2) that we have encountered throughout the book. This theorem is the simplest example of an interpolation theorem. There are whole books devoted to such theorems. We highly recommend the book [**Sad**] by Cora Sadosky[20].

12.6.1. The Riesz–Thorin Interpolation Theorem.

The Riesz–Thorin Interpolation Theorem differs from the Marcinkiewicz Interpolation Theorem in that it does not deal with operators of weak-type. However, unlike the Marcinkiewicz Interpolation Theorem, it allows the use of different L^p spaces as inputs and outputs of the operator.

Theorem 12.41 (Riesz–Thorin Interpolation). *Suppose p_1, p_2, q_1, and q_2 lie in $[1,\infty]$. For $0 \le t \le 1$ define p and q by*

$$1/p = t/p_1 + (1-t)/p_2, \qquad 1/q = t/q_1 + (1-t)/q_2. \tag{12.9}$$

Let T be a linear operator from $L^{p_1}(\mathbb{R})+L^{p_2}(\mathbb{R})$ into $L^{q_1}(\mathbb{R})+L^{q_2}(\mathbb{R})$, such that T is bounded from $L^{p_i}(\mathbb{R})$ into $L^{q_i}(\mathbb{R})$ for $i = 1, 2$: there are constants A_1 and A_2 such that

$$\|Tf\|_{q_1} \le A_1\|f\|_{p_1} \qquad \text{for all } f \in L^{p_1}(\mathbb{R}) \text{ and}$$
$$\|Tf\|_{q_2} \le A_2\|f\|_{p_2} \qquad \text{for all } f \in L^{p_2}(\mathbb{R}).$$

Then T is bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$. Moreover,

$$\|Tf\|_q \le A_1^t A_2^{1-t}\|f\|_p \qquad \text{for all } f \in L^p(\mathbb{R}).$$

A little experimentation with specific values of p_1, p_2, q_1, q_2, p, and q will help to build intuition about which combinations satisfy the hypotheses of the theorem. See Figures 12.3 and 12.4 for a geometric interpretation of the interpolated points (p,q) as those points

[20] Argentinian mathematician Cora Sadosky (1940–2010). She was President of the Association for Women in Mathematics (AWM) from 1993 to 1995.

whose reciprocals $(1/p, 1/q)$ lie in the segment in $\mathbb{R}^2$ with endpoints $(1/p_1, 1/q_1)$ and $(1/p_2, 1/q_2)$ in specific applications of the interpolation theorem.

The proof of Theorem 12.41 uses the *three-lines theorem* for analytic functions; see for example [**SW71**, Chapter 5], [**Pin**, Section 3.2.1], or [**Graf08**, Chapter 1]. It is a beautiful complex analysis argument worth understanding.

12.6.2. A festival of inequalities. In this section we use Riesz–Thorin interpolation (Theorem 12.41) to prove several fundamental inequalities of analysis: the Hausdorff–Young Inequality; a generalized Hölder Inequality; a finite-dimensional version of Hölder's Inequality; Minkowski's Inequality and a finite-dimensional analogue; Minkowski's Integral Inequality; and Young's Inequality. The book by Elliot Lieb[21] and Michael Loss[22] [**LL**] is a wonderful resource for these and many other inequalities in analysis.

12.6.3. The Hausdorff–Young Inequality. Our first inequality arose in Section 8.6 in connection with the boundedness of the Fourier transform.

Corollary 12.42 (Hausdorff–Young Inequality). *Let* $1 \le p \le 2$, $1/p + 1/q = 1$, $f \in L^p(\mathbb{R})$. *Then its Fourier transform* $\widehat{f} \in L^q(\mathbb{R})$ *and*

$$\|\widehat{f}\|_q \le \|f\|_p. \tag{12.10}$$

Proof. We prove the Hausdorff–Young Inequality by interpolating between the L^2 boundedness of the Fourier transform, expressed by Plancherel's Identity $\|\widehat{f}\|_2 = \|f\|_2$ and the L^1–L^∞ boundedness of the Fourier transform: $\|\widehat{f}\|_\infty \le \|f\|_1$. Thus the hypotheses of the Riesz–Thorin Interpolation Theorem hold, with exponents $p_1 = 2$, $q_1 = 2$, $p_2 = 1$, $q_2 = \infty$ and constants $A_1 = 1$, $A_2 = 1$.

By formulas (12.9) and since $1/p + 1/q = 1$ and $1 \le p \le 2$, the pair $(x, y) = (1/p, 1/q)$ reached by the Interpolation Theorem must lie on the line $x + y = 1$, for $x \in [1/2, 1]$. See Figure 12.3. Finally, the Interpolation Theorem gives us the desired inequality with constant

[21] American mathematical physicist Elliot H. Lieb (born 1932).

[22] Swiss mathematician Michael Loss (born 1954).

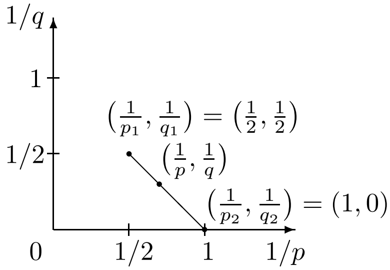


Figure 12.3. Interpolation diagram for the Hausdorff–Young Inequality. Endpoint data: $p_1 = 2$, $q_1 = 2$, $p_2 = 1$, $q_2 = \infty$.

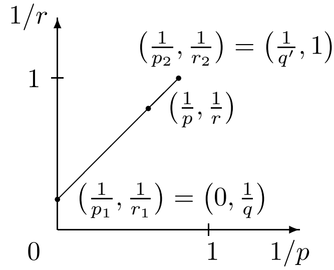


Figure 12.4. Interpolation diagram for the generalized Hölder Inequality. Endpoint data: $p_1 = \infty$, $r_1 = q$, $p_2 = q'$, $r_2 = 1$. We have used $q = 5$, $q' = 5/4$.

$A_1^t A_2^{1-t} = 1$ for all $t \in [0, 1]$: for the above range of values of p and q, $\|\widehat{f}\|_q \leq \|f\|_p$ as claimed. □

The multiplicative constant 1 on the right-hand side of inequality (12.10) is not optimal. The best possible constant in the Hausdorff–Young Inequality was found by K. I. Babenko[23] in 1961 and, independently, by Bill Beckner[24] in 1975 to be $\sqrt{p^{1/p}/q^{1/q}}$; see [**Bec**].

12.6.4. Hölder's Inequality. Our second inequality is a generalized version of the well-known Hölder Inequality[25]. It estimates the L^r norm of the product of two functions in terms of the L^p and L^q norms of the factors, for appropriate r, p, and q.

We used this generalized Hölder Inequality in Section 8.6 when considering L^p functions as distributions. Also, in the special case $q = q' = 2$, Hölder's Inequality reduces to the Cauchy–Schwarz Inequality, which we used in several arguments involving the Haar basis.

Corollary 12.43 (Generalized Hölder Inequality)**.** *Suppose f is in $L^p(\mathbb{R})$ and g is in $L^q(\mathbb{R})$. Then their product fg lies in $L^r(\mathbb{R})$, and*

$$\|fg\|_r \leq \|f\|_p \|g\|_q, \qquad \textit{where } 1/r = 1/p + 1/q.$$

[23] Russian mathematician Konstantin Ivanovich Babenko (1919–1987).
[24] American mathematician William E. Beckner (born 1941).
[25] Named after the German mathematician Otto Ludwig Hölder (1859–1937).

Note the change of notation from that used above in the Riesz–Thorin Interpolation Theorem. Here the endpoint variables are p and r, while q and q' are dual exponents: $1/q + 1/q' = 1$.

Proof. Fix q with $1 \leq q \leq \infty$, and take $g \in L^q(\mathbb{R})$. First notice that if $p = \infty$, then $r = q$. If $r = \infty$, then $r = p = q = \infty$ and $\|fg\|_\infty \leq \|f\|_\infty \|g\|_\infty$. If $1 \leq r < \infty$, then

$$\|fg\|_q = \Big(\int_{\mathbb{R}} |f(x)|^q |g(x)|^q \, dx \Big)^{1/q} \leq \Big(\|f\|_\infty^q \int_{\mathbb{R}} |g(x)|^q \, dx \Big)^{1/q}.$$

So, if $p = \infty$, then $\|fg\|_q \leq \|f\|_\infty \|g\|_q$ for all $1 \leq q \leq \infty$.

