Harmonic Analysis: from Fourier to Haar

María Cristina Pereyra

Lesley A. Ward

Department of Mathematics and Statistics, MSC03 2150, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA
E-mail address: crisp@math.unm.edu

School of Mathematics and Statistics, University of South Australia, Mawson Lakes SA 5095, Australia
E-mail address: Lesley.Ward@unisa.edu.au

2000 *Mathematics Subject Classification.* Primary 42-xx

Contents

List of Figures

List of Tables

Introduction

Over two hundred years ago, Jean Baptiste Joseph Fourier began to work on the theory of heat. His book entitled *Théorie Analytique de la Chaleur* was published in 1822. In that work, he began the development of what is possibly the single most influential body of mathematical ideas, encompassing Fourier theory and the field now known as *harmonic analysis* that has grown from it. Since that time, the subject has been exceptionally influential 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 images stored and transmitted over the Internet using the Joint Photographic Experts Group (JPEG) standard 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 undergraduate 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 (space 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, although we concentrate on the Fourier and Haar cases.

The book is organized as follows. We begin with classical Fourier series for periodic functions, discussing physical motivations, different modes of convergence of Fourier series, appropriate classes of functions, the interplay with differentiability, approximations of the identity, the Dirichlet, Fejér, and Poisson kernels, inner-product vector spaces, and the completeness of the trigonometric basis. We touch on the conceptually simpler case of Fourier analysis on finite-dimensional spaces and the celebrated Fast Fourier Transform, an algorithm discovered by Gauss almost 200 years ago and rediscovered by Cooley and Tukey in the 1960s.

Next we discuss the Fourier transform on the line, as well as surveying the windowed Fourier transform and the wavelet transform. We develop the Haar basis, the geometry of dyadic intervals, and the multiresolution analysis that they induce. We introduce the time–frequency dictionary, tempered distributions, and applications of the Fourier transform such as the Shannon sampling theorem. We describe some properties and design features of known wavelets, as well as the basics of image/signal denoising and compression. We present the Hilbert transform, probably the most important operator in harmonic analysis after the Fourier transform, as well as its connections to complex and Fourier analysis. 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 look at some interpolation results, and as a consequence we derive some of the most useful inequalities in analysis.

The book ends with some appendices containing background material, and suggestions for further reading. Chapter 12 contains student projects for more independent study, designed to complement the material in the book. Examples, exercises, figures, and some snippets of MATLAB code are included throughout the text. The notation $A := B$ means that A is defined to be the quantity B, and similarly for $A =: B$. We use the symbol $\square$ to mark the end of a proof, and the symbol $\Diamond$ to mark the end of an example, exercise, definition, or project.

The prerequisites 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 they are needed.

The mathematical typesetting software LaTeX and the version control system Subversion were invaluable as we prepared the book. We created the figures using LaTeX and MATLAB. Early versions of some figures were kindly provided by Martin Mohlenkamp.

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

Our book grew out of the lecture course we gave at the IAS/Princeton Program for Women in Mathematics on *Analysis and Partial Differential Equations*, during the period May 17–28, 2004. We thank Sun-Yung Alice Chang and Karen Uhlenbeck, the organizers, for giving us the opportunity to participate in such a remarkable program.

We warmly thank Manuela de Castro and Stephanie Molnar, our teaching assistants during the Program, for their invaluable help, and all the students for their enthusiasm and lively participation.

M. Cristina Pereyra, Albuquerque
Lesley A. Ward, Adelaide

CHAPTER 1

Fourier series: some motivation

In this book we discuss three types of Fourier analysis: 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 $n \in \mathbb{Z}$; second, *finite Fourier analysis*, where the input is a vector of length N with complex entries, and the output is another vector in $\mathbb{C}^N$; and third, the *Fourier transform*, where the input is a function on $\mathbb{R}$, and the output is another function on $\mathbb{R}$.

As an aside, we note that one can also do Fourier analysis on $\mathbb{R}^n$, on abelian groups, on graphs, and in still more abstract settings. See [**SS1**, Chapter 7]. Fourier analysis is also connected with representation theory. Some of the projects in Chapter 12 are related to these abstract settings.

1.1. Some examples and key definitions

The central idea of Fourier analysis is to break a function into a combination of simpler functions. We think of these simpler functions as *building blocks*. We will also 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 will see that our simpler functions can also correspond to pure frequencies. For example, we will shortly consider sine and cosine functions of various frequencies as our first example of building blocks. When played aloud, a given sine or cosine function produces a *pure tone*, or note, or harmonic, at a single frequency. The term *harmonic analysis* evokes this idea of separation of a sound, or in our terms a function, into pure tones.

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

We begin with an example.

Example 1.1. (*Toy Model of Voice Signal*) Suppose Amanda is in Baltimore, and she calls her mother in Vancouver, saying ‘Hi Mom, it’s Amanda, and I can’t wait to tell you what happened today’. What happens to the sound? As Amanda speaks, she creates waves of pressure in the air, which travel toward the phone receiver. The sound has duration, say 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 us try our idea of breaking Amanda's voice signal into simpler building blocks. Suppose the signal looks like the function $f(t)$ plotted in Figure 1.1. (In reality, her voice signal would be much more complicated.)

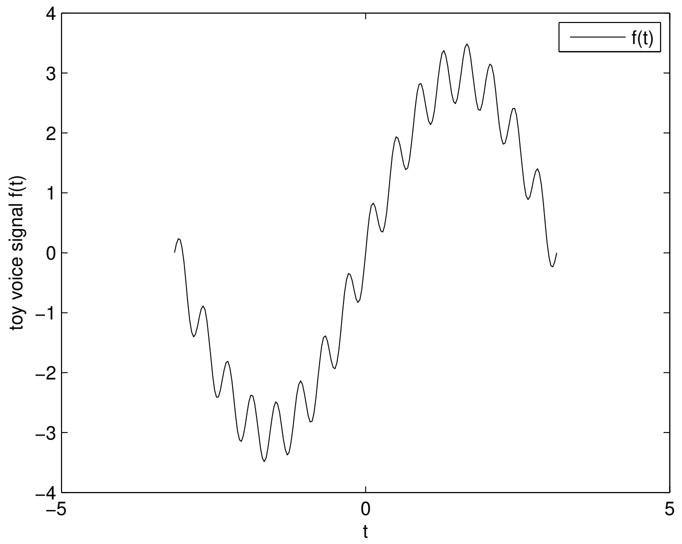


FIGURE 1.1. Toy voice signal $f(t)$.

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 Figure 1.2.

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, meaning that they complete respectively 1 and 16 full oscillations as t runs through 2π units of time.

Next we need to know how much of each building block is present in our signal. The maximum amplitudes of the large and small wiggles are 3 and 0.5 respectively, giving us the terms $3\sin t$ and $\frac{1}{2}\sin 16t$. We add these together to build the original signal:

$$f(t) = 3\sin t + \frac{1}{2}\sin 16t.$$

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 us 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 will know how to put them back together. (By the way,

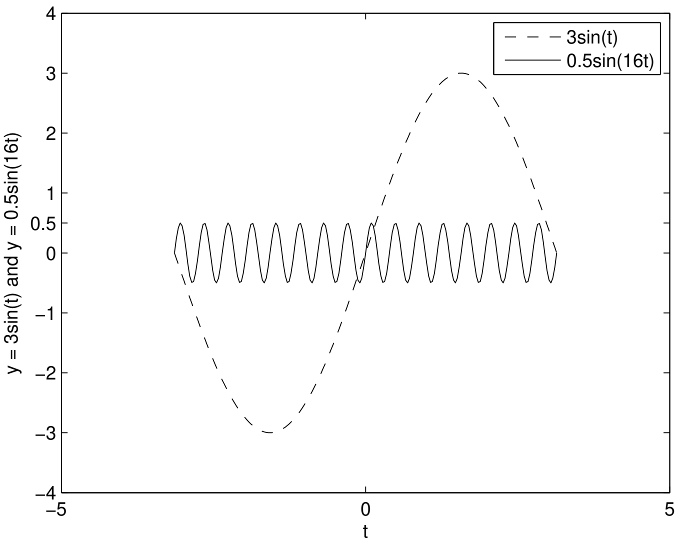


FIGURE 1.2. Building blocks of toy voice signal $f(t)$.

can you find a more efficient way of encoding the timing information?) But for our signal, we can achieve the same result with just *four* numbers. We simply send the frequencies 1 and 16, and the corresponding strengths or *amplitudes* 3 and 1/2. Once you know the code, 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 us be more ambitious. Are we willing to sacrifice a bit of the quality of Amanda's signal in order to send it more cheaply? Maybe. For instance, in our signal the strongest component is the big wiggle. What if we send only the single frequency 1 and its corresponding amplitude 3? In Vancouver, only the big wiggle will be reconstructed, and the small fast wiggle will be lost from our signal. But Amanda's mother knows the sound of Amanda's voice, and the imperfect reconstructed signal may still be recognizable. This is our first example of *compression* of a signal.

FIGURE 1.3. Plot of an actual voice saying 'Hi Mom, it's Amanda, and I can't wait to tell you what happened today'. Explain the units: time in seconds along the horizontal axis, a proxy for volume/intensity along the vertical axis. Check how long it lasts and correct the estimate (currently 'five seconds') in the first paragraph of this example.

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

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

In the interests of realism, Figure 1.3 shows a plot of an actual voice saying 'Hi Mom, it's Amanda, and I can't wait to tell you what happened today'. $\Diamond$

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

The function f in Example 1.1 is especially simple. The intuition that sines and cosines were sufficient to describe many different functions was gained from the experience of the pioneers of Fourier analysis with physical problems (for instance heat diffusion and the vibrating string) in the early eighteenth century; see Section 1.4. This intuition leads to the idea of expressing a function $f(\theta)$ as an infinite linear combination of sines and cosines, also known as a *trigonometric series*:

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

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

We can rewrite the right-hand side of (1.1) as a linear combination of exponential functions. To do so, we use *Euler's Formula*[1]: $e^{i\theta} = \cos\theta + i \sin\theta$, and the

[1] 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

corresponding formulas for sine and cosine in terms of exponentials, applied to $n\theta$ for each n,

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

We will use the following version throughout:

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

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

Exercise 1.2. Write a_n in terms of the b_n and c_n. $\Diamond$

1.2. Main questions

Our brief sketch immediately suggests several questions.

- How can we find the coefficients a_n from the corresponding function f?
- Given the $\{a_n\}$, how can we reconstruct f?
- In what sense does the series converge? Pointwise? Uniformly? or yet in a different sense?
- Does it converge to f in any of the above senses? If it does, how fast does it converge?
- Which functions can we express with a trigonometric series

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

  For example, must f be Riemann integrable, or continuous?
- What other building blocks could we use besides 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 infinite differentiable function f can be expanded into a *Taylor series*[2] centered say at $x = 0$,

$$\text{Taylor series and coefficients:} \qquad \sum_{n=0}^{\infty} c_n\, x^n, \qquad \text{where } c_n = \frac{f^{(n)}(0)}{n!}. \tag{1.3}$$

The Taylor series always converges to $f(0) = c_0$ when evaluated at $x = 0$. In the 18^{th} century mathematicians discovered that the traditional functions of calculus ($\sin x$, $\cos x$, $\ln(1+x)$, $\sqrt{1+x}$, e^x, etc.) could be expanded in Taylor series and that the series converges to the function in an open interval containing $x = 0$, 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 [**Tao2**, Section 15.6] for a quick review of complex numbers.

[2]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).

they were very good at manipulating power series and calculating with them. That led them to believe that the same would be true for all functions, which at the time meant infinitely differentiable functions. That dream was shattered by Cauchy's[3] discovery in 1821 of a counterexample.

EXAMPLE 1.3. (*Cauchy's Counterexample*) The function

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

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

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

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

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

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. ◊

A *trigonometric*[4] *polynomial of degree* M is a function of the form

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

EXERCISE 1.6. Verify that if f is a trigonometric polynomial then its coefficients $\{a_k\}$ can be calculated with the following formula:

$$a_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-ik\theta} \, d\theta.$$

◊

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).

1.3. Fourier series and Fourier coefficients

To begin answering our questions, the Fourier coefficients $\widehat{f}(n) = a_n$ are calculated using the formula suggested by Exercise 1.6:

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

ASIDE 1.7. *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, as in Theorem* 3.6 *below.*

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

[4]To be more precise, this particular trigonometric polynomial is 2π-periodic; see Section 1.3.1.

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

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

and then to proceed formally[5], 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*[6] *delta* $\delta_{n,k}$ is defined by

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

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

ASIDE 1.8. *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 the minimal requirement on f for the integral in formula (1.5) for the coefficients to exist is that the complex-valued function $f(\theta)e^{-in\theta}$ should be an

[5]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.

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

integrable[7] function on $[-\pi, \pi)$. In fact, if $|f|$ is integrable, then so is $f(\theta)e^{-in\theta}$ and

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

Aside 1.9. *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: $|a + b| \le |a| + |b|$, or the triangle inequality for integrable functions: $\int |f + g| \le \int |f| + \int |g|$. They are all animals in the same family.*

We will spend some time exploring in what sense the Fourier series (1.2) approximates the original function f. We begin with some examples.

Exercise 1.10. Find the Fourier coefficients and the Fourier series for the trigonometric polynomial

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

$\lozenge$

Example 1.11. (*Periodic Ramp Function*) Compute the Fourier coefficients and the Fourier series for the function

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

The n^{th} Fourier coefficient is given by

$$\begin{aligned}
\widehat{f}(n) = a_n &:= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\, d\theta \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta e^{-in\theta}\, d\theta \\
&= \begin{cases} (-1)^{n+1}/(in), & \text{if } n \ne 0; \\ 0, & \text{if } n = 0. \end{cases}
\end{aligned}$$

using integration by parts in the last line. Thus the Fourier series of $f(\theta) = \theta$ is given by

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

---

[7] 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 [**Tao1**, Chapters 11], [**Tao2**, Chapters 19].

These integration methods are named after the French mathematician Henri Lebesgue (1875–1941), and the German mathematician Georg Friedrich Bernhard Riemann (1826–1866).

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

EXERCISE 1.12. Fill in the details in the integration-by-parts argument in the calculation in Example 1.11. ◊

1.3.1. 2π-Periodic functions. Note that we are considering functions

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

that are complex-valued, and defined on a bounded half-open interval in the real line. We extend the function periodically to the whole of $\mathbb{R}$. For example, Figure 1.4 shows part of the graph of the periodic extension to $\mathbb{R}$ of the real-valued function f defined on $[-\pi, \pi)$ in Example 1.11.

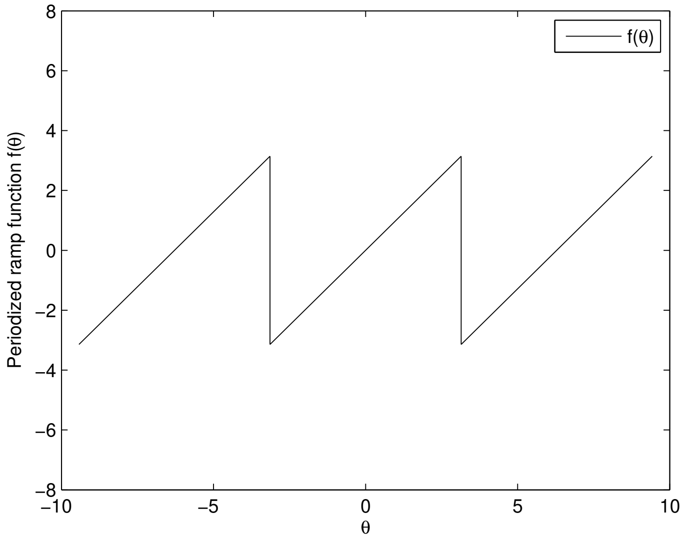


FIGURE 1.4. Periodized ramp function on $[-3\pi, 3\pi]$.