The other endpoint estimate is given by *Hölder's Inequality*, which corresponds to the case $r = 1$, $p = q'$, where $1/q + 1/q' = 1$:

$$\|fg\|_1 = \int_{\mathbb{R}} |f(x)g(x)| \, dx \leq \|f\|_{q'} \|g\|_q. \tag{12.11}$$

We outline a proof of inequality (12.11) in Exercise 12.47.

Apply the Riesz–Thorin Interpolation Theorem with $p_1 = \infty$, $q_1 = q$, $p_2 = q'$, $q_2 = 1$ and with the linear operator T given by pointwise multiplication by g as follows: $(Tf)(x) = f(x)g(x)$ for all x. This time the constants are $A_1 = A_2 = \|g\|_q$.

As shown in Figure 12.4, the point $(x, y) = (1/p, 1/r)$ lies on the line segment of slope 1 through the point $(0, 1/q)$, namely the segment $y = x + 1/q$, $0 \leq x \leq 1/q'$. Therefore $1/r = 1/p + 1/q$, and so the Interpolation Theorem implies the desired inequality

$$\|fg\|_r \leq A_1^t A_2^{1-t} \|f\|_p = \|g\|_p^t \|g\|_p^{1-t} \|f\|_p = \|f\|_p \|g\|_q. \qquad \square$$

Exercise 12.44. Verify that the constant 1 in the generalized Hölder Inequality cannot be improved. **Hint:** Consider $r = 1$, $p = q = 2$. ♢

Exercise 12.45. Show that if $1 \leq p < r < q \leq \infty$ and $h \in L^p(\mathbb{R}) \cap L^q(\mathbb{R})$, then $h \in L^r(\mathbb{R})$ and $\|h\|_r \leq \|h\|_p^t \|h\|_q^{1-t}$, where $1/r = t/p + (1-t)/q$. **Hint:** For $q < \infty$, apply Hölder's Inequality (12.11) with the dual exponents $p/(tr)$ and $q/((1-t)r)$ to the functions $f(x) := |h(x)|^{tr}$ and $g(x) := |h(x)|^{(1-t)r}$. ♢

Exercise 12.46 (*Finite-dimensional Hölder Inequality*)**.** Suppose a_i, $b_i \in \mathbb{R}$ for $i = 1, \ldots, n$. Prove that for $1/p + 1/p' = 1$, the inequality

$\sum_{i=1}^n |a_i b_i| \leq \big(\sum_{i=1}^n |a_i|^p\big)^{1/p}\big(\sum_{i=1}^n |b_i|^{p'}\big)^{1/p'}$ holds. Show that the special case of $n = 2$ and $p = p' = 2$ is a restatement of the geometric fact that $|\cos\theta| \leq 1$. **Hint:** The inequality $a^\lambda b^{1-\lambda} \leq \lambda a + (1-\lambda)b$ holds (why?) whenever $a, b \geq 0$ and $0 < \lambda < 1$. For each $j = 1, \ldots,$ n, set $a = |a_j|^p / \sum_{i=1}^n |a_i|^p$, $b = |b_j|^q / \sum_{i=1}^n |b_i|^q$, and $\lambda = 1/p$; then sum over j. ♢

Exercise 12.47. Pass from finite sums to Riemann sums and, taking limits, to integrals to obtain Hölder's Inequality (12.11) for Riemann integrable functions. Alternatively, follow the proof of the finite-dimensional Hölder Inequality, except that now for each $x \in \mathbb{R}$ set $a = |f(x)|^p/\|f\|_p^p$, $b = |g(x)|^q/\|g\|_q^q$, and $\lambda = 1/p$, and integrate with respect to x. ♢

12.6.5. Minkowski's Inequality.

From Hölder's Inequality we can deduce the *Triangle Inequality* in $L^p(\mathbb{R})$ or $L^p(\mathbb{T})$, which is one of the defining properties of a norm. In this setting the Triangle Inequality bears the name of Minkowski.

Lemma 12.48 (Minkowski's Inequality[26]). *If* $1 \leq p \leq \infty$ *and* f, $g \in L^p(\mathbb{R})$, *then* $\|f+g\|_p \leq \|f\|_p + \|g\|_p$.

Proof. The result is clear if $p = 1$, if $p = \infty$, or if $f + g = 0$. Otherwise, we estimate pointwise using the Triangle Inequality for complex numbers: $|f(x)+g(x)|^p \leq (|f(x)| + |g(x)|)\,|f(x)+g(x)|^{p-1}$. Now integrate over $\mathbb{R}$ to get

$$\|f+g\|_p^p \leq \int_{\mathbb{R}} |f(x)||f(x)+g(x)|^{p-1}\,dx + \int_{\mathbb{R}} |g(x)||f(x)+g(x)|^{p-1}\,dx.$$

Apply Hölder's Inequality, observing that if $1/p + 1/p' = 1$, then $(p-1)p' = p$. Thus the integrals are bounded by

$$\int_{\mathbb{R}} |h(x)||f(x)+g(x)|^{p-1}dx \leq \Big(\int_{\mathbb{R}} |h(x)|^p dx\Big)^{\frac{1}{p}} \Big(\int_{\mathbb{R}} |f(x)+g(x)|^p dx\Big)^{\frac{1}{p'}},$$

where $h = f$ in one case and $h = g$ in the other. Altogether we conclude that

$$\|f+g\|_p^p \leq \big(\|f\|_p + \|g\|_p\big)\|f+g\|_p^{p/p'}.$$

[26] Named after the German mathematician Hermann Minkowski (1864–1909).

Dividing both sides by the nonzero term $\|f+g\|_p^{p/p'}$, we get exactly the desired inequality. □

In fact one can have N summands or even a continuum of summands (an integral!). That is the content of our next inequality, which we mentioned in Section 7.7.

Lemma 12.49 (Minkowski's Integral Inequality). *Let $F(x,y)$ be a function of two variables such that for a.e. $y \in \mathbb{R}$, assume that the function $F_y(x) := F(x,y)$ belongs to $L^p(\mathbb{R}, dx)$ for some $p \geq 1$ and that the function $G(y) = \|F_y\|_{L^p(dx)}$ is in $L^1(\mathbb{R})$. Then the function $F_x(y) := F(x,y)$ is in $L^1(\mathbb{R}, dy)$ for a.e. $x \in \mathbb{R}$, and the function $H(x) := \int_{\mathbb{R}} F(x,y)\,dy$ is in $L^p(\mathbb{R}, dx)$. Moreover,*

$$\Big\| \int_{\mathbb{R}} F(\cdot, y)\,dy \Big\|_{L^p(dx)} \leq \int_{\mathbb{R}} \|F(\cdot, y)\|_{L^p(dx)}\,dy.$$

This inequality is a standard result in real analysis. We can think of it as the Triangle Inequality in $L^p(\mathbb{R})$ for integrals instead of sums.

We have used a more general version of Minkowski's Integral Inequality in the modern proof of Riesz's Theorem (Section 12.4.3), where the Lebesgue integral with respect to the variable y has been replaced by the integral with respect to a probability measure (the expectation E_ω): $\|E_\omega F(\cdot,\omega)\|_{L^p(dx)} \leq E_\omega \|F(\cdot,\omega)\|_{L^p(dx)}$, where ω belongs to the probability measure (Ω, dP).

In the following exercises we assume $p \geq 1$.

Exercise 12.50. Verify that if $f_1, f_2, \ldots, f_N \in L^p(\mathbb{R})$, then it is true that $\|f_1 + f_2 + \cdots + f_N\|_p \leq \|f_1\|_p + \|f_2\|_p + \cdots + \|f_N\|_p$. ◊

Exercise 12.51 (*Finite-dimensional Minkowski Inequality*). Suppose $a_i, b_i \in \mathbb{R}$, for $i = 1, \ldots, n$. Follow the proof of Minkowski's Inequality (Lemma 12.48), replacing the integrals by finite sums, to prove that $\big(\sum_{i=1}^n |a_i + b_i|^p\big)^{1/p} \leq \big(\sum_{i=1}^n |a_i|^p\big)^{1/p} + \big(\sum_{i=1}^n |b_i|^p\big)^{1/p}$. ◊

12.6.6. Young's Inequality. Our last inequality relates the L^r norm of the convolution of an L^p function and an L^q function to their norms, for appropriate r, p, and q with $1 \leq p, r, q$.

Corollary 12.52 (Young's Inequality). *Suppose f is in $L^p(\mathbb{R})$ and g is in $L^q(\mathbb{R})$. Then their convolution $f * g$ lies in $L^r(\mathbb{R})$, and*

$$\|f * g\|_r \leq \|f\|_p \|g\|_q, \qquad \text{where } 1/r + 1 = 1/p + 1/q.$$

Proof. Once again, we use interpolation. The endpoint pairs are $r = p$, $q = 1$ and $r = \infty$, $q = p'$, where $1/p + 1/p' = 1$. Fix $g \in L^p(\mathbb{R})$. We leave it as an exercise to show that if $f \in L^1(\mathbb{R})$, then

$$\|f * g\|_p \leq \|g\|_p \|f\|_1, \tag{12.12}$$

and if $f \in L^{p'}(\mathbb{R})$, then

$$\|f * g\|_\infty \leq \|g\|_p \|f\|_{p'}, \tag{12.13}$$

where $1/p + 1/p' = 1$. Then the Riesz–Thorin Interpolation Theorem applied to the linear operator T given by convolution with the fixed function $g \in L^p(\mathbb{R})$, $Tf = f * g$, implies Young's Inequality. □

The constant 1 in Young's Inequality is not optimal. Bill Beckner found the optimal constant to be $A_p A_q A_r$ where $A_p = \sqrt{p^{1/p}/(p')^{1/p'}}$; see [**Bec**].

Exercise 12.53. Sketch the interpolation diagram for the proof of Young's Inequality (similar to Figures 12.3 and 12.4). ♢

Exercise 12.54. Prove inequality (12.12), where $r = p$, $q = 1$, and inequality (12.13), where $r = \infty$, $q = p'$. The first inequality follows from Minkowski's Integral Inequality and the fact that the L^p norm is invariant under translations. The second is an application of Hölder's Inequality and, again, the translation-invariance of the L^p norm. ♢

12.7. Some history to conclude our journey

Why did mathematicians get interested in the Hilbert transform? We present two classical problems where the Hilbert transform appeared naturally. First, in complex analysis, the Hilbert transform links the Poisson kernel and the conjugate Poisson kernel on the upper half-plane. Second, in Fourier series, the convergence of the partial Fourier sums $S_N f$ to f in $L^p(\mathbb{T})$ for $1 < p < \infty$ follows from the boundedness of the periodic Hilbert transform on $L^p(\mathbb{T})$. These ideas bring us full circle back to where we started this journey: Fourier series.

In this section we identify the unit circle $\mathbb{T}$ with the interval $[0,1)$, instead of $[-\pi,\pi)$. This section is based on **[Per01]**.

12.7.1. Connection to complex analysis. Consider a real-valued function $f \in L^2(\mathbb{R})$. Let $F(z)$ be twice the analytic extension of f to the upper half-plane $\mathbb{R}^2_+ = \{z = x + iy : y > 0\}$. Then $F(z)$ is given explicitly by the well-known *Cauchy Integral Formula*[27]:

$$F(z) = \frac{1}{\pi i}\int_{\mathbb{R}} \frac{f(t)}{z-t}\,dt, \qquad z \in \mathbb{R}^2_+.$$

Note the resemblance between this integral and the Hilbert transform. No principal value is needed in the Cauchy Integral Formula, however, since the singularity is never achieved: $z \in \mathbb{R}^2_+$ cannot equal $t \in \mathbb{R}$. Separating the real and imaginary parts of the kernel, we find explicit formulas for the real and imaginary parts of $F(z) = u(z) + iv(z)$, in terms of convolutions with the *Poisson kernel* $P_y(x)$ and the *conjugate Poisson kernel* $Q_y(x)$ of the project in Section 7.8:

$$u(x+iy) = f * P_y(x), \qquad v(x+iy) = f * Q_y(x).$$

The function u is called the *harmonic extension* of f to the upper half-plane, while v is called the *harmonic conjugate* of u.

Exercise 12.55 (*The Poisson and Conjugate Poisson Kernels*)**.** Show that the Poisson kernel $P_y(x)$ and the conjugate Poisson kernel $Q_y(x)$ are given by $P_y(x) = y/\pi(x^2+y^2)$, $Q_y(x) = x/\pi(x^2+y^2)$. Calculate the Fourier transform of Q_y for each $y > 0$, and show that $\widehat{Q_y}(\xi) = -i\,\mathrm{sgn}(\xi)\exp(-2\pi|y\xi|)$. Therefore, as $y \to 0$, $\widehat{Q_y}(\xi) \to -i\,\mathrm{sgn}(\xi)$. Show that as $y \to 0$, $Q_y(x)$ approaches the principal value distribution p.v. $1/(\pi x)$ of Hf, meaning that for all test functions $\phi \in \mathcal{S}(\mathbb{R})$, $\lim_{y\to 0}\int_{\mathbb{R}} Q_y(x)\phi(x)\,dx = \lim_{\varepsilon\to 0}\int_{|x|>\varepsilon}\frac{\phi(x)}{x}\,dx$. ◊

The limit as $y \to 0$ of $u = P_y * f$ is f, both in the L^2 sense and almost everywhere, because the Poisson kernel is an approximation of the identity (Section 7.5). On the other hand, as $y \to 0$, $v = Q_y * f$ approaches the Hilbert transform Hf in $L^2(\mathbb{R})$. Therefore, by continuity of the Fourier transform in $L^2(\mathbb{R})$, we conclude that

$$(Hf)^\wedge(\xi) = -i\,\mathrm{sgn}(\xi)\widehat{f}(\xi),$$

[27]Named after the same Cauchy as the counterexample in Example 1.3 and as Cauchy sequences (Subsection 2.1.2).

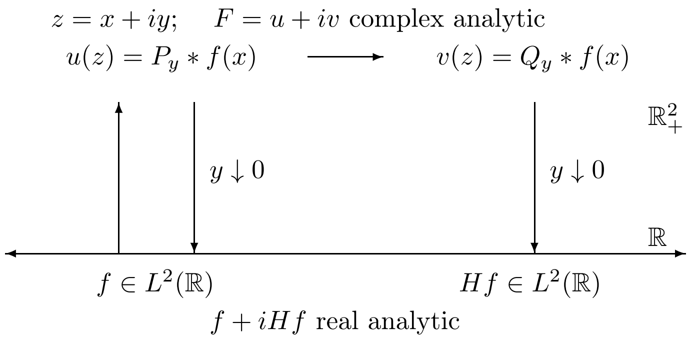


Figure 12.5. The Hilbert transform via the Poisson kernel. To sum up: the Hilbert transform Hf of f gives the boundary values (on $\mathbb{R}$) of the harmonic conjugate $v(x,y) := Q_y * f(x)$ of the harmonic extension $u(x,y) := P_y * f(x)$ of $f(x)$ to the upper half-plane.

which agrees with our original definition of $\widehat{Hf}$ (Definition 12.1).

The connection between the Hilbert transform and the Poisson kernel is illustrated in Figure 12.5.