EXERCISE 1.13. Run the following MATLAB script to reproduce the plot in Figure 1.4.

```
clear all
plot([-3*pi,-pi],[-pi,pi],'k',[-pi,-pi],[-pi,pi],'k',...
  [-pi,pi],[-pi,pi],'k',[pi,pi],[-pi,pi],'k',[pi,3*pi],[-pi,pi],'k')
axis([-10,10,-8,8])
xlabel('\theta')
ylabel('Periodized ramp function f(\theta)')
legend('f(\theta)')}
```

◊

EXERCISE 1.14. Modify the MATLAB script from Exercise 1.13 to plot over $[-3\pi, 3\pi]$ the periodic extension to $\mathbb{R}$ of the function defined by $f(x) = x^2$ for $x \in [-\pi, \pi)$. ◊

In mathematical language, a function defined on $\mathbb{R}$ is 2π-*periodic* if $f(x+2\pi) = f(x)$ for all $x \in \mathbb{R}$. The building blocks we are using are 2π-periodic functions: $\sin(n\theta)$, $\cos(n\theta)$, $e^{-in\theta}$. Finite linear combinations of 2π-periodic functions are 2π-periodic.

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

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

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

$$g(z) := f(\theta).$$

Note that if g is to be continuous on the unit circle $\mathbb{T}$, we will need $f(-\pi) = f(\pi)$. 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', ..., $f^{(k)}$ 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 $0 \le \ell \le k$.

Note that if f is 2π-periodic, 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 1.15. Verify that if f is 2π-periodic, then for all $a \in \mathbb{R}$

$$\int_a^{a+2\pi} f(\theta)\, d\theta = \int_{-\pi}^{\pi} f(\theta)\, d\theta.$$

◊

In general, 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 its Fourier series coincides with f itself. See Exercises 1.6 and 1.17.

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

An *L-periodic function* f defined on $\mathbb{R}$ is a function such that $f(x) = f(x + L)$ for all $x \in \mathbb{R}$. We can define the Fourier coefficients and the Fourier series for an

L-periodic function f. The formulae come out slightly differently to account for the rescaling:

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

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

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

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

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

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

Exercise 1.17. Let f be an L-periodic trigonometric polynomial of degree N, 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. A little history, and motivation from the physical world

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

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

For the last 200 years, Fourier Analysis has been of immense practical importance, both for theoretical mathematics (especially in the subfield now called

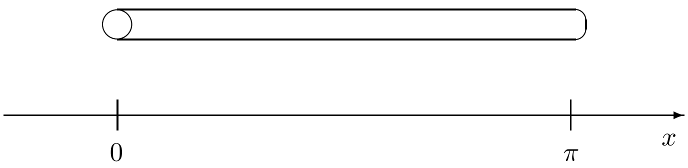


FIGURE 1.5. Sketch of a one-dimensional rod.

harmonic analysis), and in understanding, modeling, predicting, and controlling the behavior of physical systems. Older applications in mathematics, physics and engineering include the way that heat diffuses in a solid object, and the wave motion undergone by a string or a two-dimensional surface such as a drumhead. Recent applications include telecommunications (radio, wireless phones) and more generally data storage and transmission (the JPEG format for images on the Internet).

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

1.4.1. Temperature distribution in a one-dimensional rod. Our task is to find the temperature $u(x, t)$ in a rod of length π, where $x \in [0, \pi]$ represents the position of a point on the rod, and $t \in [0, \infty)$ represents time. See Figure 1.5. We assume that the initial temperature (when $t = 0$) is given by a known function $f(x) = u(x, 0)$, and that at both endpoints of the rod the temperature is held at zero, giving the boundary conditions $u(0, t) = u(\pi, t) = 0$. There are many other plausible boundary conditions.

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

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}. \tag{1.9}$$

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

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

By equation (1.9), we conclude that

$$\alpha(x)\beta'(t) = \alpha''(x)\beta(t),$$

and so

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

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 1.18. *A note on notation: Here we are using a typical 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*

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

Finite linear combinations of these solutions will also be 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.

1.4.2. Other physical models. In Table 1.1 we present schematically several physical models. In each case, one can do an heuristic analysis similar to the one performed in Section 1.4.1 for the heat equation in a one-dimensional rod.

EXERCISE 1.19. Find solutions to the physical models presented in Table 1.1 by an analysis similar to the one used in Section 1.4.1. $\Diamond$

The initial or boundary conditions in the table specify the value of u at time $t = 0$, that is $u(x, 0) = f(x)$.

TABLE 1.1. Physical models.

Model	Initial or Boundary Condition	Solution
Vibrating string $\boxed{u_{tt} = u_{xx}}$ $u(0,t) = u(\pi,t) = 0$	$f(x) = \sum_n b_n \sin nx$	$u(x,t) = \sum_n b_n e^{int} \sin nx$
Temperature in a rod $\boxed{u_t = u_{xx}}$ $u_x(0,t) = u_x(\pi,t)$	$f(x) = \sum_n c_n \cos nx$	$u(x,t) = \sum_n c_n e^{-n^2 t} \cos nx$
Steady-state temp. in a disk $\boxed{u_{rr} + r^{-1}u_r + r^{-2}u_{\theta\theta} = 0}$	$f(\theta) = \sum_n a_n e^{in\theta}$	$u(r,\theta) = \sum_n a_n r^{\lvert n\rvert} e^{in\theta}$
Steady-state temp. in a rectangle $\boxed{u_{xx} + u_{yy} = 0}$	$f_0(x) = \sum_k A_k \sin kx$ $f_1(x) = \sum_k B_k \sin kx$	$u(x,y) = \sum_k \left(\frac{\sinh[k(1-y)]}{\sinh k} A_k + \frac{\sinh ky}{\sinh k} B_k\right) \sin kx$

CHAPTER 2

Interlude

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

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

2.1. Nested classes of functions on bounded intervals

Here we introduce some classes of functions on bounded intervals I (the intervals 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, following the exposition in the book by Terence Tao[1] [**Tao1**, Chapter 11]. We highly recommend Tao's book to our readers.

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 $\{J_k\}_{k=1,\dots,n}$ such that $I = \bigcup_{k=1}^n J_i$. The intervals J_i 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:

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

$\Diamond$

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

FIGURE 2.1. Graph of the step function $h = 3\chi_{[-1,2)} - \chi_{[2,4]} + 2\chi_{(5,6)}$ in Example 2.4.

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

DEFINITION 2.3. A function h defined on I is a *step function* if there are a partition P of the interval I, and 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. ♢

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

EXAMPLE 2.4. The function $h = 3\chi_{[-1,2)} - \chi_{[2,4]} + 2\chi_{(5,6)}$ is a step function defined on the interval $[-1, 6)$. Its graph is shown in Figure 2.1. ♢

If we ask the reader to compute the integral of this function over the interval $[-1, 6)$, she will add (with appropriate signs) the areas of the rectangles in the picture to get $(3 \times 3) + (-1 \times 2) + (2 \times 1) = 9$. More generally, we can 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 could be associated to more than one partition, and one can verify that this definition of the integral is independent of the partition chosen. Notice that the integral of the characteristic function of a subinterval J of I is its length $|J|$.

EXERCISE 2.5. Given an interval $J \subset I$, show that

♢ $$\int_I \chi_J(x)\,dx = |J|.$$

♢

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*. ♢

EXERCISE 2.7. Verify that step functions are always bounded functions. ♢

DEFINITION 2.8. A function $f : I \to \mathbb{R}$ is *Riemann integrable* if f is bounded on I, and if the supremum of the integrals of all step functions whose graphs lie below the graph of f is equal to the infimum of the integrals of all step functions whose graphs lie above that of f. In mathematical notation,

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

FIGURE 2.2. This figure illustrates the approximation of a function f from below and from above by step functions h_1 and h_2 respectively, as in the definition of the Riemann integral.

where the supremum and infimum are taken over step functions $h_1(x) \leq f(x) \leq h_2(x)$ 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 $\mathcal{R}(I)$. See Figure 2.2 illustrating the approximation of f from below and from above by step functions. $\Diamond$

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 $\int_I f < \int_I g$. The *triangle inequality*

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

holds for Riemann-integrable functions. See [**Tao1**, Section 11.4].

EXERCISE 2.9. (*The Riemann Integral is Linear*) Given two step functions h_1 and h_2 defined on I, verify that for c, $d \in \mathbb{R}$, the linear combination $ch_1 + dh_2$ is again a step function. Also check that

$$\int_I \big(ch_1(x) + dh_2(x)\big)\,dx = c\int_I h_1(x)\,dx + d\int_I h_2(x)\,dx.$$

Warning: 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*).

Use Definition 2.8 to show that this linearity property is inherited by the Riemann-integrable functions. $\Diamond$

One can deduce from Definition 2.8 the following very useful proposition.

PROPOSITION 2.10. *A bounded function f is Riemann integrable over I if and only if for each $\varepsilon > 0$ there are step functions h_1 and h_2 defined on I below and above f, that is $h_1(x) \leq f(x) \leq h_2(x)$ for all $x \in I$, 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

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

FIGURE 2.3. The graph illustrates the integral of the difference of the two step functions in Figure 2.2.

See Figure 2.3 illustrating the integral of the difference of the two step functions in Figure 2.2.

ASIDE 2.11. *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 coincides with a* lower *(respectively* upper*)* Riemann sum *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:

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

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

EXERCISE 2.12. Verify that if $f : [-\pi, \pi) \to \mathbb{C}$ is Riemann integrable then the function $g : [-\pi, \pi) \to \mathbb{C}$ defined by $g(\theta) = f(\theta)e^{-in\theta}$ is Riemann integrable. $\Diamond$

Here are some examples of familiar functions that are Riemann integrable and for which we can define Fourier coefficients. See Theorem 2.28 for a complete characterization of Riemann-integrable functions.

EXAMPLE 2.13. Uniformly continuous functions on intervals are Riemann integrable. In particular continuous functions on closed intervals are Riemann integrable; see [**Tao1**, Section 11.5]. Monotone[2] bounded functions on intervals are Riemann integrable; see [**Tao1**, Section 11.6]. $\Diamond$

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

EXAMPLE 2.14. (*Dirichlet's Example*) 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 $[0, 1]$. Notice also that it is discontinuous everywhere. $\Diamond$

EXERCISE 2.15. Show that the Dirichlet function is not Riemann integrable. $\Diamond$

[2]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 $f(x) \geq f(y)$.

2.1.2. Lebesgue integrable functions and L^p-spaces. There are other more general notions of integral. We will consider the *Lebesgue integral*. The reader could consult the book by Robert G. Bartle [**Bart2**] for other notions.

The Lebesgue integral is studied in depth in courses on *measure theory*. For a brief introduction see [**Tao2**, Chapters 18, 19]. For more in-depth presentations at a level intermediate between advanced undergraduate and graduate students see [**SS2**] or [**Bart1**]. Classical graduate textbooks include the books by Gerald B. Folland [**Fol**], and Halsey L. Royden [**Roy**].

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.

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 2.14 is Lebesgue integrable but not Riemann integrable, with Lebesgue integral equal to zero; see the last paragraph in Section 2.1.4. 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*[3] *function* of a set $A \subset \mathbb{R}$ is defined, similarly to the characteristic function of an interval, to be

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

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

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

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

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*, defined for each real number p such that $1 \leq p < \infty$ by

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

and for $p = \infty$ by

$$L^\infty(I) := \{f : I \to \mathbb{C} : f \text{ is essentially bounded}\}.$$

The L in L^p and L^∞ stands for Lebesgue.

All the functions involved are assumed to be measurable. In Definition 2.29 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 functions on I:

$$B(I) := \{f : I \to \mathbb{C} : f \text{ is bounded}\}.$$

There are essentially bounded functions that are not bounded, so $B(I) \subset L^\infty(I)$, but $L^\infty(I)$ is larger.

These spaces are *normed spaces* (see Appendix B for the definition) with the L^p-norm defined to be

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

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

In Definition 2.30 we state precisely what *essential supremum of f* means.

The *L^p-metric* or *L^p-distance d_p* is the metric induced by the L^p-norm, for each p with $1 \leq p \leq \infty$. The L^p-distance between two functions f, $g \in L^p(I)$ is given by the norm of their difference:

$$d_p(f, g) := \|f - g\|_{L^p(I)}. \tag{2.5}$$

A property of normed spaces that we use repeatedly is the *triangle inequality.* In our setting it states that for all f, $g \in L^p(I)$, $1 \leq p \leq \infty$,

$$\|f + g\|_{L^p(I)} \leq \|f\|_{L^p(I)} + \|g\|_{L^p(I)}. \tag{2.6}$$

The Riemann-integrable functions are not only contained in all L^p-spaces for $1 \leq p \leq \infty$, but they are also *dense* in each $L^p(I)$ in its L^p-metric; see Definition 2.59 below. In other words, given any function $f \in L^p(I)$ we can find a Riemann-integrable function arbitrarily close to f in the L^p-metric. Furthermore, for each p the space $L^p(I)$ is the *completion* of $\mathcal{R}(I)$ in the L^p-metric. Being the completion of $\mathcal{R}(I)$ automatically implies that $\mathcal{R}(I)$ is dense in $L^p(I)$, but is a stronger statement. It also implies that the limit of each convergent sequence in $L^p(I)$ must be a function in $L^p(I)$. Finally, we can think of a *Cauchy*[4] *sequence* Cauchy sequence as being a sequence that has the right to converge. Every Cauchy sequence in $L^p(I)$ does converge, and its limit is a function in $L^p(I)$.

---

[4]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 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.

Informally, the *completion* Z of a subspace Y of a metric space X in the metric of X is the smallest subspace of X that contains Y and that also contains all limits of Cauchy sequences in Z.

ASIDE 2.16. *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* [**Tao2**, Exercise 12.4.8].

In fact, we could have *defined* the Lebesgue L^p-spaces as the completion of $\mathcal{R}(I)$ in the L^p-metric, instead of attempting to define the Lebesgue integral via measure theory. The important thing is that both processes lead to the same spaces. This fact is so important that we state it (without proof) as a theorem, for future reference.

THEOREM 2.17. *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*-metric. In particular, every Cauchy sequence in* $L^p(I)$ *converges to a function in* $L^p(I)$.

Therefore the L^p-spaces with the L^p-norms are *complete normed spaces*. Such spaces are so common and important that they have a special name, *Banach*[5] *spaces*. See Appendix B for the definition of a complete normed space, and see Section 2.4 for some results about density of one space in another. One learns about Banach spaces in a course on functional analysis. Classical graduate textbooks include [**Fol**] and the book by Martin Schechter [**Sch**].

Among the L^p-spaces, both L^1 and L^∞ play special roles. The space

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

of *square-integrable* functions 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*

$$\langle f, g\rangle_{L^2(I)} := \int_I f(x)\overline{g(x)} \, dx, \tag{2.7}$$

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*[6] *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. See Appendix B for a more detailed description of inner-product vector spaces and Hilbert spaces. Another reference is [**SS2**, Chapters 4, 5].

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

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

[5]Named after the Polish mathematician Stefan Banach (1892–1945).

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

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 2.18. Verify that

$$\langle e_n, e_m\rangle = \delta_{n,m}.$$

In particular, $\|e_n\|_{L^2(\mathbb{T})} = 1$. In the language of Chapter 5, the trigonometric system is an *orthonormal* system. $\Diamond$

2.1.3. Ladder of function spaces on $\mathbb{T}$. So far in this chapter, 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, all these function spaces are nicely nested. The inclusions are summarized in Figure 2.4 below.

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.)

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 $0 \le j \le k$. This boundary condition implies that when we restrict f to the closed interval $[-\pi, \pi]$ it is C^k, in particular it is 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 $C^\infty(\mathbb{T})$ up to the large space $L^1(\mathbb{T})$:

$$\begin{aligned} &C^\infty(\mathbb{T}) \subset C^k(\mathbb{T}) \subset C^2(\mathbb{T}) \subset C^1(\mathbb{T}) \subset C(\mathbb{T}) \\ &\qquad \subset \mathcal{R}(\mathbb{T}) \subset B(\mathbb{T}) \subset L^\infty(\mathbb{T}) \subset L^{p_1}(\mathbb{T}) \subset L^{p_2}(\mathbb{T}) \subset L^1(\mathbb{T}), \end{aligned} \tag{2.8}$$

for integers $k \ge 2$ and real numbers p_1 and p_2 such that $p_2 < p_1$. In particular, notice that

$$L^2(\mathbb{T}) \subset L^1(\mathbb{T}).$$

Warning: For functions defined on the real line, the L^p-spaces are no longer nested, nor are the continuous functions necessarily bounded or integrable.

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 $N = 0$, 1, 2, ... We denote these spaces by $\mathcal{P}_N(\mathbb{T})$. More precisely,

$$\mathcal{P}_N(\mathbb{T}) := \Big\{ \sum_{|n|\le N} a_n e^{in\theta} : a_n \in \mathbb{C},\ |n| \le N \Big\}.$$

These spaces are clearly nested and increasing as N increases, and they are certainly subsets of $C^\infty(\mathbb{T})$. Thus we can extend the above ladder to one starting at $\mathcal{P}_0(\mathbb{T})$ (the constant functions) and climbing all the way to $C^\infty(\mathbb{T})$, by including the nested spaces

$$\mathcal{P}_0(\mathbb{T}) \subset \mathcal{P}_{N_1}(\mathbb{T}) \subset \mathcal{P}_{N_2}(\mathbb{T}) \subset C^\infty(\mathbb{T}),$$

for all N_1, N_2 with $0 \le N_1 \le N_2$.

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

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

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

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

Exercise 2.19. Compute the L^p-norm of a step function defined on an interval I. Compute the L^p-norm of a simple function defined on an interval I. Use the notation $m(A)$ for the measure of a measurable set $A \subset I$. $\Diamond$

We will encounter other classes of functions on $\mathbb{T}$, $[a,b]$, or $\mathbb{R}$, such as functions that are *piecewise continuous* or *piecewise smooth*, *monotone functions*, functions of *bounded variation*, *Lipschitz*[7] *functions*, and *Hölder*[8]*-continuous 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 in to the ladder of spaces in Figure 2.4, which will start looking more like a tree with branches and leaves.

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

Exercise 2.20. Let I be a bounded interval. Show that if $f : I \to \mathbb{C}$ is bounded by $M > 0$ and $f \in L^p(I)$, then $f \in L^q(I)$ for all q such that $p \le q < \infty$. Furthermore,

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

and so in this situation the L^q-norm is controlled by the L^p-norm. In the language of norms, the preceding inequality can be written as

$$\|f\|_{L^q(I)}^q \le \|f\|_{L^\infty(I)}^{q-p}\,\|f\|_{L^p(I)}^p,$$

since $\|f\|_{L^\infty(I)}$ is the infimum of the numbers M that bound f. $\Diamond$

For the previous exercise you may find *Hölder's inequality* useful.

Lemma 2.21 (Hölder's Inequality). *Suppose I is a bounded interval,* $1 \le s \le \infty$, $\dfrac{1}{s} + \dfrac{1}{t} = 1$, *$f \in L^s(I)$, and $g \in L^t(I)$. Then*

$$\left|\int_I f(x)g(x)\,dx\right| \le \|f\|_{L^s(I)}\|g\|_{L^t(I)}; \tag{2.9}$$

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

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

in other words the integral of the product of f and g is smaller in absolute value than the product of the L^s-norm of f and the L^t-norm of g.

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 celebrated *Cauchy–Schwarz*[9] *inequality*. Inequalities are inherent to the way of analysis, and we will see them over and over again throughout the book.

EXERCISE 2.22. 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

$$\|h\|_{L^p(I)} \leq |I|^{1/p-1/q}\|h\|_{L^q(I)}.$$

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 an exceptional set of points x. Such exceptional sets are small. We now introduce the appropriate notion of smallness, given by the concept of a *set of measure zero*.

DEFINITION 2.23. 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 $\{I_j\}$ 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*. ♢

In particular, every countable set has measure zero. There are uncountable sets that have measure zero. The most famous example is the *Cantor*[10] *set*, discussed in Example 2.25.

EXERCISE 2.24. Show that finite subsets of $\mathbb{R}$ have measure zero. Show that countable subsets of $\mathbb{R}$ have measure zero. In particular the set $\mathbb{Q} \cap \mathbb{R}$ has measure zero. ♢

EXAMPLE 2.25. (*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 $(1/9, 2/9)$ and $(7/9, 8/9)$ of the remaining closed intervals. We are left with four closed intervals: $[0, 1/9]$, $[2/9, 1/3]$, $[2/3, 7/9]$, and $[8/9, 1]$. Delete the four

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

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

middle-third intervals $(1/27, 2/27)$, $(7/27, 8/27)$, $(19/27, 20/27)$, and $(25/27, 26/27)$ of the remaining closed intervals.

Continuing in this way, at the n^{th} stage we are left with 2^n disjoint closed intervals each of length 3^{n-1}. Removing the 2^n middle-third intervals, each of length $1/3^n$, we are left with 2^{n+1} disjoint closed intervals of length $1/3^n$. 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 remain after A is deleted: $C := [0, 1] \setminus A$.

The Cantor set can be seen to coincide with the set of numbers in $[0, 1]$ whose *ternary expansion* (base three) uses only the digits 0 and 2. Once this identification is made, one can also verify that C is in a one-to-one correspondence (except for certain points for which the correspondence is two-to-one) with the interval $[0, 1]$. Therefore C has the same cardinality as $\mathbb{R}$, and so C is an uncountable set. However, C has measure zero, because its complement A has *measure one*[11]: the sum of the lengths of the disjoint intervals removed is

$$m(A) := 1 \times \frac{1}{3} + 2 \times \frac{1}{9} + 4 \times \frac{1}{27} + \cdots = \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = 1.$$

The last equality holds because we are dealing with a geometric sum with ratio $r = 2/3 < 1$. $\Diamond$

EXERCISE 2.26. Verify that the Cantor set C is closed. Use Definition 2.23 to verify that C has measure zero. $\Diamond$

ASIDE 2.27. *We assume that the reader is familiar with basic notions of point set topology on* $\mathbb{R}$*, and in particular with the concepts of* open *and* closed *sets in* $\mathbb{R}$*. To review these ideas, see for example* [**Tao2**, Chapter 12].

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

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

DEFINITION 2.29. 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 2.30. 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 defined to be the smallest constant M such that $|f(x)| \leq M$ a.e. on the interval I. In other words,

$$\operatorname*{ess\,sup}_{x \in I} |f(x)| := \inf\{M \geq 0 : |f(x)| \leq M \text{ a.e. in } I\}.$$

$\Diamond$

[11] 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.

Exercise 2.31. 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 $\operatorname{ess\,sup}_{x\in I}|f(x)| = \sup_{x\in I}|f(x)|$. $\Diamond$

Remark 2.32. 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 2.14 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 $1 \le p \le \infty$.

Now that we have given the relevant definitions, we summarize in Figure 2.4 the ladder of function spaces discussed in Section 2.1.3. In Figure 2.4, 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 $N_1 \le N_2$. 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 2.22).

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

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 seven 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); and in the sense of convergence in L^p. 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. The last three of these modes are defined in terms of integrals of powers of the difference $|g_n - g|$. Some of these modes of convergence are stronger than others, uniform convergence being the strongest. Figure 2.5 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*.

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

FIGURE 2.4. Ladder of nested classes of functions $f : \mathbb{T} \to \mathbb{C}$.

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

$\swarrow$ $\searrow$

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

L^2 *(mean-square) convergence*:
$\int_{\mathbb{T}} |g_n(x) - g(x)|^2 \, dx \to 0.$

$\Downarrow$ $\Downarrow$

Pointwise a.e. convergence:
$|g_n(x) - g(x)| \to 0$
for almost all $x \in \mathbb{T}$.
For each $\varepsilon > 0$ and almost every $x \in \mathbb{T}$
there is an $N = N(x, \varepsilon)$ that works.

L^1 *(mean) convergence*:
$\int_{\mathbb{T}} |g_n(x) - g(x)| \, dx \to 0.$

FIGURE 2.5. Relations between five of the seven 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.

See [**Bart1**, 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 modes of convergence that are defined in terms of individual points x.

DEFINITION 2.33. A sequence of functions $g_n : I \to \mathbb{C}$ *converges pointwise* to $g : I \to \mathbb{C}$ if for each $x \in I$,

$$\lim_{n\to\infty} g_n(x) = g(x).$$

Equivalently, g_n converges pointwise to g if for each $\varepsilon > 0$ and for each $x \in I$ there is an $N = N(\varepsilon, x) > 0$ such that for all $n > N$,

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

$\Diamond$

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

Definition 2.34. A sequence of functions $g_n : I \to \mathbb{C}$ *converges almost everywhere*, or *pointwise a.e.*, to $g : I \to \mathbb{C}$ if it converges pointwise to g except on a set of measure zero:

$$\lim_{n\to\infty} g_n(x) = g(x) \quad \text{a.e. in } I.$$

Equivalently, g_n converges a.e. to g if there exists a set E of measure zero, $E \subset I$, such that for each $x \in I \setminus E$ and for each $\varepsilon > 0$ there is an $N = N(\varepsilon, x) > 0$ such that for all $n > N$,

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

$\Diamond$

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

Definition 2.35. A sequence of bounded functions $g_n : I \to \mathbb{C}$ *converges uniformly* to $g : I \to \mathbb{C}$ if given $\varepsilon > 0$ there is an $N = N(\varepsilon) > 0$ such that for all $n > N$,

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

Equivalently, g_n converges uniformly to g if

$$\lim_{n\to\infty} \sup_{x\in I} |g_n(x) - g(x)| = 0.$$

$\Diamond$

Here is a slightly weaker version of uniform convergence.

Definition 2.36. A sequence of L^∞-functions $g_n : I \to \mathbb{C}$ *converges uniformly outside a set of measure zero*, or *in L^∞*, to $g : I \to \mathbb{C}$ if

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

We write $g_n \to g$ in $L^\infty(I)$. $\Diamond$

Remark 2.37. If the functions g_n and g are continuous, then

$$\|g_n - g\|_{L^\infty(I)} = \sup_{x\in I} |g_n(x) - g(x)|,$$

and in this case convergence in $L^\infty(I)$ and uniform convergence coincide. This is why the L^∞-norm for continuous functions on closed bounded intervals 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 [**Tao2**, Section 4.4]. In particular, if the functions g_n in Definition 2.35 are continuous and converge uniformly, then the limiting function is itself continuous; see Theorem 2.53.

We move to the modes of convergence defined via integrals. The two most important are convergence in L^1 and in L^2; these are special cases of convergence in L^p.

The integral used in these definitions is the Lebesgue integral, and by integrable we mean Lebesgue integrable.

DEFINITION 2.38. A sequence of integrable functions $g_n : I \to \mathbb{C}$ *converges in the mean*, or *in* $L^1(I)$, to $g : I \to \mathbb{C}$ if for each $\varepsilon > 0$ there is a number $N > 0$ such that for all $n > N$

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

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

DEFINITION 2.39. 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 to g if $\lim_{n\to\infty} \|g_n - g\|_{L^2(I)} = 0$. We write $g_n \to g$ in $L^2(I)$. $\Diamond$

Convergence in $L^1(I)$ and in $L^2(I)$ are particular cases of convergence in $L^p(I)$.

DEFINITION 2.40. For p such that $1 \le p < \infty$, a sequence of L^p-integrable functions $g_n : I \to \mathbb{C}$ *converges in* $L^p(I)$ to $g : I \to \mathbb{C}$ if given $\varepsilon > 0$ there exists $N > 0$ such that for all $n > N$

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

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

We have already noted that the L^p-spaces are complete normed spaces, also known as Banach spaces. Therefore the limit functions in Definition 2.40 have no choice but to be in $L^p(I)$ themselves. In particular, in Definitions 2.38 and 2.39 the limit functions must be in $L^1(I)$ and $L^2(I)$ respectively.

EXERCISE 2.41. Verify that uniform convergence on a bounded interval I implies the other modes of convergence on I. $\Diamond$

EXERCISE 2.42. Show that for $1 \le q < p \le \infty$ convergence in $L^p(I)$ implies convergence in $L^q(I)$. **Hint:** Hölder's Inequality. $\Diamond$

The following examples show that in Figure 2.5 the implications go only in the directions indicated. In these examples we work on the unit interval $[0, 1]$ rather than on $\mathbb{T}$.

EXAMPLE 2.43. (*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 or to any other function. See Figure 2.6. $\Diamond$

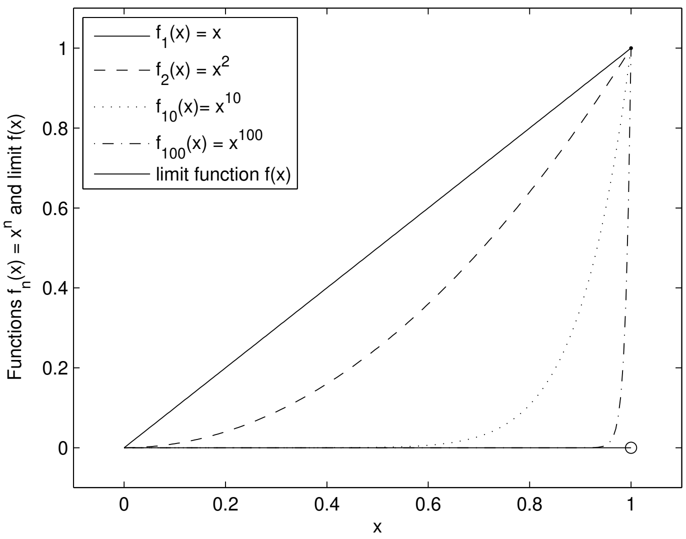


FIGURE 2.6. The functions $f_n(x) = x^n$ converge pointwise, but not uniformly, on $[0,1]$ to the function $f(x)$ that is 0 for $x \in [0,1)$ and 1 at $x = 1$.

EXAMPLE 2.44. (*Pointwise but not in the mean*) The sequence of functions $g_n : [0,1] \to \mathbb{R}$ defined by

$$g_n(x) = \begin{cases} n, & \text{if } 0 < x < 1/n; \\ 0, & \text{otherwise,} \end{cases}$$

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 2.7. $\Diamond$

EXAMPLE 2.45. (*In the mean but not pointwise*) Given a positive integer n, let m, k be the unique non-negative integers such that $n = 2^k + m$ and $0 \leq m < 2^k$. The sequence of functions $h_n(x) : [0,1] \to \mathbb{R}$ defined by

$$h_n(x) = \begin{cases} 1, & \text{if } \dfrac{m}{2^k} < x < \dfrac{m+1}{2^k}; \\ 0, & \text{otherwise,} \end{cases}$$

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 function*. See Figure 2.8. $\Diamond$

Despite these examples, there is a connection between pointwise convergence and convergence in L^p. For p such that $1 \leq p < \infty$, if $f_n, f \in L^p(\mathbb{T})$, and if $f_n \to f$ in L^p, then there is a subsequence $\{f_{n_k}\}$ that converges to f a.e. on $\mathbb{T}$. See [**SS2**, Corollary 2.2].