An analogous result holds on the unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| \leq 1\}$. By a limiting procedure similar to the one described on the upper half-plane, one shows that the boundary values of the harmonic conjugate of the harmonic extension to $\mathbb{D}$ of a periodic, real-valued, continuously differentiable function f on the unit circle $\mathbb{T}$ are given by the periodic Hilbert transform of f,

$$H_P f(\theta) = \text{p.v.}\frac{1}{\pi}\int_0^1 f(t)\cot\big(\pi(\theta - t)\big)\,dt. \tag{12.14}$$

Here we are identifying the function $f(\theta)$ with $f(z)$, for $z = e^{2\pi i\theta}$. This is the periodic analogue of the Hilbert transform on the unit circle from Section 12.2. The analogy is reinforced since

$$(H_P f)^\wedge(n) = -i\ \text{sgn}(n)\widehat{f}(n). \tag{12.15}$$

Exercise 12.56. Show that for $H_P f$ defined on the time domain by equation (12.14), the Fourier coefficients are indeed given by equation (12.15). Assume f is smooth. $\Diamond$

Exercise 12.57 (*Exercise* 12.10 *Revisited*)**.** With the complex analysis picture in mind, write a two-line proof showing that $H_P c(\theta) = s(\theta)$ and $H_P s(\theta) = -c(\theta)$, for $s(\theta) = \sin(2\pi\theta)$ and $c(\theta) = \cos(2\pi\theta)$. $\lozenge$

For more on the periodic Hilbert transform and its connections to the conjugate function and analytic functions on the unit disc, see the books by Katznelson [**Kat**, Chapter III], Koosis [**Koo**, Chapter I.E], Torchinsky [**Tor**, Sections III.5, III.6], and Pinsky [**Pin**, Section 3.3].

12.7.2. Connection to Fourier series.

The *periodic Hilbert transform* H_P is an example of a *Fourier multiplier* T_m on the circle. These operators are linear transformations on $L^2(\mathbb{T})$, defined on the Fourier side by multiplication of the Fourier coefficients $\widehat{f}(n)$ of a given $f \in L^2(\mathbb{T})$, by a complex number $m(n)$ depending on the frequency n:

$$(T_m f)^\wedge(n) = m(n)\widehat{f}(n). \tag{12.16}$$

The sequence $\{m(n)\}_{n\in\mathbb{Z}}$ is called the *symbol* of T_m. For H_P, $m(n) = -i\,\mathrm{sgn}(n)$ (equation (12.15)).

The sequence $\{m(n)\}_{n\in\mathbb{Z}}$ is said to be a *bounded sequence*, written $\{m(n)\}_{n\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z})$, if there is a constant $M > 0$ such that $|m(n)| \leq M$ for all $n \in \mathbb{Z}$. For a bounded sequence $\{m(n)\}_{n\in\mathbb{Z}}$, we have

$$\begin{aligned} \|T_m f\|^2_{L^2(\mathbb{T})} &= \sum_{n\in\mathbb{Z}} |(T_m f)^\wedge(n)|^2 = \sum_{n\in\mathbb{Z}} |m(n)\widehat{f}(n)|^2 \\ &\leq M^2 \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2 = M^2\|f\|^2_{L^2(\mathbb{T})}, \end{aligned}$$

where the first and last equalities hold by Parseval's Identity. Thus if the sequence $\{m(n)\}_{n\in\mathbb{Z}}$ is in $\ell^\infty(\mathbb{Z})$, then the multiplier T_m with symbol $\{m(n)\}_{n\in\mathbb{Z}}$ is bounded on $L^2(\mathbb{T})$. The converse is also true and easy to prove in the periodic case.

Lemma 12.58. *A Fourier multiplier T_m on the circle $\mathbb{T}$ with symbol $\{m(n)\}_{n\in\mathbb{Z}}$ is bounded on $L^2(\mathbb{T})$ if and only if the sequence $\{m(n)\}_{n\in\mathbb{Z}}$ lies in $\ell^\infty(\mathbb{Z})$.*

Proof. We have just proved the "if" direction. For the "only if" direction, suppose that T_m is bounded on $L^2(\mathbb{T})$. Then there exists a constant $C > 0$ such that $\|T_m f\|_{L^2(\mathbb{T})} \leq C\|f\|_{L^2(\mathbb{T})}$ for all $f \in L^2(\mathbb{T})$.

In particular this inequality holds for $f(\theta) = e_n(\theta) = e^{2\pi in\theta}$, $n \in \mathbb{Z}$. The exponential functions form an orthonormal basis in $L^2(\mathbb{T})$; each of them has norm 1. Their Fourier coefficients have the simple form $\widehat{e_n}(k) = 1$ for $k = n$ and 0 otherwise. Therefore, using Parseval's Identity twice, for each $n \in \mathbb{Z}$ we have

$$|m(n)|^2 = \sum_{k\in\mathbb{Z}} |m(k)\widehat{e_n}(k)|^2 = \|T_m e_n\|^2_{L^2(\mathbb{T})} \leq C^2\|e_n\|^2_{L^2(\mathbb{T})} = C^2.$$

Thus $|m(n)| \leq C$ for all $n \in \mathbb{Z}$, so $\{m(n)\}_{n\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z})$. □

The operators H and H_P share similar boundedness properties.

Theorem 12.59 (Riesz's Theorem on the Circle)**.** *The periodic Hilbert transform H_P is bounded on $L^p(\mathbb{T})$ for $1 < p < \infty$: there is a constant $C_p > 0$ such that for all $f \in L^p(\mathbb{T})$,*

$$\|H_P f\|_{L^p(\mathbb{T})} \leq C_p\|f\|_{L^p(\mathbb{T})}.$$

The operation of taking the N^{th} partial Fourier sum of a nice function f, $S_N f(\theta) = \sum_{|n|\leq N} \widehat{f}(n)\, e^{2\pi in\theta}$, is another important example of a Fourier multiplier on the unit circle $\mathbb{T}$. Specifically,

$$(S_N f)^\wedge(n) = m_N(n)\widehat{f}(n),$$

where the symbol $\{m_N(n)\}_{n\in\mathbb{Z}}$ of S_N is the sequence $m_N(n) = 1$ if $|n| \leq N$ and $m_N(n) = 0$ if $|n| > N$.

The next two exercises show how to write S_N in terms of modulations of the periodic Hilbert transform, and they show that these modulation operators preserve the L^p norm of a function.

Exercise 12.60 (*Partial Fourier Sums as Modulations of the Periodic Hilbert Transform*)**.** Check that if $|n| \neq N$, then we can write $m_N(n) = \big(\operatorname{sgn}(n+N) - \operatorname{sgn}(n-N)\big)/2$. If $|n| = N$, then we can write $m_N(n) = \operatorname{sgn}(n+N) - \operatorname{sgn}(n-N)$. (Sketching these step functions will help.) Let M_N denote the modulation operator, $M_N f(\theta) := f(\theta)e^{2\pi i\theta N}$. Show that $i(M_N H_P M_{-N})^\wedge(n) = \operatorname{sgn}(n+N)\widehat{f}(n)$ for all $N \in \mathbb{Z}$. (The Fourier transform converts modulations into translations.) Apply the inverse Fourier transform to deduce that

$$\begin{aligned} S_N f(\theta) \;&=\; (i/2)\big(M_N H_P M_{-N} f(\theta) - M_{-N} H_P M_N f(\theta)\big) \\ &\quad + (1/2)\big(\widehat{f}(N)e^{iN\theta} + \widehat{f}(-N)e^{-iN\theta}\big). \end{aligned} \tag{12.17}$$

◊

Exercise 12.61 (*Modulation Operator Preserves L^p Norms*)**.** Show that $\|M_N f\|_{L^p(\mathbb{T})} = \|f\|_{L^p(\mathbb{T})}$, for all $N \in \mathbb{Z}$. ◊

We are ready to prove that $S_N f$ converges to f in $L^p(\mathbb{T})$.

Theorem 12.62. *The N^{th} partial Fourier sums $S_N f$ of a function $f \in L^p(\mathbb{T})$ converge to f in the L^p norm as $N \to \infty$, for each p with $1 < p < \infty$. That is,* $\lim_{N\to\infty} \|S_N f - f\|_{L^p(\mathbb{T})} = 0$.