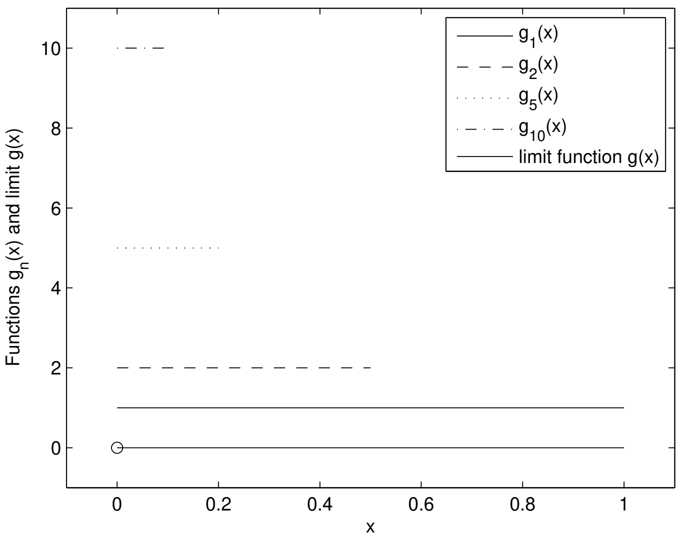


FIGURE 2.7. The functions g_n defined in Example 2.44 converge pointwise, but not in the mean, in L^2, or uniformly.

EXERCISE 2.46. Verify that the sequences described in Examples 2.43, 2.44, and 2.45 do have the convergence properties described there. Adapt the sequences in the examples to have the same convergence properties on the interval $[a, b]$. $\diamond$

EXERCISE 2.47. Where would the two missing modes, namely convergence outside a set of measure zero and L^p convergence for a given p with $1 \leq p \leq \infty$, fit in Figure 2.4? Prove the implications that involve these two modes, and devise examples to show that other potential implications involving these two modes do not hold. $\diamond$

2.3. Interchanging limit operations

Uniform convergence on closed bounded intervals is a very strong form of convergence which is amenable to interchange with other limiting operations and properties defined by limits, such as integration, differentiation and continuity. We list here without proof several results from advanced calculus that will be used throughout the book. Some results do involve uniform convergence as one of the limit operations, some others involve other limit operations (like interchanging two integrals).

THEOREM 2.48 (Interchange of Limits and Integrals). *If f_n converges uniformly to f in $[a, b]$ then*

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

In other words,

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

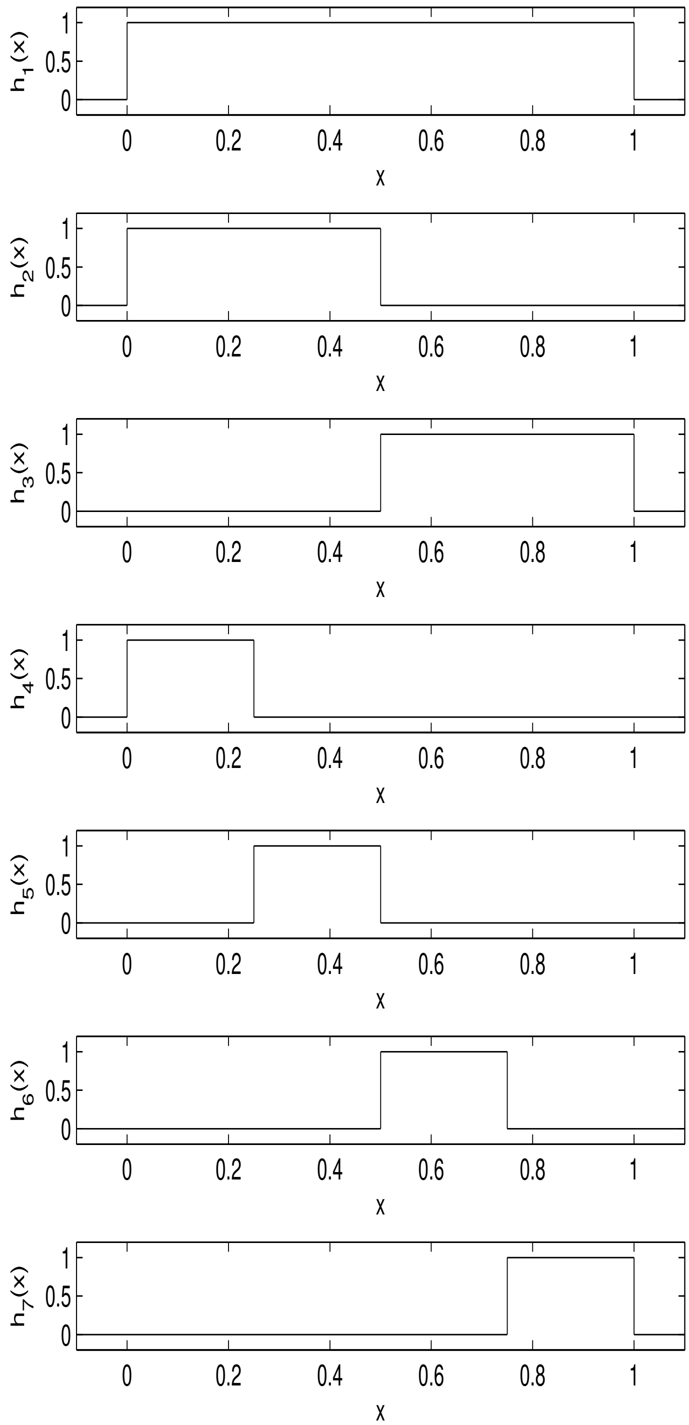


FIGURE 2.8. The functions $h_n(x)$ defined in Example 2.45 converge in mean, but do not converge pointwise at any x.

Notice that the functions in Example 2.44 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

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

EXERCISE 2.49. Given a closed bounded interval I and 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

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

$\Diamond$

We can interchange infinite summation and integration provided the convergence of the partial sums is uniform.

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

THEOREM 2.50 (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 2.51. Show that the hypotheses in Theorem 2.50 are necessary. $\Diamond$

EXERCISE 2.52. What are the necessary hypotheses on a series $\sum_{n=0}^{\infty} g_n(x)$ so that it is legal to interchange sum and differentiation? In other words, when can we guarantee that

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

$\Diamond$

Here is another classical and useful result in the same vein, already mentioned in Remark 2.37.

THEOREM 2.53. *The uniform limit of continuous functions on $[a,b]$ is continuous on $[a,b]$.*

Notice that this is the same as saying that if $g_n \to g$ uniformly, and the functions g_n are continuous, then for all $x_0 \in [a,b]$,

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

Perhaps inserting a few more equalities will make the above statement clear:

$$\begin{aligned}\lim_{x\to x_0}\lim_{n\to\infty} g_n(x_0) &= \lim_{x\to x_0} g(x)\\ &= g(x_0)\\ &= \lim_{n\to\infty} g_n(x_0)\\ &= \lim_{n\to\infty}\lim_{x\to x_0} g_n(x).\end{aligned}$$

Example 2.43 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[12] Theorem, to Theorem 2.53. It says that the pointwise limit of a decreasing sequence of functions is also their uniform limit.

Theorem 2.54 (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 2.55. Prove Dini's Theorem, and show that hypothesis (ii) is essential. A continuous version of Example 2.44, where the "steps" are replaced by "tents", might help. ♢

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[13]. Here is one such theorem, valid for continuous functions.

Theorem 2.56 (Fubini's Theorem for Continuous Functions). *Let I and J be closed intervals. 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 2.56 that is stated and proved in every measure theory book. You can find a two-dimensional version in [**Tao2**, Section 19.5].

Interchanging limit operations is a delicate maneuver that cannot always be accomplished. We illustrate throughout the book settings where this interchange is allowed and settings where it is illegal, such as the ones just described involving uniform convergence. If we insist on carrying on the interchange, false conclusions could be deduced.

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

---

[12] Ulisse Dini, Italian mathematician (1845–1918).

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

THEOREM 2.57 (Lebesgue Dominated Convergence Theorem). *Consider a sequence of measurable functions f_n defined on the interval I, converging pointwise to a function f. Suppose there exists a dominating function $g \in L^1(I)$, meaning that $|f_n(x)| \leq g(x)$ a.e. for each $n > 0$. Then*

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

EXERCISE 2.58. Verify that the functions in Example 2.44 cannot be dominated by an integrable function g. $\Diamond$

2.4. Density

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

DEFINITION 2.59. 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

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

$\Diamond$

One can deduce from the definition of the Riemann-integrable functions on a bounded interval I that the step functions (see Definition 2.3) are dense in $\mathcal{R}(I)$ with respect to the L^1-norm; see Proposition 2.10. We state this result as a theorem, for future reference.

THEOREM 2.60. *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 2.61. We can choose the step functions that approximate a given Riemann-integrable function to lie entirely below the function, or entirely above it.

In Chapter 4 we will learn that

> *Trigonometric polynomials can approximate uniformly continuous functions on $\mathbb{T}$.*

This is the celebrated Weierstrass theorem (see Theorem 3.4), proved below.

In Chapter 9 we will learn that on closed bounded intervals,

> *Step functions can approximate uniformly continuous functions, despite the fact that step functions are not continuous.*

It is not hard to deduce directly from the fact that continuous functions can be approximated uniformly by step functions that they can also be approximated in the L^p-norm by step functions.

> *Step functions can approximate continuous functions in the L^p-norm.*

FIGURE 2.9. The graph shows a step function whose corners have been rounded.

What is perhaps surprising is that

Continuous functions can approximate step functions pointwise and in the L^p-norm.

We verify this result in Chapter 4, when we introduce the concepts of convolution and approximations of the identity. If you draw a picture of a step function and then round off the corners, this result seems more reasonable (see Figure 2.4).

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

THEOREM 2.62. *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. □

REMARK 2.63. We will use a very precise version of this result for $p = 1$ in Chapter 4 (see Lemma 4.8), 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 $f \in \mathcal{R}(I)$ and $\varepsilon > 0$, there exists a continuous function g on I (by Theorem 2.62) 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 $|g(x)| \leq M$, and $g(x) = M$ otherwise. 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 2.64. Show that the function g_0 defined in Remark 2.63 satisfies the properties claimed there. ◊

EXERCISE 2.65. 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 2.66. (*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, in Section 2.1 we learned that $C(I) \subset L^2(I) \subset L^1(I)$. Suppose we knew that the continuous functions are dense in $L^2(I)$ (with respect to the L^2-norm), and the square-integrable functions are dense in $L^1(I)$ (with respect to the L^1-norm). Even so, we could not use transitivity to conclude that the continuous functions are dense in $L^1(I)$. We need to know more about the relation between the L^1- and the L^2-norms. In fact, to complete the argument, it suffices to know that $\|f\|_{L^1(I)} \leq C\|f\|_{L^2(I)}$ (see Exercise 2.22), 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 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*.

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 below, 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; see Theorem 2.17. 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 2.67. *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 2.68. Given $f \in L^p(I)$, $1 \leq p < \infty$, and $\varepsilon > 0$, and assuming all the theorems and statements in this section, show that there exists a step function h on I such that

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

$\Diamond$

CHAPTER 3

Pointwise convergence of Fourier series

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

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

3.1. Pointwise convergence: why do we care?

For the periodic ramp function in Example 1.11, the Fourier expansion converges *pointwise* to $f(\theta)=\theta$ everywhere except at the odd multiples of π, where it converges to zero (the halfway height in the vertical jump). For the toy model of a voice signal in Example 1.1, we were able to reconstruct the signal perfectly using only four numbers (two frequencies and their corresponding amplitudes), because the signal was very simple. In practice, complicated signals such as real voice signals transmitted through phone lines require many more amplitude coefficients and frequencies for accurate reconstruction, sometimes thousands or millions. Should we use all of these coefficients in reconstructing our function? Or perhaps we could get a reasonable approximation to our original signal using only a subset of the coefficients? If so, which ones should we keep? One simple approach is to truncate at specific frequencies $-N$ and N, and to use only the coefficients corresponding to frequencies such that $|n|\leq N$. Instead of the full Fourier series, we obtain the N^{th} *partial Fourier sum of* f,

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

which is a trigonometric polynomial of degree N.

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

For the periodic ramp function f in Example 1.11, the partial Fourier sums are

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

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

Exercise 3.1. Use the following Matlab script [***script has to be made] 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

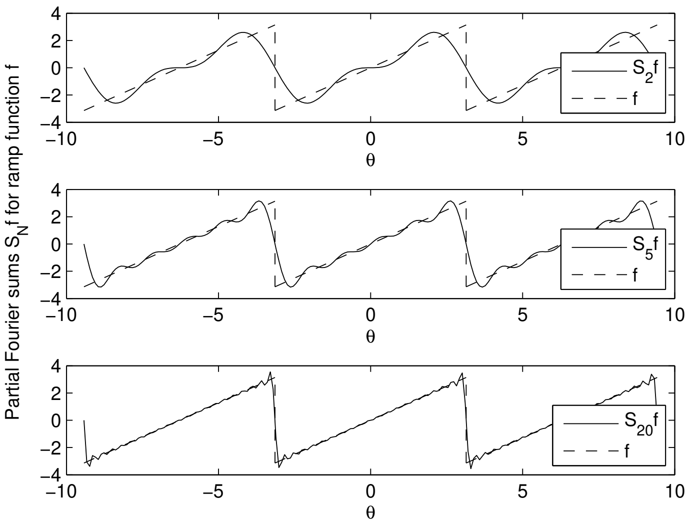


FIGURE 3.1. Plots of the periodic ramp function f and its partial Fourier sums $S_N f$, for $N = 2$, 5, and 20.

N increases. This is the so-called *Gibbs*[1] *phenomenon*. Repeat the exercise for the *periodic step function*

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

¡¡¡¡¡¡ .mine See Project **??** for more explorations. ♢======= Project 12.4 outlines a more detailed investigation of the Gibbs phenomenon. ♢¿¿¿¿¿¿¿ .r75

EXERCISE 3.2. Compute the partial Fourier sums for the trigonometric polynomial in Exercise 1.10. ♢

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 $S_M f = S_{M+1} f = \cdots = f$. See Exercise 1.17.

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

The partial Fourier sums $S_N f(\theta)$ sums converge uniformly to $f(\theta)$ for all trigonometric polynomials f.

EXERCISE 3.3. Verify that the Taylor polynomials $P_N(f, 0)$ converge uniformly for all polynomials f. See (1.4) and Exercise 1.5. $\Diamond$

Approximating with partial Fourier sums presents some problems. For example, there are continuous functions for which one does not get pointwise convergence 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 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 is allowed to combine the Fourier coefficients in ways different from the partial sums $S_N f$. For example, averaging over the partial Fourier sums provides a smoother truncation method in terms of the *Cesáro*[4] *means*,

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

that we will study more carefully in Section 4.5. In particular, we will prove Fejér's[5] Theorem 4.27 that we quote now.

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

Therefore the following *uniqueness principle* holds:

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

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

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

THEOREM 3.4 (Weierstrass's Approximation Theorem for trigonometric polynomials). *Continuous functions 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* [**Tao2**, 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].

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

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

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

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

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

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

3.2. Smoothness vs convergence

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

3.2.1. A first convergence result.

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. 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; afterwards we argue that the limit must be $f(\theta)$.

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

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

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

PROOF. The partial Fourier sums of f are given by

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

Fix θ, and notice that the Fourier series of f evaluated at θ is a power series in the 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.

We can find new expressions for the Fourier coefficients $\widehat{f}(n)$, for $n \neq 0$, by integrating by parts twice:

$$\begin{aligned}
\widehat{f}(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta \qquad \text{(by parts with } u = f(\theta),\ v = -\frac{1}{in}e^{-in\theta}\text{)} \\
&= \frac{1}{2\pi}\left[-\frac{1}{in}f(\theta)e^{-in\theta}\right]_{-\pi}^{\pi} - \frac{1}{2\pi}\int_{-\pi}^{\pi} -\frac{1}{in}f'(\theta)e^{-in\theta}\,d\theta \\
&= \frac{1}{2\pi}\frac{1}{in}\int_{-\pi}^{\pi} f'(\theta)e^{-in\theta}\,d\theta \\
&= \frac{1}{in}\widehat{f'}(n) \\
&= -\frac{1}{n^2}\widehat{f''}(n) \qquad \text{(by the same argument applied to } f'\text{)}.
\end{aligned}$$

The integration by parts is permissible since f is assumed to have two continuous derivatives. The boundary term in the second line vanishes because of 2π-periodicity. We conclude that

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

We are now ready to make a direct estimate of the coefficients:

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

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

$$|\widehat{f}(n)| \leq \frac{C}{n^2}.$$

Therefore, by the comparison test, $\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|$ converges, and by the *Weierstrass M-Test*, the Fourier series $\sum_{n=-\infty}^{\infty} \widehat{f}(n)e^{inx}$ does converge uniformly and hence pointwise if $f \in C^2(\mathbb{T})$. □

ASIDE 3.7. *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 Theorem 3.6 as follows.