Proof. It follows from formula (12.17) that the partial sums S_N are uniformly (in N) bounded on $L^p(\mathbb{T})$, for each p with $1 < p < \infty$. To see this, take L^p norms on both sides of (12.17) and apply the Triangle Inequality in $L^p(\mathbb{T})$. Note that the Fourier coefficients of a function in $L^p(\mathbb{T})$ are uniformly bounded by $\|f\|_{L^p(\mathbb{T})}$. Therefore the last two terms in formula (12.17) are uniformly bounded by $\|f\|_{L^p(\mathbb{T})}$. Thus we can bound $\|S_N f\|_{L^p(\mathbb{T})}$ by

$$\big(\|M_N H_P M_{-N} f\|_{L^p(\mathbb{T})} + \|M_{-N} H_P M_N f\|_{L^p(\mathbb{T})}\big)/2 + \|f\|_{L^p(\mathbb{T})}.$$

Next, by the boundedness properties of M_N and H_P,

$$\begin{aligned} \|S_N f\|_{L^p(\mathbb{T})} &\leq \big(\|H_P M_{-N} f\|_{L^p(\mathbb{T})} + \|H_P M_N f\|_{L^p(\mathbb{T})}\big)/2 + \|f\|_{L^p(\mathbb{T})} \\ &\leq (C_p/2)\big(\|M_{-N} f\|_{L^p(\mathbb{T})} + \|M_N f\|_{L^p(\mathbb{T})}\big) + \|f\|_{L^p(\mathbb{T})} \\ &= (C_p + 1)\|f\|_{L^p(\mathbb{T})}. \end{aligned}$$

The bound $(C_p + 1)$ is independent of N, as claimed.

The trigonometric polynomials are dense in $L^p(\mathbb{T})$, and for each $f(\theta) = \sum_{|m|\leq M} a_m e^{2\pi i m\theta}$, for $N \geq M$ we have $S_N f = f$. Thus, the partial sum operators S_N converge to the identity operator on a dense subset of $L^p(\mathbb{T})$. It follows, as in the proof of Theorem 9.36 (page 243) that $\lim_{N\to\infty} \|S_N f - f\|_p = 0$. □

Thus the convergence in $L^p(\mathbb{T})$ of the partial Fourier sums follows from the boundedness of the periodic Hilbert transform on $L^p(\mathbb{T})$.

Exercise 12.63. Take $f \in L^p(\mathbb{R}) \subset L^1(\mathbb{T})$. Use Hölder's Inequality (12.11) to show that $|\widehat{f}(n)| \leq (2\pi)^{-1/p}\|f\|_{L^p(\mathbb{T})}$. ◊

Formula (12.17) also implies that the periodic Hilbert transform cannot be bounded on $L^1(\mathbb{T})$, for otherwise the partial Fourier sums $S_N f$ would be uniformly bounded on $L^1(\mathbb{T})$, which is false.

Exercise 12.64. Show that there is no constant $C > 0$ for which $\|S_N f\|_{L^1(\mathbb{T})} \leq C\|f\|_{L^1(\mathbb{T})}$ for all $N > 0$. It may help to recall that the Dirichlet kernels D_N do not have uniformly bounded $L^1(\mathbb{T})$ norms, since $\|D_N\|_{L^1(\mathbb{T})} \sim \log N$; see Chapter 4 and Exercise 9.48. ♢

Exercise 12.65. Show that if $f \in \mathcal{S}(\mathbb{R})$, then the partial Fourier integrals $S_R f$ can be written as $2S_R f = M_R H M_{-R} f - M_{-R} H M_R f$, in terms of the Hilbert transform H and the modulations $M_R f(x) := e^{2\pi i R x} f(x)$ defined on $\mathbb{R}$. This decomposition is analogous to formula (12.17) for the partial Fourier sums. Use this decomposition, the boundedness on $L^p(\mathbb{R})$ of the Hilbert transform for $1 < p < \infty$, and the fact that $\|M_R f\|_p = \|f\|_p$ to show that $S_R f$ converges to f in the L^p norm. ♢

12.7.3. Looking further afield. In this book we have considered Fourier analysis and wavelets only in one dimension: on $\mathbb{T}$, $\mathbb{R}$, or discrete one-dimensional versions of these spaces. There is also a rich theory in $\mathbb{R}^n$ of multiple Fourier series, Fourier transforms, and wavelet transforms, with many fascinating applications (we briefly touched on the two-dimensional wavelet transform and its applications to image processing). The reader may also like to investigate the Riesz transforms and their connections to partial differential equations. The Riesz transforms are natural extensions of the Hilbert transform to higher dimensions.

We encourage the reader to start his or her excursion into multidimensional Fourier theory by reading the accounts in [**DM**, Sections 2.10 and 2.11], [**Kör**, Chapter 79 and above], [**Pin**, Chapters 4 and 5], [**Pre**, Chapters 6, 7, and 8], and [**SS03**, Chapter 6]. In these books you will also find applications of harmonic analysis in many areas including physics, differential equations, probability, crystallography, random walks, Radon transforms, radioastronomy, and number theory.

We have sampled a few of the many applications of wavelets and have encouraged the reader to explore more applications through the projects. There is a lot of potential for wavelets, and there is life beyond wavelets as well. Nowadays in the analysis of large data sets (think internet and genome), ideas from harmonic analysis and spectral graph theory are being used in novel ways and new tools are being invented (compressed sensing [**Can**], diffusion geometry [**CMa08**], [**CMa06**]).

In our book we have moved from Fourier series to wavelet series, from analytic formulas to fast algorithms, from matrices to operators, from dyadic to continuous, from Fourier transform to Hilbert transform, from Fourier to Haar, from Haar functions to Daubechies wavelets. We closed the circle by moving from Hilbert to Fourier. We ended with the venerable Hilbert transform, reconnecting with our initial theme of Fourier series. We gave a glimpse of what lies ahead in the field of harmonic analysis, and we saw that even for the now familiar Hilbert transform, new and surprising results are still being discovered, such as its representation in terms of Haar shifts: from Hilbert to Haar.

We thank the reader for joining us on our journey through harmonic analysis, and we wish you *bon voyage*!

12.8. Project: Edge detection and spectroscopy

Here are two practical applications of the Hilbert transform.

(a) One can use the discrete Hilbert transform (see the project in Section 6.8) to detect edges in images and to detect discontinuities in signals. The paper [**Eng**] describes "three edge detectors that work using spectral data (i.e. Fourier coefficients) about a function to 'concentrate' the function about its discontinuities. The first two detectors are based on the Discrete Hilbert Transform and are minimally invasive; unfortunately they are not very effective. The third method is both invasive and effective." The techniques used in all three methods are elegant and elementary. Read the paper [**Eng**], and make sense of these statements. Do some numerical experiments and comparisons between the three methods, and search further in the literature for other methods to detect edges, notably wavelet-based ones. See also the paper [**LC**].

(b) An Internet search for *Hilbert spectroscopy* in 2011 led to the following text in a Wikipedia entry citing the paper [**LDPU**]: *It is a technique that uses the Hilbert transform to detect signatures of chemical mixtures by analyzing broad spectrum signals from gigahertz* *to terahertz frequency radio. One suggested use is to quickly analyze liquids inside airport passenger luggage.* Read the paper [**LDPU**]. Explain the paragraph above, both mathematically and chemically, to a mathematical audience. Think of some other possible uses of the technique of Hilbert spectroscopy.

12.9. Projects: Harmonic analysis for researchers

The four projects gathered in this final section serve as an invitation to explore some advanced concepts in harmonic analysis, which are of central importance for researchers. We mentioned these concepts briefly in the text; if that sparked your curiosity, here is the opportunity to learn more. Elias Stein's book [**Ste93**] is an invaluable resource for researchers in this field.

12.9.1. Project: Cotlar's Lemma. In the project in Section 6.8 we defined discrete analogues of the Hilbert transform and used Fourier techniques to prove their boundedness. Here we introduce a technique, known as *Cotlar's Lemma*, or the *almost orthogonality lemma*, that works in settings where Fourier analysis is not available. We illustrate the use of this technique for the Hilbert transform and its discrete analogues.

Cotlar's Lemma says that if a given operator on a Hilbert space is a sum of operators that are uniformly bounded and *almost orthogonal* in a specific sense, then the original operator is bounded. If the pieces are actually orthogonal, then this is simply Plancherel's Identity or the Pythagorean Theorem. The content of the lemma is that the sum operator is still bounded even if the orthogonality of the pieces is significantly weakened.

(a) Find a precise statement and a proof of Cotlar's Lemma in the literature. Work through the proof and fill in the details. The proof given in [**Duo**, Lemma 9.1] is accessible and uses a beautiful counting argument. See also [**Cot**] and [**Burr**].

(b) Prove a version of the Hausdorff–Young Inequality for convolution of sequences: given sequences $x \in \ell^2(\mathbb{Z})$ and $y \in \ell^1(\mathbb{Z})$, then $x * y \in \ell^2(\mathbb{Z})$ and $\|x * y\|_{\ell^2(\mathbb{Z})} \leq \|y\|_{\ell^1(\mathbb{Z})}\|x\|_{\ell^2(\mathbb{Z})}$.

(c) Recall that $H^d x(n) = k * x(n)$ is given by discrete convolution with the kernel $k(n) = (1/n)\chi_{\{n\in\mathbb{Z}:n\neq 0\}}(n)$, which is in $\ell^2(\mathbb{Z})$ but not in $\ell^1(\mathbb{Z})$. Although the Hausdorff–Young Inequality does not imply that H^d is bounded on $\ell^2(\mathbb{Z})$, it does imply that certain truncated pieces are. Consider $H^d_j x(n) := k_j * x(n)$ where, for $j \geq 0$, k_j is the truncated Hilbert kernel given by $k_j(n) = (1/n)\chi_{S_j}(n)$, where the set $S_j = \{n \in \mathbb{Z} : 2^j \leq |n| < 2^{j+1}\}$. Then $H^d = \sum_{j\geq 0} H^d_j$. Show that the pieces H^d_j are uniformly bounded in $\ell^2(\mathbb{Z})$. To estimate the ℓ^2 norm of H^d_j, it suffices to estimate the ℓ^1 norm of each k_j (why?). Show that the ℓ^1 norms of the truncated kernels k_j have the following uniform bound: $\|k_j\|_{\ell^1} \leq 4\ln 2$.

(d) Use Cotlar's Lemma to show that the discrete Hilbert transform H^d is bounded on $\ell^2(\mathbb{Z})$. It will be useful to know that the following conditions on the truncated kernels hold for all $j \geq 0$ and $m \in \mathbb{Z}$: a *cancellation condition* $\sum_{n\in\mathbb{Z}} k_j(n) = 0$ and a *smoothness condition* $\sum_{n\in\mathbb{Z}} |k_j(m-n) - k_j(n)| \leq C2^{1-j}|m|$. To estimate the ℓ^2 norm of $H^d_i H^d_j$, it suffices to estimate the ℓ^1 norm of $k_i * k_j$ (why?). Verify that $\|k_i * k_j\|_{\ell^1} \leq C2^{-|i-j|}$.

12.9.2. Project: Maximal functions. The *dyadic maximal function* $M^d f$ is defined by $M^d f(x) = \sup_{I\ni x, I\in\mathcal{D}} (1/|I|) \int_I |f|$, where the supremum is over only dyadic intervals. Compare $M^d f$ with the Hardy–Littlewood maximal function Mf (see Section 12.5.3).

The dyadic maximal operator M^d and the Hardy–Littlewood maximal operator M are of weak type $(1,1)$; see Theorem 12.37. The reader is asked to explore one or more of the proofs of this theorem.

(a) Given a function $f \in L^1(\mathbb{R})$, you want to estimate the size of the set $\{x \in \mathbb{R} : M^d f(x) > \lambda\}$. Consider the *maximal* dyadic intervals for which $m_I|f| > \lambda$; these are exactly the intervals $\{J_i\}$ given by the *Calderón–Zygmund decomposition*. (Investigate what this decomposition is about.) Conclude that $\{x \in \mathbb{R} : M^d f(x) > \lambda\} = \bigcup J_i$ and $|E^d_\lambda| = \sum |J_i| \leq \|f\|_1/\lambda$. Hence M^d is of weak-type $(1,1)$.

(b) Most proofs are via *covering lemmas*. See for example [**Duo**, Section 2.4], [**Pin**, Section 3.5], and [**SS05**, Section 3.1.1]. Another proof uses a beautiful argument involving the *Rising Sun Lemma*; see [**Koo**, p. 234]. In G. H. Hardy and J. E. Littlewood's original paper, they first consider a discrete case, noting that *"the problem is most easily grasped when stated in the language of cricket, or any other game in which a player compiles a series of scores of which an average is recorded."* Their proof uses decreasing rearrangements and can be found in Zygmund's book [**Zyg59**] and in the original article [**HL**].

The best weak-type $(1, 1)$ constant for the centered maximal operator was shown to be the largest root of the quadratic equation $12x^2 - 22x + 5 = 0$; see [**Mel**]. For the uncentered Hardy–Littlewood maximal function, the best constant in L^p was found to be the unique positive solution of the equation $(p-1)x^p - px^{p-1} - 1 = 0$. See [**Graf08**, Exercise 2.1.2]. Other references are [**GM**], [**Ler**], and [**Wien**].

12.9.3. Project: BMO **and John–Nirenberg.** The space BMO, or BMO($\mathbb{R}$), of functions of *Bounded Mean Oscillation* (BMO) is the collection of locally integrable functions $b : \mathbb{R} \to \mathbb{R}$ such that $\|b\|_{\mathrm{BMO}} := \sup_I (1/|I|) \int_I |b(x) - m_I b|\, dx < \infty$, where the supremum is taken over all intervals I (this is a Banach space when considered modulo the constant functions).

(a) Understand the definition of BMO. Show that bounded functions are in BMO. Show that $\log|x|$ is in BMO, but the function $\chi_{x>0}(x)\log|x|$ is not in BMO, [**Graf08**, Examples 7.1.3 and 7.1.4].

(b) The *John–Nirenberg Inequality*[28] says that the function $\log|x|$ is, in terms of the measure of its λ-level sets, typical of unbounded BMO functions. More precisely, given a function $b \in$ BMO, an interval I, and a number $\lambda > 0$, there exist constants C_1, $C_2 > 0$ (independent of b, I, and λ) such that $|\{x \in I : |b(x) - m_I b| > \lambda\}| \leq C_1 |I| e^{-C_2\lambda/\|b\|_{\mathrm{BMO}}}$. Show that if a function b satisfies the John–Nirenberg Inequality, then $b \in$ BMO. Work through a proof of the John–Nirenberg Inequality (see [**Duo**], [**Per01**], or [**Graf08**]).

[28] Named after the German mathematician Fritz John (1910–1994) and the Canadian-born American mathematician Louis Nirenberg (born 1925). Nirenberg is the first recipient of the Chern Medal Award, given at the opening ceremony of the International Congress of Mathematicians, ICM 2010, Hyderabad, India, for his role in the formulation of the modern theory of nonlinear elliptic partial differential equations and for mentoring numerous students and post-docs in this area.

(c) One can deduce from the John–Nirenberg Inequality the following *self-improvement* result: If $b \in \mathrm{BMO}(\mathbb{R})$, then for each $p > 1$ there exists a constant $C_p > 0$ such that $\big((1/|I|)\int_I |b(x) - m_I b|^p\,dx\big)^{1/p} \leq C_p\|b\|_{\mathrm{BMO}}$, for all intervals I. Notice that the reverse inequality is exactly Hölder's Inequality. Therefore the left-hand side gives an alternative definition for the BMO norm. These *reverse Hölder Inequalities* can be thought of as *self-improvement* inequalities. Being a BMO function in $L^1(I)$, we conclude that we are also in $L^p(I)$ for all $p > 1$. This result is usually false without the BMO hypothesis; the reverse is always true. Prove the self-improvement inequality. Use Theorem 12.39 to write the L^p norm of $(b - m_I b)\chi_I$ in terms of its distribution function, and then use the John–Nirenberg Inequality.