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

REMARK 3.8. We have not yet established whether the Fourier series evaluated at θ converges to $f(\theta)$ or to some other number.

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 3.9. *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*

$$\lim_{n\to\infty} \frac{a_1 + a_2 + \cdots + a_n}{n} = \lim_{n\to\infty} a_n.$$

EXERCISE 3.10. Prove Lemma 3.9. Give an example to show that if the sequence of averages converges, that does not imply that the original sequence converges. $\lozenge$

In Section 4.5 we show that for continuous functions f on $\mathbb{T}$, the Cesáro means $\sigma_N f(\theta)$, see (3.2), which are averages of the partial Fourier sums, do converge uniformly to $f(\theta)$ as $N \to \infty$. Lemma 3.9 shows that since for $f \in C^2(\mathbb{T})$ the partial Fourier sums converge (Theorem 3.6), then the Cesáro sums also converge, and to the same limit. It follows that if $f \in C^2(\mathbb{T})$, then the partial sums $S_N f(\theta)$ of the Fourier series for f do converge pointwise to $f(\theta)$, for each $\theta \in \mathbb{T}$.

Summing up, we have established the following theorem (modulo Fejér's Theorem 4.27).

THEOREM 3.11. *If $f \in C^2(\mathbb{T})$, then its Fourier series converges uniformly to f on $\mathbb{T}$.*

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

THEOREM 3.12. *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,

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

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 $\mathbb{T}$:

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

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: $|\widehat{f^{(k)}}(n)| < C$. What we just have shown is the following,

Theorem 3.13. *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 3.14. 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 $\mathbb{T} \setminus [0, 1]$, 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. ◊

Example 3.15. In Section 4.3 we introduce the concept of *convolution* of 2π-periodic functions, and its properties. Among the most important properties is its interplay with the Fourier transform: *The n^{th}-Fourier coefficient of the convolution of two functions is the product of the n^{th} Fourier coefficients of the functions convolved.* See Theorem 4.7 (v). This is another instance of the time–frequency dictionary on $\mathbb{T}$. See Table 4.1, and Section 7.3 for the analogous dictionary on $\mathbb{R}$. Denote by g_k the convolution of $\chi_{[0,1]}$ with itself k times; see Example 4.5 and Exercise 4.6. By the rule just described,

$$|\widehat{g_k}(n)| = \left|\widehat{\chi}_{[0,1]}(n)\right|^k \leq (\pi n)^{-k}, \tag{3.4}$$

so the Fourier coefficients decay like n^{-k}. However, it can be shown that g_k is not C^{k-1}. ◊

The function g_k described in Example 3.15 is not C^{k-1}. However, it follows from Corollary 3.17 that g_k is C^{k-2} for $k \geq 2$. (For a direct verification at least for $k = 2$ and 3, see Exercise 4.6.)

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 $k \geq 2$, then f is C^ℓ, where $\ell = k - 2$ if $k \in \mathbb{Z}$, 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 $\ell = 0$, the hypothesis is $\sum_{n\in\mathbb{Z}} |\widehat{f}(n)| < \infty$, which in turn implies the uniform convergence of the partial Fourier sums $S_N f$ to some periodic and continuous function F. Féjer's Theorem 4.27 shows that the averages of the partial Fourier sums, the Cesàro means $\sigma_N f$, see (3.2), converge uniformly to $f \in C(\mathbb{T})$, finally Exercise 3.10 implies that $F = f$, hence the partial Fourier sums converge uniformly to f.

For the case $\ell = 1$, see Exercise 3.19.

Now prove the general case by induction on ℓ. □

EXERCISE 3.19. Let $f : \mathbb{T} \to \mathbb{C}$ be continuous. Assume 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

$$\lim_{N\to\infty} \sum_{|n|\leq N} in\widehat{f}(n)e^{in\theta} = f'(\theta).$$

In other words, $h = f'$. Hint: Theorem 2.50 about interchanging limits and differentiation might be useful. ♢

In this section we learned that the Fourier coefficients of smooth functions go to zero as $|n| \to \infty$. The question is, at what rate? We found that

> *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

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

where we have made the change of variable $\alpha = \theta - 1/(2\pi n)$. But since the integrand is 2π-periodic, and we are integrating over an interval of length 2π, we can shift the integral back to $\mathbb{T}$ without altering its value, and conclude that

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

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

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

Hence
$$|\widehat{f}(n)| \leq \frac{1}{4\pi}\int_{-\pi}^{\pi}\left|f(\theta) - f\left(\theta + \frac{\pi}{n}\right)\right| d\theta.$$
We can now take the limit as $|n| \to \infty$, and since f is uniformly continuous on $[-\pi, \pi]$, the sequence $g_n(\theta) := \left|f(\theta) - f\left(\theta + \frac{\pi}{n}\right)\right|$ converges uniformly to zero. Therefore we can interchange the limit and the integral (see Theorem 2.48), obtaining
$$0 \leq \lim_{|n|\to\infty}|\widehat{f}(n)| \leq \frac{1}{4\pi}\int_{-\pi}^{\pi}\lim_{|n|\to\infty}\left|f(\theta) - f\left(\theta + \frac{\pi}{n}\right)\right| d\theta = 0,$$
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 $g_n(\theta) := \left|f(\theta) - f\left(\theta + \frac{\pi}{n}\right)\right|$ converge uniformly to zero. ◊

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

3.2.3. Differentiability vs rate of convergence of partial Fourier sums. It can be shown that

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

More precisely, it can be proven that if $f \in C^k(\mathbb{T})$ for $k \geq 2$, then there exists a constant $C > 0$ independent of θ such that
$$|S_N f(\theta) - f(\theta)| \leq \frac{C}{N^{k-1}}. \tag{3.5}$$

If $f \in C^1(\mathbb{T})$, then there exists a constant $C > 0$ independent of θ such that
$$|S_N f(\theta) - f(\theta)| \leq \frac{C}{\sqrt{N}}. \tag{3.6}$$

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

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 3.22 for $k = 2$. It is not difficult to modify that example for the case $k > 2$.

Example 3.22. (*Plucked String*) Consider the function $f : \mathbb{T} \to \mathbb{R}$ given by $f(\theta) = \pi/2 - |\theta|$, whose graph is shown in Figure 3.2. This function is continuous but not differentiable at $\theta = 0, \pm\pi$. 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 3.23. Compute the Fourier coefficients of the plucked string function in Example 3.22. Verify that the inequalities (3.7) hold. Hint: It might help to remember the following inequalities

$$\frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)} \leq \frac{1}{n^2} \leq \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}.$$

The left- and right-hand sides in the string of inequalities, when added up, form telescoping series. This observation allows you to obtain the following bounds:

$$\frac{1}{N} \sum_{n \geq N} \frac{1}{n^2} \leq \frac{1}{N-1}.$$

If this hint is not enough, see [**Kor**, Example 10.1, p.35]. $\Diamond$

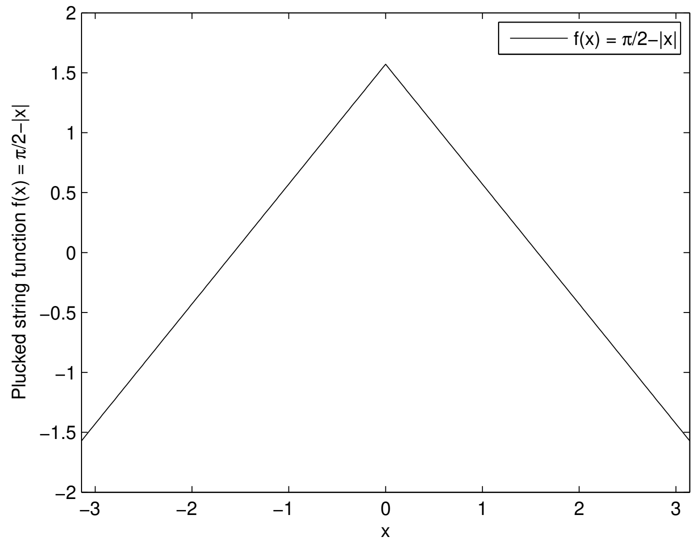


FIGURE 3.2. The plucked string function f.

EXERCISE 3.24. Construct an example of a function that satisfies equation (3.5) but is not C^{k-1}. $\Diamond$

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 try to 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 might be a good time to revisit Chapter 2. In this section we will use notation and concepts introduced in Chapter 2, such as sets of measure zero, almost everywhere convergence, Lebesgue-integrable functions and convergence in L^p.

Project 12.6 investigates in more detail some of the results below.

THEOREM 3.25 (Dirichlet, 1829). *Suppose that, except perhaps at finitely many points, $f : \mathbb{T} \to \mathbb{C}$ is continuous and its derivative f' is continuous and bounded. Then for all $\theta \in \mathbb{T}$,*

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

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

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

THEOREM 3.26 (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 later in the book in Section 9.3.4. For a constructive proof see [**Kor**, Chapter 18, Theorem 18.1]; see also [**SS1**, Chapter 3, Section 2.2].

A reproduction of Dirichlet's paper can be found in [**KL**, Chapter 4], as well as a comparative analysis with Jordan's[7] result that we quote next.

THEOREM 3.27 (Jordan, 1881). *If f is a function of* bounded variation[8] *on $\mathbb{T}$, then the Fourier series of f converges to $\frac{1}{2}[f(\theta^+) + f(\theta^-)]$ 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 3.25 imply bounded variation, and so Jordan's Theorem follows from Dirichlet's Theorem. A proof of Dirichlet's Theorem can be found in [**Kor**, 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.

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

[8]This means that 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, so are monotone functions, and you can compute precisely their variation.

Theorem 3.28 (Dini, 1878). *Let f be a 2π-periodic function. Suppose that for some θ there exists a $\delta > 0$ such that*

$$\int_{|t|<\delta} \left| \frac{f(\theta+t) - f(\theta)}{t} \right| dt < \infty;$$

in particular f must be continuous at θ. Then

$$\lim_{N\to\infty} S_N f(\theta) = f(\theta).$$

A 2π-periodic function f is said to satisfy 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 $C^\alpha(\mathbb{T})$. It is clear that $C(\mathbb{T}) \subset C^\alpha(\mathbb{T}) \subset C^\beta(\mathbb{T})$ for all $0 < \alpha \leq \beta \leq 1$. If $f \in C^\alpha(\mathbb{T})$ for $0 < \alpha \leq 1$, then it satisfies uniformly Dini's condition 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 [**SS1**, Chapter 3, Exercises 15–16]. A proof of Dini's result can be found in [**Graf**, Theorem 3.3.6, p.189].

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

Theorem 3.29 (Riemann Localization Principle). *For an integrable function f, the convergence of the Fourier series to $f(\theta)$ only depends on the values of f in a neighborhood of θ.*

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

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

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

Theorem 3.30 (Kolmogorov, 1926). *There is a Lebesgue-integrable function $f : \mathbb{T} \to \mathbb{C}$ such that*

$$\limsup_{N\to\infty} |S_N f(\theta)| = \infty \qquad \textit{for all } \theta \in \mathbb{T}.$$

Recall that the minimal requirement on f so that we can compute its Fourier coefficients is that f is Lebesgue-integrable $\big(f \in L^1(\mathbb{T})\big)$. In that case $S_N f$ is a well-defined trigonometric polynomial for each $N \in \mathbb{N}$. 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 points θ. Carleson's result is one of the deepest theorems in analysis in the twentieth century.

THEOREM 3.31 (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 $\mathbb{T}$.

This result 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 3.32 (Kahane[9] and 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*

$$\limsup_{N\to\infty} |S_N f(\theta)| = \infty \qquad \text{for all } \theta \in E.$$

The proof of Theorem 3.32 can be found in Katznelson's classical book *An introduction to harmonic analysis* [**Kat**, Chapter 3]. The same chapter includes a construction of a Kolmogorov-type example.

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

THEOREM 3.33 (Hunt, 1967). *Carleson's Theorem holds for all $f \in L^p(\mathbb{T})$ and for all $p > 1$.*

It took almost 40 years for the Carleson-Hunt Theorem to make its way into graduate textbooks in more detail than an annotation similar to ours. There are now several more accessible accounts of this famous theorem, for instance in [**Graf**] and in [**Ari**].

Now we see where to draw the line between those functions for which the partial Fourier series converge pointwise for all x to $f(x)$, 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 will get convergence almost everywhere, and given any set of measure zero there is a continuous function whose partial Fourier series diverges on that set.

As Körner writes [**Kor**, p.75],

[9] Jean-Pierre Kahane, French mathematician, 1926–.

[10] Yitzhak Katznelson, Israeli mathematician, 1934–.

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.

CHAPTER 4

Summability methods: Dirichlet kernels and Fejér kernels

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 4.1), and *approximations of the identity*, also known as *good kernels* (Sections 4.4 and 7.5). Roughly speaking, an approximation of the identity is a sequence of functions whose mass is concentrating near the origin (very much like a delta function[1]); a precise definition is given below. We discuss these techniques in detail in the context of 2π-periodic functions. Later we will revisit these ideas in the context of functions defined on the line (Section 7.5).

Here we describe several kernels that arise naturally in the theory of Fourier series. Some of them (the Fejér and Poisson[2] kernels) are good kernels that generate approximations of the identity; another (the Dirichlet kernel) 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.

For a captivating account of the history surrounding the use of summability methods for Fourier series, see [**KL**, Chapter 7].

4.1. Partial Fourier sums and the Dirichlet kernel D_N

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 notations 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

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

where the Fourier coefficients are calculated using the formula

$$a_n := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) e^{-iny} dy.$$

[1]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.

[2]Named after the French mathematician Simeón Denis Poisson (1781–1840).

Inserting the formula for a_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 $f(y)$:

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

Here we were led to define the function

$$D_N(\theta) := \sum_{|n|\leq N} e^{in\theta}. \tag{4.1}$$

This function D_N is called the *Dirichlet kernel*. 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 D_N is 2π-periodic.

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

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

We study convolution in more detail in Section 4.3. With these definitions, we have shown that the N^{th} partial Fourier sum of f is given by

$$S_N f(\theta) = (f * D_N)(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_N(\theta - y)\, dy. \tag{4.2}$$

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

ASIDE 4.1. *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. We do use in our proofs the assumption that the functions are bounded. If the functions are Riemann integrable then they are bounded by definition, but this is not true for Lebesgue-integrable functions. In those instances where we relied on boundedness a different argument is required to verify the result for all Lebesgue-integrable functions, and we point the reader to appropriate references.*

[3] Riemann-integrable functions on bounded intervals are closed under translations, reflections and products. Hence $\mathcal{R}(\mathbb{T})$ 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.

4.2. Properties of 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

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

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

Next, we obtain a convenient and concise closed formula for D_N as a quotient of two sines:

$$\begin{aligned} D_N(\theta) &= \sum_{|n|\leq N} e^{in\theta} \\ &= e^{-iN\theta} \sum_{n=0}^{2N} (e^{i\theta})^n \\ &= e^{-iN\theta}\,\frac{(e^{i\theta})^{2N+1} - 1}{e^{i\theta} - 1} \qquad \text{(sum of a finite geometric series)} \\ &= \frac{e^{i(N+1)\theta} - e^{-iN\theta}}{e^{i\theta} - 1} \cdot \frac{e^{-i\theta/2}}{e^{-i\theta/2}} \\ &= \frac{e^{i(2N+1)\theta/2} - e^{-i(2N+1)\theta/2}}{e^{i\theta/2} - e^{-i\theta/2}}, \end{aligned}$$

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

$$D_N(\theta) = \frac{\sin\left((2N+1)\theta/2\right)}{\sin\left(\theta/2\right)}. \tag{4.3}$$

The quotient of sines has a removable singularity at the origin $\theta = 0$. 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 $\mathbb{T}$.

In Figure 4.1 we see that the oscillations of D_N, as $\sin\left((2N+1)\theta/2\right)$ varies between 1 and -1, lie within the envelope $\pm 1/\sin\left(\theta/2\right)$. Also, D_N takes some positive and some negative values, D_N is even, and D_N achieves its maximum value of $2N + 1$ at $\theta = 0$.

THEOREM 4.2. *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*:*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(\theta)\, d\theta = 1. \tag{4.4}$$

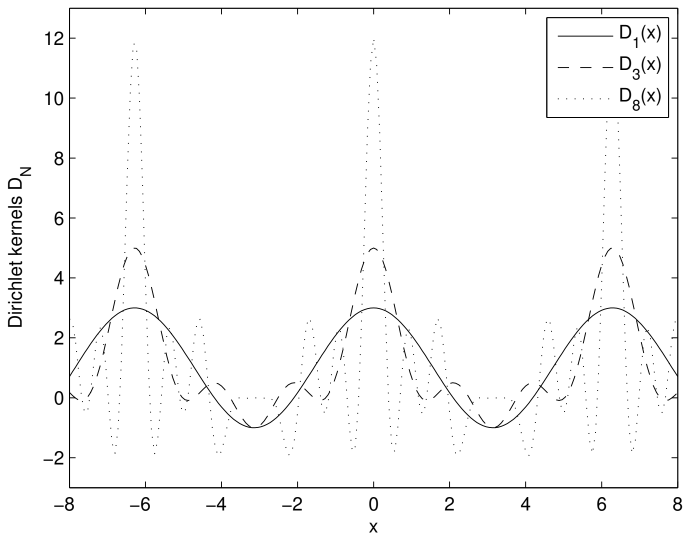


FIGURE 4.1. Graphs of Dirichlet kernels D_N for $N = 1$, 3, and 8.

(iv) *However, the integral of the absolute value of D_N depends on N, and in fact*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |D_N(\theta)|\,d\theta \approx C\,\log N. \tag{4.5}$$

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.

PROOF. We have already established properties (i) and (ii). For property (iii), note that

(4.6)

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(\theta)\,d\theta = \frac{1}{2\pi}\left[\int_{-\pi}^{\pi}\sum_{|n|\le N} e^{in\theta}\right]d\theta = \sum_{|n|\le N}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{in\theta}\,d\theta\right] = 1,$$

since the integrals in the second last term are zero unless $n = 0$, and 1 in that case.

Verifying property (iv) is left to the reader; see Exercise 4.3. See Figure 4.2 for numerical evidence. □

EXERCISE 4.3. Justify equation (4.5) analytically. It may be useful to remember that by the integral test,

$$\sum_{N=1}^{N}\frac{1}{n} \sim \int_1^N \frac{1}{x}\,dx = \log N.$$

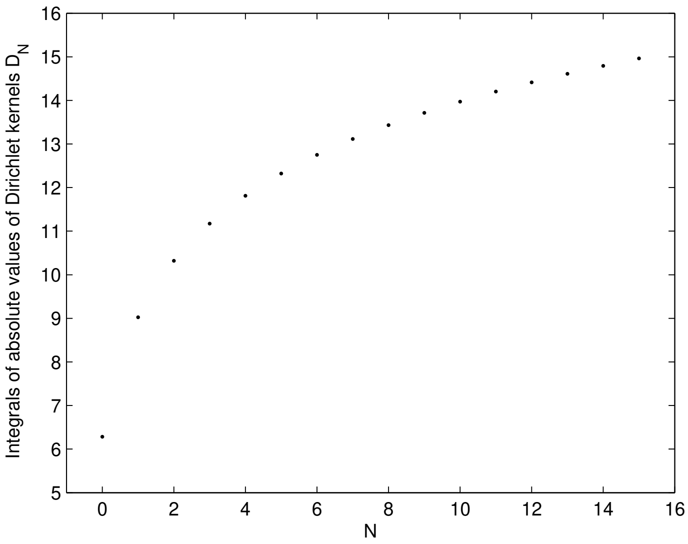


FIGURE 4.2. The value of the integral of the absolute value of the Dirichlet kernel D_N, for $0 \leq N \leq 15$. Note the logarithmic growth.

$\Diamond$

4.3. 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 4.4. 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

$$(f * g)(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)g(\theta - y)\, dy. \tag{4.7}$$

$\Diamond$

Notice that 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. See Figures 4.3 and 4.4.

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. See Section 2.1.1. This last statement is not 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

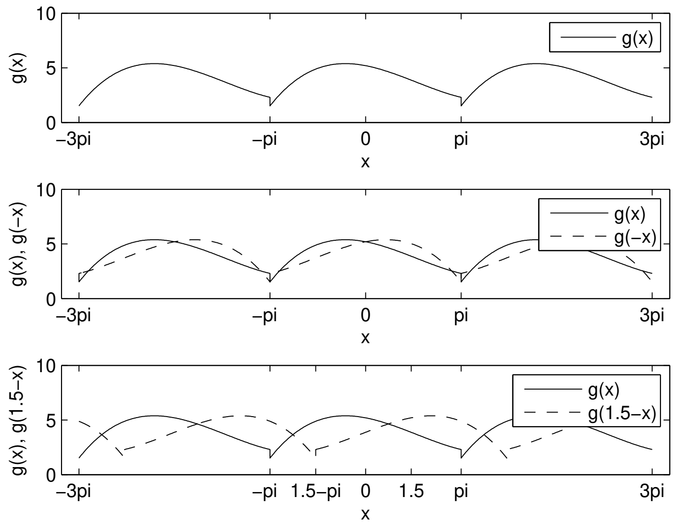


FIGURE 4.3. Reflection and translation. Graphs over $[-3\pi, 3\pi]$ of a function $g(x)$, its reflection $g(-x)$, and its translated reflection $g(1.5 - x)$. To clarify their 2π-periodic nature, the functions are graphed over $[-3\pi, 3\pi]$ rather than over $[-\pi, \pi]$.

inequality[4] holds:

$$\|f * g\|_{L^1(\mathbb{T})} \leq \|f\|_{L^1(\mathbb{T})} \|g\|_{L^1(\mathbb{T})}. \tag{4.8}$$

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

EXAMPLE 4.5. (*Convolving with a Characteristic Function*) We compute a convolution in the special case where one of the convolution factors is the characteristic function $\chi_{[a,b]}(x)$ of a closed interval $[a, b]$ in $\mathbb{T}$. See Figure 4.5. Recall that the characteristic function is defined by

$$\chi_{[a,b]}(\theta) = \begin{cases} 1, & \text{if } \theta \in [a, b]; \\ 0, & \text{otherwise.} \end{cases} \tag{4.9}$$

Using the observation that

$$a \leq \theta - y \leq b \qquad \Longleftrightarrow \qquad \theta - b \leq y \leq \theta - a,$$

[4]This is an instance of Young's inequality; see Chapters 7 and 11.

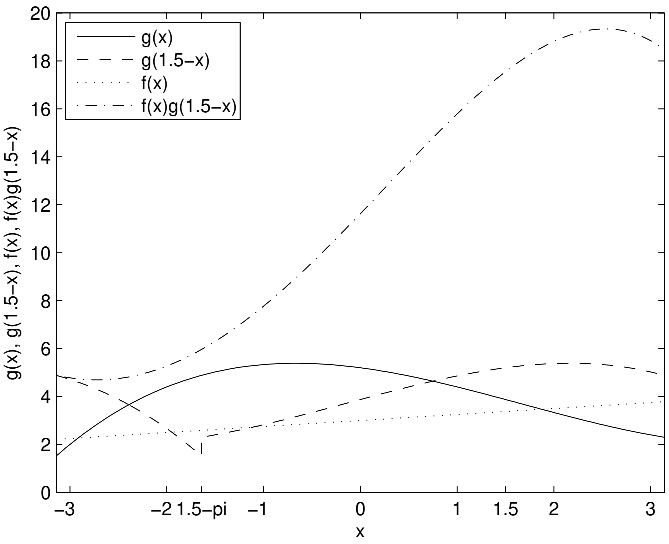


FIGURE 4.4. The ingredients of a convolution. Graphs over $[-\pi, \pi]$ of the same function $g(x)$ as in Figure 4.3, its translated reflection $g(1.5-x)$, another function $f(x)$, and the product $f(x)g(1.5-x)$. The value $f * g(1.5)$ of the convolution $f * g(t)$ at $t = 1.5$ is the area under the graph of $f(x)g(1.5-x)$, divided by $2 * \pi$.

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 a multiple of the average value of f on the reflected and translated interval $[\theta - b, \theta - a]$. ◊

EXERCISE 4.6. (*Convolving Again and Again*) Let $f = \chi_{[0,1]}$. Compute $f * f$ and $f * f * f$. Use the following MATLAB script [***script has to be made] to plot f, $f * f$, and $f * f * f$. Notice how the smoothness improves as we take more convolutions. Recall Example 3.15. ◊

This is an example of the so-called *smoothing properties* of convolution. See Section 4.3.2.

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

THEOREM 4.7. *Let f, g, $h : \mathbb{T} \to \mathbb{C}$ be 2π-periodic integrable functions, and let $c \in \mathbb{C}$ be a constant. Then*

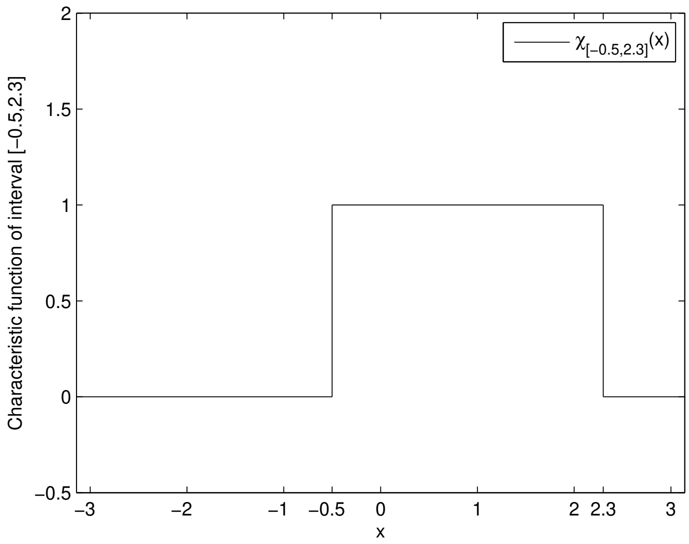


FIGURE 4.5. Graph of the characteristic function $\chi_{[a,b]}$ of the interval $[a,b]=[-0.5,2.3]$.

(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 4.8. *Suppose $f:\mathbb{T}\to\mathbb{C}$ is an integrable and bounded function. Let B be a bound for f. In other words, $|f(\theta)|\le B$ 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) *$\sup_{\theta\in\mathbb{T}}|f_k(\theta)|\le B$ 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 already met in Section 2, that the f_k converge to f in $L^1(\mathbb{T})$.

[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 already 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 basis and wavelet bases.

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 2.62, in particular Remark 2.63.

LEMMA 4.9. *Suppose $f, g : \mathbb{T} \to \mathbb{C}$ are 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 4.8. 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 4.9 after we prove Theorem 4.7.

PROOF OF THEOREM 4.7. Property (i), rewritten slightly, says that

$$\begin{aligned} (f * g)(\theta) &:= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) g(\theta - y)\, dy \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta - y) g(y)\, dy, \end{aligned} \tag{4.10}$$

so that 'the $\theta - y$ can go with g *or* with f'. Equation (4.10) 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, coupled with the commutativity Property (i), one deduces the linearity in the first variable as well. Hence convolution 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 2.56 for continuous functions, obtaining

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

In the second last line we changed variables to $\theta' = \theta - y$. Since the integrand is 2π-periodic we may keep the limits of integration unchanged.

We have established property (v) for continuous functions. Now suppose f and g are integrable and bounded. Take sequences $\{f_k\}$ and $\{g_k\}$ of continuous functions such that $f_k \to f$, $g_k \to g$, as in the Approximation Lemma (Lemma 4.8).

First, $\widehat{f_k}(n) \to \widehat{f}(n)$ for each $n \in \mathbb{Z}$, since by Lemma 4.8(ii)

$$|\widehat{f_k}(n) - \widehat{f}(n)| = \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} [f_k(\theta) - f(\theta)]\, e^{-in\theta}\, d\theta \right| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_k(\theta) - f(\theta)|\, d\theta \to 0$$

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,

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

for each $n \in \mathbb{Z}$ and each $k \in \mathbb{N}$.

Third, we show that as $k \to \infty$

$$\widehat{f_k * g_k}(n) \to \widehat{f * g}(n)$$

for each $n \in \mathbb{Z}$. For

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

Here we can move the limit inside the integral since, by Lemma 4.9, the integrands converge uniformly on $\mathbb{T}$ (see Theorem 2.48).

Hence

$$\widehat{f * g}(n) = \lim_{k\to\infty} \widehat{f_k * g_k}(n) = \left(\lim_{k\to\infty} \widehat{f_k}(n)\right) \left(\lim_{k\to\infty} \widehat{g_k}(n)\right) = \widehat{f}(n)\, \widehat{g}(n),$$

as required.

For Lebesgue-integrable functions, the L^1-version of Fubini's Theorem is needed to carry out the calculation we did for continuous functions. We then 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. $\square$

Exercise 4.10. (*Convolution is Associative*) Prove property (iv). $\diamond$

Proof of Lemma 4.9. We show that $f_k * g_k \to f * g$ uniformly on $\mathbb{T}$. Recall that f_k, g_k and f, g are all periodic functions bounded by $B > 0$ on $\mathbb{T}$, and that

$\int_{-\pi}^{\pi} |f_k - f| \to 0$, and $\int_{-\pi}^{\pi} |g_k - g| \to 0$ as $k \to \infty$ by Lemma 4.8, hence

$$\begin{aligned}
&|(f_k * g_k)(\theta) - (f * g)(\theta)| \\
&\quad \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_k(y) g_k(\theta - y) - f(y) g(\theta - y)|\, dy \\
&\quad \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_k(y)|\, |g_k(\theta - y) - g(\theta - y)|\, dy \\
&\qquad\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_k(y) - f(y)|\, |g(\theta - y)|\, dy \\
&\quad \leq \frac{B}{2\pi} \left(\int_{-\pi}^{\pi} |g_k(\theta - y) - g(\theta - y)|\, dy + \int_{-\pi}^{\pi} |f_k(y) - f(y)|\, dy \right) \\
&\quad = \frac{B}{2\pi} \left(\int_{-\pi}^{\pi} |g_k(y) - g(y)|\, dy + \int_{-\pi}^{\pi} |f_k(y) - f(y)|\, dy \right) \\
&\quad \to 0
\end{aligned}$$

as $k \to \infty$. The last integrals are independent of θ because of the periodicity of g and g_k, and so the convergence is uniform. Also, the second inequality is a consequence of the *Triangle Inequality*[6] for complex numbers and the *additivity*[7] of the integral. □

4.3.2. 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 4.13). (This is how in theory you would like marriages to be: the marriage combines the smoothness of each partner to produce an even smoother couple).

The following result is a precursor of that principle. We start with at least one continuous function and the 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 not necessarily continuous, and their convolution will be continuous.