(d) There is a dyadic version $\mathrm{BMO}_d(\mathbb{R})$ of the space of functions of bounded mean oscillation, in which the oscillation of the function over only dyadic intervals ($I \in \mathcal{D}$) is controlled. Specifically, a locally integrable function $b : \mathbb{R} \to \mathbb{R}$ is in $\mathrm{BMO}_d(\mathbb{R})$ if $\|b\|_{\mathrm{BMO}_d} := \sup_{I\in\mathcal{D}}(1/|I|)\int_I |b(x) - m_I b|dx < \infty$. A process called *translation-averaging* connects the two spaces. Given a suitable family of functions in the dyadic BMO space $\mathrm{BMO}_d(\mathbb{R})$, their translation-averaging is a new function that lies not only in $\mathrm{BMO}_d(\mathbb{R})$ but also in the strictly smaller space $\mathrm{BMO}(\mathbb{R})$. The averaging result extends to the more general setting of multiparameter BMO. While translation-averaging does not have the analogous improving effect on the related class of *dyadic A_p weight functions*, a modified version (*geometric-arithmetic averaging*) does. Explore these ideas in the papers [**GJ**], [**PW**], [**Treil**], [**War**], and [**PWX**].

12.9.4. Project: Hilbert as an average of Haar. We have seen the Hilbert transform on the frequency side as a Fourier multiplier, and on the time side as a singular integral operator. We discussed in Section 12.3.3 a striking new representation of the Hilbert transform, discovered by Stefanie Petermichl [**Pet**]. She proved that the Hilbert transform can be expressed as an average of translations and dilations of Haar multiplier operators, giving a new connection between continuous and dyadic aspects of harmonic analysis. In this project we want you to think more deeply about this representation.

(a) Read Petermichl's paper [**Pet**] and some of the references below. See also Sections 12.3 and 12.4. It is important to understand that the only bounded operators on $L^2(\mathbb{R})$ that commute with translations and dilations and anticommute with reflections are constant multiples of the Hilbert transform. See [**Hyt08**], [**Lac**], and [**Vag**].

(b) Develop (if possible; if not, explain what obstacles you encountered) an analogous theory for the discrete Hilbert transforms (both finite and sequential) defined in the project in Section 6.8.

(c) Compare Petermichl's representation of the Hilbert transform to the nonstandard decomposition for Calderón–Zygmund operators introduced in [**BCR**].

Appendix

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 spaces, 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).

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*: $\lambda_1(\lambda_2 x) = (\lambda_1\lambda_2)x$. ♢

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. ♢

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. ♢

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 Section A.3). ♢

Example A.5. The space $\ell^2(\mathbb{Z})$ of sequences defined by equation (5.1) in Section 5.1, with componentwise addition and scalar multiplication. See also Example A.14. ♢

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 $\|x\|_p = \left(\sum_{k=1}^n |x_k|^p \right)^{1/p}$. ♢

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 interval I (or $\mathbb{R}$) with the L^p norm. See Chapter 2 and Section A.3 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) $\langle x, x \rangle = 0 \iff x = 0$ ((iii) and (iv) *together: strictly positive-definite*).

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 $y \in Y$. $\Diamond$

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 $n \neq m$. $\Diamond$

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). $\Diamond$

Example A.20. The trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are *orthonormal* in $L^2(\mathbb{T})$ (see Chapter 5). $\Diamond$

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

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 $\{\lambda_1, \ldots, \lambda_n\} \subset \Lambda$ 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 $\lim_{N\to\infty} \left\| x - \sum_{n=1}^{N} a_n\, x_n \right\| = 0$. 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: $a_n = \langle x, x_n\rangle$. ◊

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* (Section A.2.1). In this context we discuss the notion of a (Schauder) basis and unconditional bases in Banach spaces (Section A.2.2). We define Hilbert spaces and we give criteria to decide when a given orthonormal system is a complete system (Section A.2.3). We discuss orthogonal complements and orthogonal projections onto close subspaces of a Hilbert space (Section A.2.4).

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

Definition A.29. Let V be a normed space. The sequence $\{x_n\}_{n\in\mathbb{N}}$ in V *converges to* $x \in V$ if $\lim_{n\to\infty} \|x_n - x\| = 0$. ◊

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

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, ... 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.31. 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*. ◊

Example A.32. 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]. ◊

Example A.33. 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. ◊

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.34 (*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. ◊

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 $1 \le p < \infty$, are bases under any permutation. Bases with the property that they remain bases 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.35 (*Unconditional Basis Equivalent Conditions*)**.** *Let X be a (real or complex) Banach space and let $\{x_n\}_{n\ge1}$ be a basis of X. Then the following are equivalent.*

(i) *$\{x_{\pi(n)}\}_{n\ge1}$ 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 $\{a_n\}_{n\ge1}$ and every sequence of scalars $\{\sigma_n\}_{n\ge1}$ of modulus at most 1, we have the inequality*

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

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.36. 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 $\mathcal{H}$. ◊

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* $\{f_n\}_{n\in\mathbb{N}}$ *of functions.*

Here are some canonical examples of Hilbert spaces.

Example A.37. $\mathbb{C}^n$, a finite-dimensional Hilbert space. ◊

Example A.38. $\ell^2(\mathbb{Z})$, an infinite-dimensional Hilbert space. ◊

Example A.39. $L^2(\mathbb{T})$, 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. ♢

Example A.40. $L^2(\mathbb{R})$, the Lebesgue square-integrable functions on $\mathbb{R}$. See Section A.3. ♢

A Hilbert space is called *separable* if it has a basis that is at most countable. Such a basis can always be orthonormalized. Examples A.37–A.40 are all separable Hilbert spaces.

How can we tell whether a given countable orthonormal family X in a separable Hilbert space $\mathcal{H}$ is an orthonormal basis for $\mathcal{H}$? One criterion for completeness of the orthonormal system is that the only vector orthogonal to all the vectors in $\mathcal{H}$ is the zero vector. Another is that Plancherel's Identity must hold. These ideas are summarized in Theorem A.41, which holds for all separable Hilbert spaces. In Chapter 5 we prove Theorem A.41 for the special case when $\mathcal{H} = L^2(\mathbb{T})$ with the trigonometric basis. We leave it as an exercise to extend the same proofs to the general case.

Theorem A.41. *Let $\mathcal{H}$ be a separable Hilbert space and let $X = \{x_n\}_{n\in\mathbb{N}}$ be an orthonormal family in $\mathcal{H}$. Then the following are equivalent:*

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

(ii) *If $x \in \mathcal{H}$ and $x \perp X$, 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.2.4. 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 $\mathcal{H}$.* ♢

Lemma A.43. *Given a Hilbert space $\mathcal{H}$ and a closed subspace V, 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 is defined to be $V \oplus W = \{x + y : x \in V, y \in W\}$. ♢

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,*

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

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 $\mathcal{P}_n(\mathbb{T})$. ♢

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 onto V. 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 $\mathcal{H}$, 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 $Q_1 x := P_2 x - P_1 x$ is the orthogonal projection of $x \in \mathcal{H}$ onto W_1.*

Proof. First observe that both $P_2 x$ and $P_1 x$ belong to the subspace V_2, and so does their difference, $Q_1 x \in V_2$. Assume that $Q_1 x$ is orthogonal to $P_1 x$ and that $Q_1 x \in W_1$ (the reader should justify these assertions). To show that $Q_1 x$ is the orthogonal projection of x onto W_1, we need only to check that $x - Q_1 x$ is orthogonal to $Q_1 x$. But $\langle x - Q_1 x, Q_1 x\rangle = \langle x - P_2 x + P_1 x, Q_1 x\rangle = \langle x - P_2 x, Q_1 x\rangle$, where the last equality holds because $P_1 x \perp Q_1 x$. Finally $P_2 x$ is the orthogonal projection of x onto V_2, so $x - P_2 x$ is orthogonal to V_2, in particular to $Q_1 x \in V_2$. We conclude that $\langle x - Q_1 x, Q_1 x\rangle = 0$. 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 $L^p(\mathbb{R})$ space 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 $E \subset \bigcup_{j\in\mathbb{N}} I_j$ 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 $L^p(\mathbb{R}) \subset L^1_{\text{loc}}(\mathbb{R})$. 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 $\mathbb{R} = \bigcup_{n\in\mathbb{N}} A_n$, 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. The 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 248. In this proof one in fact reduces the problem to the Weierstrass Approximation Theorem on the compact intervals $[-N, N]$. □

The results in Section 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 about interchanging limit operations and integrals that we have used in this book. 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 the project in Section 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 $\mathbb{R}^n$.