Lemma 4.11. *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. Given $f \in C(\mathbb{T})$, and g integrable on $\mathbb{T}$. First, f is uniformly continuous. That is, given $\varepsilon > 0$ there is $\delta > 0$ such that if $|h| < \delta$, then $|f(\theta + h) - f(\theta)| < \varepsilon$ for all $\theta \in \mathbb{T}$. Second, by the linearity of the integral,

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

[6]Namely, for a, $b \in \mathbb{C}$, $|a + b| \leq |a| + |b|$.

[7]Additivity means that if f, $g \in L^1(\mathbb{T})$, then $\int_{-\pi}^{\pi} (f + g) = \int_{-\pi}^{\pi} f + \int_{-\pi}^{\pi} g$.

Third, using the triangle inequality for integrals[8], and the uniform continuity of f, we conclude that for all $|h| < \delta$

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

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

ASIDE 4.12. *Let f and g be integrable and bounded functions, and let f_k, g_k continuous and bounded functions approximating f and g like in Lemma 4.8. Then $f_k * g_k$ is continuous (by Lemma 4.11), and $f*g$ is the uniform limit of continuous functions (by Lemma 4.9). We conclude, by Theorem 2.53 that $f*g$ is continuous.*

We have shown (see Aside 4.12)

> *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 4.13. (*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:

$$(f*g)^{(k+m)} = f^{(k)} * g^{(m)}.$$

It suffices to assume that the functions $f^{(k)}$ and $g^{(m)}$ are bounded and integrable to conclude that $f * g \in C^{k+m}(\mathbb{T})$.

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. ◊

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

EXERCISE 4.14. (*Convolution with a Trigonometric Polynomial Yields a Trigonometric Polynomial.*) Show that if f is integrable and $P \in \mathcal{P}_N$ then $f * P \in \mathcal{P}_N$. ◊

Property (v) in Theorem 4.7 is another instance of the time–frequency dictionary. Earlier, 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 4.15. (*Translation Corresponds to Modulation.*) Prove that the process of computing Fourier coefficients converts *translation* to *modulation.* In other

[8] That is, the fact that $|\int_T f| \leq \int_T |f|$.

TABLE 4.1. A time–frequency dictionary for Fourier series. Note that f is treated as periodic.

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 $\widehat{\tau_h f}(n) = e^{-isn}\widehat{f}(n)$

words, if f is 2π-periodic, and integrable on $\mathbb{T}$, and the translation operator τ_h is defined by

$$\tau_h f(\theta) := f(\theta - h) \qquad \text{for } h \in \mathbb{R},$$

then $\tau_h f$ is 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 $n \in \mathbb{Z}$. ◊

We summarize the *time–frequency dictionary* for Fourier coefficients of periodic functions in Table 4.1.

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.

4.4. Good kernels, or approximations of the identity

DEFINITION 4.16. 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) The K_n all have mean value 1:

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

(b) There is some $M > 0$ such that

$$\int_{-\pi}^{\pi} |K_n(\theta)|\, d\theta \leq M \qquad \text{for all } n \in \mathbb{N}.$$

(c) For each $\delta > 0$,

$$\int_{\delta \leq |\theta| \leq \pi} |K_n(\theta)|\, d\theta \to 0 \qquad \text{as } n \to \infty.$$

◊

For brevity, we may simply say that 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 4.20 below, we explain why.

Condition (b) says that the mean values 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 4.17. *Some authors use a more restrictive definition of approximations of the identity, replacing conditions* (b) *and* (c) *by*

(b') K_n *is positive:*

$$K_n(\theta) \geq 0 \qquad \text{for all } \theta \in \mathbb{T}, \text{ for all } n \in \mathbb{N}.$$

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

EXERCISE 4.18. 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 condition (c') implies condition (c). $\Diamond$

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

EXERCISE 4.19. Suppose K is a continuous function on $\mathbb{R}$ that is zero for all $|\theta| \geq \pi$. Assume $\int_{-\pi}^{\pi} |K(\theta)|\, d\theta = 1$. 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 $\mathbb{T}$. $\Diamond$

Why is a family of kernels satisfying conditions (a), (b), and (c) called *good*? In the context of Fourier series, 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 4.5. 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 4.20. *Let* $\{K_n(\theta)\}_{n=1}^{\infty}$ *be a family of good kernels,* $K_n : \mathbb{T} \to \mathbb{R}$, *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 4.13. The second part of the Theorem 4.20 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 4.14 The second part of the Theorem 4.20 will say that we can approximate uniformly continuous functions by trigonometric polynomials, and Weierstrass' Theorem 3.4 will be proved.

FIGURE 4.6. Plot of $K_n(y)$ for large n - mass concentrates at origin. Plot of $K_n(\theta - y)$ - mass concentrates at θ. Plot of $f(y)K_n(\theta - y)$ illustrating how the integral only 'sees' the part of f near $y = \theta$.

Before proving Theorem 4.20, let us indulge in a flight of fancy. Suppose for a moment that the Dirichlet kernel D_N were a good kernel. (Aside: 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 4.20 would imply that the Fourier partial sums $S_N f$ of a continuous function f on $\mathbb{T}$ would converge to f, since we would have

$$S_N f(\theta) = \sum_{|n| \le N} \widehat{f}(n) = (f * D_N)(\theta) \to f(\theta)$$

at all points $\theta \in \mathbb{T}$.

We already 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 sums, discussed in Section 4.5 below, and already mentioned in Chapter 3.

EXERCISE 4.21. (*The Dirichlet Kernel is Not a Good Kernel*) Determine which of the properties of a good kernel fail for the Dirichlet kernel D_N. $\Diamond$

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 4.6.

Here is a model example.

EXAMPLE 4.22. (*A Good Kernel Made up of Characteristic Functions*) Consider the kernels

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

They are pictured in Figure 4.7 for several values of n.

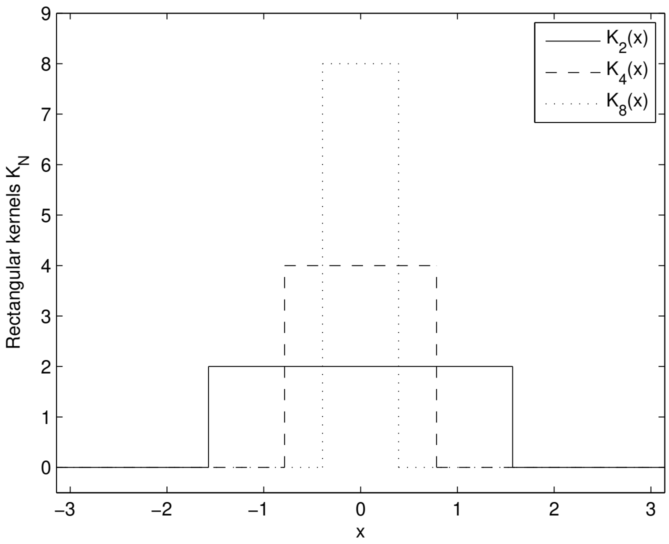


FIGURE 4.7. Graphs of rectangular kernels $\widetilde{K_2}$, $\widetilde{K_4}$, and $\widetilde{K_8}$.

It is straightforward to check that these functions $\widetilde{K_n}$ are a family of good kernels. Further, for a fixed $\theta \in \mathbb{T}$,

$$\begin{aligned}(f * \widetilde{K_n})(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)\, n\chi_{[-\pi/n,\pi/n]}(\theta - y)\, dy \\ &= \frac{n}{2\pi}\int_{\theta-(\pi/n)}^{\theta+(\pi/n)} f(y)\, dy \\ &= \text{average value of } f \text{ on } \left[\theta - \frac{\pi}{n}, \theta + \frac{\pi}{n}\right],\end{aligned}$$

as for convolution with the characteristic function in Example 4.5. 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 $F'(\theta) = f(\theta)$. ♢

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

PROOF OF THEOREM 4.20. 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

$$|f(\theta - y) - f(\theta)| < \varepsilon \qquad \text{whenever } |y| < \delta. \tag{4.11}$$

Now

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

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.

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 $\delta \leq |y| \leq \pi$:

$$\begin{aligned}
|(K_n * f)(\theta) - f(\theta)| &\leq \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)\,[f(\theta - y) - f(\theta)]\,dy\right| \\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |K_n(y)|\,|f(\theta - y) - f(\theta)|\,dy \\
&= \frac{1}{2\pi}\int_{|y|\leq\delta} |K_n(y)|\,|f(\theta - y) - f(\theta)|\,dy \\
&\quad + \frac{1}{2\pi}\int_{\delta\leq|y|\leq\pi} |K_n(y)|\,|f(\theta - y) - f(\theta)|\,dy. \qquad (4.12)
\end{aligned}$$

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

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

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

Condition (c) on good kernels implies that for our δ,

$$\int_{\delta\leq|y|\leq\pi} |K_n(y)|\,dy \to 0$$

as $n \to \infty$. 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 second integral in the last line of equation (4.12), for $n \geq N$, by

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

Hence for $n \geq N$, we have

$$|(K_n * f)(\theta) - f(\theta)| \leq (M + 2B)\, \varepsilon/(2\pi). \tag{4.13}$$

This proves the first part of the theorem.

Assume now that f is continuous on $\mathbb{T}$, then it is uniformly continuous. This means that fixed $\varepsilon > 0$ there exists $\delta > 0$ such that (4.11) holds for all $\theta \in \mathbb{T}$. We can repeat verbatim the previous argument and conclude that the estimate (4.13) holds for all $\theta \in \mathbb{T}$, that is $K_n * f \to f$ uniformly. □

EXERCISE 4.23. Check that if F is continuously differentiable at θ and $h > 0$, then

$$\lim_{h \to 0} \frac{F(\theta + h) - F(\theta - h)}{2h} = F'(\theta).$$

◊

For an arbitrary integrable function f, the limit as n goes to infinity of the convolution with the good kernel $\tilde{K}_n$ does pick out the value of f at points of continuity θ (in other words, at all points θ where f is continuous). The L^1-version of this averaging limiting result goes by the name *Lebesgue Differentiation Theorem.*

THEOREM 4.24 (Lebesgue Differentiation Theorem). *If $f \in L^1(\mathbb{T})$ then*

$$\lim_{h \to 0} \frac{1}{2h} \int_{\theta - h}^{\theta + h} f(y) \, dy = f(\theta) \qquad \text{for a.e. } \theta \in \mathbb{T}.$$

Proofs of this result require a thorough understanding of measure theory. For example, see [**SS2**, Chapter 3, Section 1.2].

4.5. 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 about ten years earlier. 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

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

where a_n is the n^{th}-Fourier coefficient of f. 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 4.14). Interchanging the order of summation, and counting the number of appearances of each Fourier summand $a_k e^{ik\theta}$ 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,

$$\sigma_N f(\theta) = \sum_{|k| \le N-1} \frac{N - |k|}{N} a_k e^{ik\theta}.$$

The *Fejér kernel* $F_N(\theta)$ is defined to be the average of the Dirichlet kernels,

$$F_N f(\theta) := \frac{D_0(\theta) + D_1(\theta) + \cdots + D_{N-1}(\theta)}{N}. \tag{4.14}$$

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 $|k| \le N$, where the coefficients corresponding to smaller frequencies are weighted most heavily. More precisely,

$$F_N f(\theta) = \sum_{|n| \le N} \left(\frac{N - |n|}{N} \right) e^{in\theta}.$$

What is not so obvious is that there is a closed formula for the Fejér kernel, namely

$$F_N f(\theta) = \frac{1}{N} \left[\frac{\sin(N\theta/2)}{\sin(\theta/2)} \right]^2. \tag{4.15}$$

EXERCISE 4.25. Verify (4.15). You might want to use the closed formula (4.3) for the Dirichlet kernel. ♢

Compare with formula (4.3) for the Dirichlet kernel in terms of sines. Notice that, unlike the Dirichlet kernel, the Fejér kernel is non-negative. See Figure 4.8.

EXERCISE 4.26. (*The Fejér kernel is a Good Kernel*) Verify that the Fejér kernel F_N is a good kernel, and that

$$\sigma_N f = F_N * f.$$

♢

By Theorem 4.20 on good kernels, it follows that the Cesàro sums of f converge to f wherever f is continuous.

THEOREM 4.27 (Fejér, 1900). *If $f : \mathbb{T} \to \mathbb{C}$ is integrable and bounded, then*

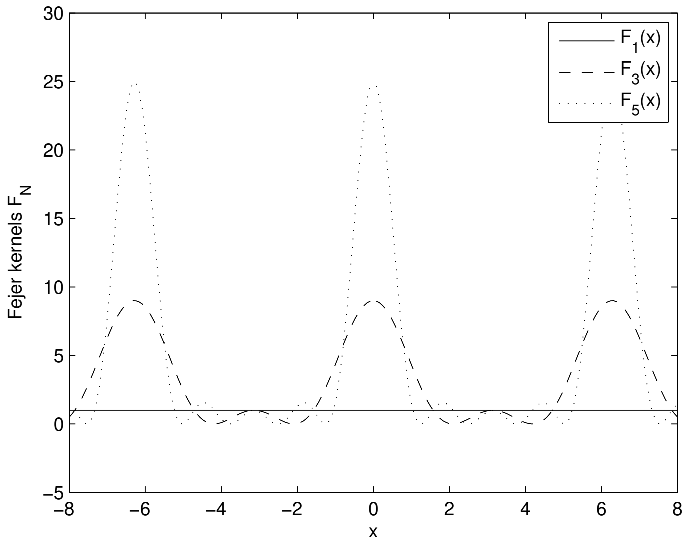


FIGURE 4.8. Graphs of Fejér kernels $F_N(x)$ for $N = 1$, 3, and 5. **Maybe lower horizontal line to x-axis.

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 celebrated 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. More precisely, if $f, g \in C(\mathbb{T})$ and $\widehat{f}(n) = \widehat{g}(n)$ for all $n \in \mathbb{Z}$, then $f = g$. It suffices to verify the above statement for $g = 0$, which we did at the end of Section 3.1.

EXERCISE 4.28. The uniqueness principle for continuous functions on the circle states that

If $f \in C(\mathbb{T})$ and all its Fourier coefficients vanish, then $f = 0$.

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, $f(\theta) = \lim_{N\to\infty} \sigma_N f(\theta) = 0$. We reproduce the argument in this setting: suppose $f : \mathbb{T} \to \mathbb{C}$ is Riemann-integrable, and all its Fourier coefficients are zero, then all the partial Fourier sums are zero, and so are the Cesàro means, $\sigma_N f = 0$. But the Cesàro means converge to f, at all points of continuity, and we are done. Lebesgue's Theorem 2.28 says that f is Riemann integrable if and only if it 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 4.29 (Uniqueness Principle). *If $f \in L^1(\mathbb{T})$ and $\widehat{f}(n) = 0$ for all n, then $f = 0$ a.e.*

4.6. 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, 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 [**SS1**, Sections 5.3, 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 4.30. The *Poisson kernel* $P_r(\theta)$ is defined for $r \in [0, 1)$ by

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

$\Diamond$

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. We are interested now in the behavior as r increases towards 1, instead of $N \to \infty$. The Poisson kernel is non-negative (for $r \in [0, 1)$). The series for $P_r(\theta)$ is absolutely convergent and uniformly convergent.

EXERCISE 4.31. Verify that the two formulae in our definition of the Poisson kernel are actually equal. $\Diamond$

EXERCISE 4.32. (*The Poisson Kernel is a Good Kernel*) Verify that the Poisson kernel $P_r(\theta)$ is a good kernel. (Modify the definition of *good kernel* to take account of the change in index from $n \in \mathbb{N}$ to $r \in [0, 1)$.) $\Diamond$

DEFINITION 4.33. Let f be a function with Fourier series $\sum_{n\in\mathbb{Z}} a_n e^{in\theta}$. The r^{th}-*Abel mean* of f is defined by

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

for $r \in [0, 1)$. $\Diamond$

Thus the Abel mean is formed from f by multiplying the Fourier coefficients a_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.