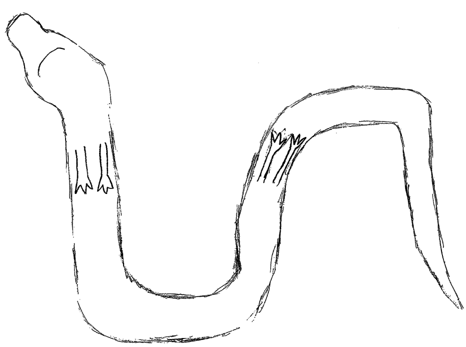
Interchanging limit operations is a delicate maneuver that is not always justified. We have illustrated throughout this book settings where this interchange is allowed and settings where it is illegal, such as the one just described involving uniform convergence.

Unlike the Riemann integral, Lebesgue theory allows for the interchange of a pointwise limit and an integral. There are several landmark theorems that one learns 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, and V. Rokhlin, *Fast wavelet transform and numerical algorithms*, Comm. Pure Appl. Math. **44** (1991), 141–183.

[Blå] V. Blåsjö, *The evolution of the isoperimetric problem*, Amer. Math. Monthly **112** (2005), 526–566.

[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.

[Burke] B. Burke Hubbard, *The world according to wavelets: The story of a mathematical technique in the making*, second edition, A K Peters/CRC Press, Natick, MA, 1998.

[Burkh] D. L. Burkholder, *Boundary value problems and sharp inequalities for martingale transforms*, Ann. Probab. **12**, No. 3 (1984), 647–702.

[Burr] 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, and C. Perez, *Sharp weighted estimates for approximating dyadic operators*, Electron. Res. Announc. Math. Sci. **17** (2010), 12–19.

[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.

[Dav] B. Davis, *On the weak-type* $(1, 1)$ *inequality for conjugate functions*, Proc. Amer. Math. Soc. **44** (1974), 307–311.

[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.

[GJ] J. B. Garnett and P. W. Jones, BMO *from dyadic* BMO, Pacific Math. J. **99**, No. 2 (1982), 351–371.

[Gib1898] J. W. Gibbs, *Fourier's series*, letter in Nature **59**, No. 1522 (December 29, 1898). Letter dated November 29, 1898.

[Gib1899] J. W. Gibbs, *Fourier's series*, letter in Nature **59**, No. 1539 (April 27, 1899). Letter dated April 12, 1899.

[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.

[Gre] 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. Hernández 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, *An 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 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., Washington, DC, 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**, No. 1 (2010), 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, 8 pp.

[Mall89] S. Mallat, *Multiresolution approximations and wavelet orthonormal bases for $L^2(\mathbb{R})$*, Trans. Amer. Math. Soc. **315** (1989), 69–87.

[Mall09] S. Mallat, *A wavelet tour of signal processing. The sparse way*, third edition, with contributions from G. Peyré, Elsevier/Academic Press, Amsterdam, 2009.

[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**, No. 1510 (October 6, 1898). Letter dated September 6, 1898.

[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.

[Per01] M. C. Pereyra, *Lecture notes on dyadic harmonic analysis*, Contemp. Math. **289** (2001), Amer. Math. Soc., Ch. I, 1–61.

[Per12] M. C. Pereyra, *Weighted inequalities and dyadic harmonic analysis.* Chapter in Excursions in harmonic analysis, T. Andrews, R. V. Balan, J. J. Benedetto, W. Czaja, and K. Okoudjou, to be published by the Springer imprint Birkhäuser, 2012.

[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.

[PW] J. Pipher and L. A. Ward, BMO *from dyadic* BMO *on the bidisc*, J. London Math. Soc. **77**, No. 2 (2008), 524–544.

[PWX] J. Pipher, L. A. Ward, and X. Xiao, *Geometric-arithmetic averaging of dyadic weights*, Rev. Mat. Iberoamericana **27**, No. 3 (2011), 953–976.

[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*, second 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.

[Ste70] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Univ. Press, Princeton, NJ, 1970.

[Ste93] E. M. Stein, *Harmonic analysis: Real-variable methods, orthogonality, and oscillatory integrals*, Princeton Univ. Press, Princeton, NJ, 1993.

[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., Extra Vol. (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.

[Treib] A. Treibergs, *Inequalities that imply the isoperimetric inequality*, `http://www.math.utah.edu/~treiberg/isoperim/isop.pdf` dated March 4, 2002.

[Treil] S. Treil, *H^1 and dyadic H^1*, in Linear and complex analysis: Dedicated to V. P. Havin on the occasion of his 75th birthday (ed. S. Kislyakov, A. Alexandrov, A. Baranov), Advances in the Mathematical Sciences **226** (2009), Amer. Math. Soc., Providence, RI.

[Vag] A. Vagharshakyan, *Recovering singular integrals from Haar shifts*, Proc. Amer. Math. Soc. **138**, No. 12 (2010), 4303–4309.

[VK] M. Vetterli and J. Kovačević, *Wavelets and subband coding*, Prentice Hall, Upper Saddle River, NJ, 1995.

[War] L. A. Ward, *Translation averages of good weights are not always good weights*, Rev. Mat. Iberoamericana **18**, No. 2 (2002), 379–407.

[Wic] V. Wickerhauser, *Adapted wavelet analysis from theory to software*, A K Peters, Wellesley, 1994.

[Wieg] J. Wiegert, *The sagacity of circles: A history of the isoperimetric problem*, `http://mathdl.maa.org/mathDL/23/` (keyword: sagacity), 2006.

[Wien] N. Wiener, *The ergodic theorem*, Duke Math. J. **5** (1939), 1–18.

[Wil] H. Wilbraham, *On a certain periodic function*, Cambridge & Dublin Math. J. **3** (1848), 198–201.

[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.

[Zho] D.-X. Zhou, *Stability of refinable functions, multiresolution analysis, and Haar bases*, SIAM J. Math. Anal. **27**, No. 3 (1996), 891–904.

[Zyg59] A. Zygmund, *Trigonometric Series*, second 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".

Titles in This Subseries

63 **María Cristina Pereyra and Lesley A. Ward,** Harmonic Analysis: From Fourier to Wavelets, 2012

58 **Álvaro Lozano-Robledo,** Elliptic Curves, Modular Forms, and Their L-functions, 2011

51 **Richard S. Palais and Robert A. Palais,** Differential Equations, Mechanics, and Computation, 2009

49 **Francis Bonahon,** Low-Dimensional Geometry, 2009

33 **Rekha R. Thomas,** Lectures in Geometric Combinatorics, 2006

32 **Sheldon Katz,** Enumerative Geometry and String Theory, 2006

7 **Judy L. Walker,** Codes and Curves, 2000

3 **Roger Knobel,** An Introduction to the Mathematical Theory of Waves, 2000

2 **Gregory F. Lawler and Lester N. Coyle,** Lectures on Contemporary Probability, 1999

Courtesy of Harvey Mudd College

In the last 200 years, harmonic analysis has been one of the most influential bodies of mathematical ideas, having been exceptionally significant both in its theoretical implications and in its enormous range of applicability throughout mathematics, science, and engineering.

In this book, the authors convey the remarkable beauty and applicability of the ideas that have grown from Fourier theory. They present for an advanced undergraduate and beginning graduate student audience the basics of harmonic analysis, from Fourier's study of the 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). While concentrating on the Fourier and Haar cases, the book touches on aspects of the world that lies between these two different ways of decomposing functions: time–frequency analysis (wavelets). Both finite and continuous perspectives are presented, allowing for the introduction of discrete Fourier and Haar transforms and fast algorithms, such as the Fast Fourier Transform (FFT) and its wavelet analogues.

The approach combines rigorous proof, inviting motivation, and numerous applications. Over 250 exercises are included in the text. Each chapter ends with ideas for projects in harmonic analysis that students can work on independently.

ISBN 978-0-8218-7566-7

9 780821 875667

STML/63

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-63