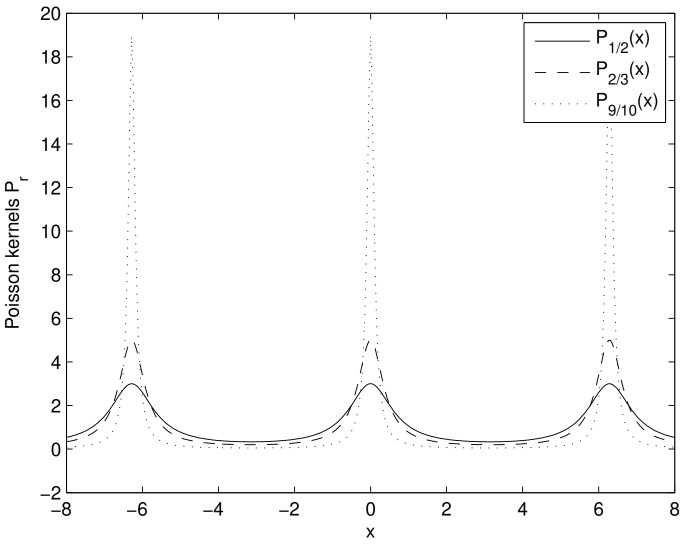


FIGURE 4.9. Graphs of Poisson kernels $P_r(x)$ for $r = 1/2$, $2/3$, and $9/10$.

EXERCISE 4.34. Verify that for integrable functions $f : \mathbb{T} \to \mathbb{C}$,

$$A_r f(\theta) = (P_r * f)(\theta). \tag{4.16}$$

♢

It follows from Theorem 4.20 (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

$$\begin{aligned} f(\theta) &= \lim_{r \to 1^-} (f * P_r)(\theta) \\ &= \lim_{r \to 1^-} A_r f(\theta). \end{aligned}$$

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 the convergence is uniform.

EXERCISE 4.35. To plot the kernels $\widetilde{K}_n$ and the Féjer kernels $\{F_n\}$, we used the following MATLAB script [****script has to be written]. Modify the script to plot the Poisson kernels. Hint: replace the continuous parameter r by the discrete parameter $r_n = 1 - 1/n$. ♢

4.7. 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. For example, convergence in $L^2(\mathbb{T})$, and also in $L^p(\mathbb{T})$.

THEOREM 4.36. *If $f \in C(\mathbb{T})$, $\{K_n\}_{n\geq 1}$ is a family of good kernels, then*

$$\lim_{n\to\infty} \|f * K_n - f\|_{L^p(\mathbb{T})} = 0.$$

EXERCISE 4.37. Prove Theorem 4.36. Just notice that you can control the L^p-norm with the L^∞-norm on $\mathbb{T}$; see Exercise 2.20. Recall the relations between different modes of convergence illustrated in Figure 2.5. ♢

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 4.38. *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-convergence of the partial Fourier sums of f. Since we have information about the mean-square convergence of the Cesàro means, if we had an analogue to Lemma 3.9 in $L^2(\mathbb{T})$, and we knew that $S_N f$, the partial Fourier sums, converge in L^2-sense, we could deduce indirectly, as we did in Section 3.2.1 that they have no other choice than to converge to f, the limit of the Cesàro sums. We will prove that $\|S_N f - f\|_{L^2(\mathbb{T})} \to 0$, and Theorem 4.38 will play a rôle, however we will not follow the route sketched in this paragraph. We will first understand better 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".

Just for fun, here is the analogue of Lemma 3.9 in L^p.

LEMMA 4.39. *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of functions in $L^p(\mathbb{T})$. Assume the sequence is convergent in $L^p(\mathbb{T})$ to f (since $L^p(\mathbb{T})$ is complete, the limit function f must be in $L^p(\mathbb{T})$). Then the sequence of averages*

$$F_n(\theta) = \frac{f_1(\theta) + \cdots + f_n(\theta)}{n}$$

converges in $L^p(\mathbb{T})$ and to the same limit function f.

EXERCISE 4.40. Prove Lemma 4.39. ♢

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 4.41 (Riemann-Lebesgue Lemma for Integrable Functions). *If $f \in L^1(\mathbb{T})$, then*

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

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 2.67. In particular their Fourier coefficients are close: for each fixed n,

$$|\widehat{f}(n) - \widehat{f_k}(n)| \le \frac{1}{2\pi}\|f_k - f\|_{L^1(\mathbb{T})} \to 0 \qquad \text{as } k \to \infty. \tag{4.17}$$

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 for all $k > K$

$$\|f_k - f\|_{L^1(\mathbb{T})} < \pi\varepsilon.$$

Hence, by inequality (4.17), for all $n > 0$ and all $k > K$

$$|\widehat{f}(n) - \widehat{f}_k(n)| < \varepsilon/2.$$

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 $|n| > N$,

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

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. □

Since $L^p(\mathbb{T}) \subset L^1(\mathbb{T})$, the same must hold for functions in $L^p(\mathbb{T})$. One could run a direct proof in complete analogy to the proof for $L^1(\mathbb{T})$, provided (4.17) holds with L^1-norm replaced by L^p-norm, which is the content of the next exercise.

EXERCISE 4.42. Prove that if $f \in L^p(\mathbb{T})$ and $f_k \in C(\mathbb{T})$ with $\|f_k - f\|_{L^p(\mathbb{T})} \to 0$ as $|n| \to \infty$, then for $p \neq 2$

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

For $p = 1$ this proves (4.17). For $p = 2$ the constant is 1 instead of $(2\pi)^{1/2}$, because of the normalization constant in the definition of the L^2-norm. Hint: Hölder's Inequality 2.21 may be helpful. ◊

CHAPTER 5

Mean-square convergence of Fourier series

This chapter has two main themes: convergence in the L^2-sense of the partial Fourier sums $S_N f$ to f, and the completeness of the trigonometric functions in $L^2(\mathbb{T})$. They are intimately related. 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 have learned that continuity of f is not sufficient to ensure pointwise convergence; we need something more. In particular we showed that if f has two continuous derivatives, then we get uniform convergence. If instead of taking the partial Fourier sums we take their averages (Cesàro sums), then we get uniform convergence for all continuous functions.

In this chapter we discuss *mean-square convergence*, also known as *convergence in* $L^2(\mathbb{T})$. Among other things, we will learn that the N^{th}-partial Fourier sum $S_N f$ is the trigonometric polynomial of degree less than or equal to N which is *closest* to f in the L^2-norm. In fact it is the *orthogonal projection* onto the subspace of $L^2(\mathbb{T})$ consisting of N^{th}-degree trigonometric polynomials. In particular, this holds for continuous functions on $\mathbb{T}$ and more generally for functions in $L^2(\mathbb{T})$. The *energy* or L^2*-norm* of the function can be recovered from just the knowledge of the Fourier coefficients via the celebrated *Parseval's Identity*, which is an infinite-dimensional analogue of the Pythagorean Theorem.

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* $L^2(\mathbb{T})$.

We can see immediately that the trigonometric functions are orthonormal on $\mathbb{T}$, because

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{in\theta}e^{-im\theta}\,d\theta = \delta_{n,m},$$

see (1.7). What is really noteworthy here is that the family of trigonometric functions (with discrete integer values of n) 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 $e^{in\theta}$.

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 function $f : \mathbb{T} \to \mathbb{C}$ such that

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

The integration here is in the Lebesgue sense, but for most purposes we can think of it as Riemann integration.

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 non-negative, 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 this space, let us take a step ahead and describe 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

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

THEOREM 5.1. *If $f \in L^2(\mathbb{T})$, then its partial Fourier sums converge to f in the L^2-sense, meaning that*

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

We can restate the mean-square convergence theorem in terms of norms. Namely, 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[1] shows.

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

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

We prove Theorems 5.1 and 5.2, and more, in the course of this chapter. Parseval's Identity immediately gives us an insight into the behavior of Fourier coefficients of (at least) 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 lemma.

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

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

EXERCISE 5.4. Deduce Corollary 5.3 from Parseval's Identity. Note that we proved this result for the larger class $L^1(\mathbb{T})$ in Lemma 4.41. ◊

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

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

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

$$\|\{a_n\}\|_{\ell^2(\mathbb{Z})} = \left(\sum_{n=-\infty}^{\infty} |a_n|^2 \right)^{1/2}.$$

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

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

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

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.

In Appendix B, you will find the definition 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[2]. The Fourier mapping also preserves the inner products. In fact, the Fourier mapping is an *isometry* between these two Hilbert spaces. In Section 5.2 we describe the space $L^2(\mathbb{T})$ in more detail than we did in Chapter 2.

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

We denote by $L^2(\mathbb{T})$ the space of square-integrable functions on $\mathbb{T}$. The space $L^2(\mathbb{T})$ 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, but not all, inner-product vector spaces satisfy another property, 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, while $\mathcal{R}(\mathbb{T})$ with the L^2-inner product is not a Hilbert space.

DEFINITION 5.5. A sequence of functions $f_n \in L^2(\mathbb{T})$ is a Cauchy sequence if for every $\varepsilon > 0$ there exists $N > 0$ such that $\|f_n - f_m\|_{L^2(\mathbb{T})} < \varepsilon$ for all n, $m > N$. $\Diamond$

We say that a sequence of functions $\{f_n\}$ 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 is converging to a limit, then as $n \to \infty$ the functions in the sequence are getting closer to each other, in the sense that they form a *Cauchy sequence*. The converse is true in the case of $L^2(\mathbb{T})$, a remarkable fact that we state without proof now (see [**SS2**] for example).

THEOREM 5.6. *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 $L^2(\mathbb{T})$.*

[2]For this statement to be true in the case of $L^2(\mathbb{T})$, we really need the Lebesgue integral. If we insist on using only the Riemann integral, then we still get an inner-product vector space. However, that space is not complete: there are Cauchy sequences of Riemann square-integrable functions that converge to functions that are not Riemann square integrable, but are necessarily Lebesgue square integrable. See Section 5.2.

Notice that here we are using the word "complete" with a different meaning than in Definition 5.21 where we will consider "complete systems of orthonormal functions". It will be clear from the context whether we are talking about a "complete space" or a "complete system $\{f_n\}$ of functions".

More can be said.

THEOREM 5.7. *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 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.*

EXAMPLE 5.8. The space of square-summable sequences $\ell^2(\mathbb{Z})$ is also an *inner-product vector space* with inner-product

$$\langle \{a_n\}, \{b_n\} \rangle_{\ell^2(\mathbb{Z})} = \sum_{n=-\infty}^{\infty} a_n \overline{b_n}.$$

Moreover, $\ell^2(\mathbb{Z})$ is complete. Therefore it is a Hilbert space. ◊

EXERCISE 5.9. Prove that $\ell^2(\mathbb{Z})$ is complete. ◊

Recall that Lebesgue square-integrable functions on $\mathbb{T}$ are always Lebesgue integrable on $\mathbb{T}$. In other words, $L^2(\mathbb{T}) \subset L^1(\mathbb{T})$[3]; see Section 2.1. On the other hand, there are functions in $L^1(\mathbb{T})$ that are not in $L^2(\mathbb{T})$. Such functions must be unbounded. In fact if f is a bounded function in $L^1(\mathbb{T})$, then f is in $L^2(\mathbb{T})$; see Exercise 2.20. In particular, Riemann-integrable functions are necessarily bounded, and are Lebesgue integrable, and so they are in $L^2(\mathbb{T})$. That is, $\mathcal{R}(\mathbb{T}) \subset L^2(\mathbb{T})$. Thus in each theorem in this section we could replace $f \in L^2(\mathbb{T})$ by $f \in \mathcal{R}(\mathbb{T})$, obtaining a less general theorem that does not involve the notion of Lebesgue integration.

5.2.1. Three key results involving orthogonality. *Orthogonality* is one of the most important ideas in the context of inner-product vector spaces. It generalizes, to infinite-dimensional spaces, the familiar idea of two perpendicular vectors in the plane.

We recall here the definitions for the specific cases of $L^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$ that we are studying. In Appendix B we recall the definitions in the setting of a general inner-product vector space. The reader will find that 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 5.10. Two functions f, $g \in L^2(\mathbb{T})$ are *orthogonal* if $\langle f, g\rangle = 0$. Sometimes we use the notation $f \perp g$ to indicate that f and g are orthogonal. 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 of 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 $L^2(\mathbb{T})$. In other words, not only are they an orthogonal set, but each element in the set has L^2-norm equal to one (the vectors have been "normalized"). Checking orthonormality

[3]**Warning:** The analogous statement for $\mathbb{R}$ is false. The function $f(x) = (1 + |x|)^{-1}$ is not integrable on $\mathbb{R}$, but it is square integrable on $\mathbb{R}$. So $L^2(\mathbb{R}) \not\subset L^1(\mathbb{R})$.

of the trigonometric system is equivalent to verifying that

$$\langle e_k, e_m \rangle = \delta_{k,m}. \tag{5.3}$$

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

THEOREM 5.11 (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 5.12 (Cauchy–Schwarz Inequality). *For all f, $g \in L^2(\mathbb{T})$,*

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

THEOREM 5.13 (Triangle Inequality). *For all f, $g \in L^2(\mathbb{T})$,*

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

EXERCISE 5.14. Prove Theorems 5.11, 5.12, and 5.13. In particular, show that $\|f\|_{L^2(\mathbb{T})} = \sqrt{\langle f, f \rangle}$ is a norm on $L^2(\mathbb{T})$. ♢

EXERCISE 5.15. 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 L^2 and the inner product (or in language that we will meet later in the book, that the inner product is *continuous*).

PROPOSITION 5.16. *Fix $g \in L^2(\mathbb{T})$. If a sequence of L^2-functions $\{f_n\}_{n \in \mathbb{N}}$ converges to f in L^2-sense, then*

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

In other words, we can interchange the limit and the inner product with g.

PROOF. Assume 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,

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

We conclude that

$$\lim_{n \to \infty} \langle g, f_n - f \rangle = 0,$$

and so

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

by the linear properties of the inner product. This proves the proposition. □

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

COROLLARY 5.17. *Fix $g \in L^2(\mathbb{T})$. Suppose $f = \sum_{k=1}^{\infty} h_k$ in the L^2-sense. Then*

$$\langle g, \sum_{k=1}^{\infty} h_k \rangle = \sum_{k=1}^{\infty} \langle g, h_k \rangle.$$

EXERCISE 5.18. Prove Corollary 5.17 by setting $f_n = \sum_{k=1}^{n} h_k$ in Proposition 5.16. ♢

EXERCISE 5.19. Verify that if $f = \sum_{n\in\mathbb{Z}} a_n e_n$ and $g = \sum_{n\in\mathbb{Z}} b_n e_n$, where the equalities are in L^2-sense, and $\{e_n\}$ are the trigonometric functions, then

$$\langle f, g\rangle = \sum_{n\in\mathbb{N}} a_n \overline{b_n}.$$

$\Diamond$

EXERCISE 5.20. Suppose $\{f_\lambda\}_{\lambda\in\Lambda}$, with Λ an arbitrary index set, is an orthogonal family in $L^2(\mathbb{T})$. Assume all the functions f_λ are non-zero in $L^2(\mathbb{T})$. Show that they are *linearly independent*; in other words, for any finite subset of indices $\{\lambda_1, \dots, \lambda_n\} \subset \Lambda$,

$$a_1 f_{\lambda_1} + \cdots a_n f_{\lambda_n} = 0, \quad \text{if and only if} \quad a_1 = \cdots = a_n = 0.$$

$\Diamond$

5.2.2. Orthonormal bases. Orthogonality implies *linear independence* (see Exercise 5.20), and if our geometric intuition is right, orthogonality is in some sense the "most" linearly independent a set could be. In a finite-dimensional vector space of dimension N, if we locate N linearly independent vectors we have located a basis of the space. Similarly if we have an N-dimensional inner-product vector space and we find N orthonormal vectors, we have found an orthonormal basis.

The trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are an orthonormal set of $L^2(\mathbb{T})$. Having an infinite orthonormal family (hence a linearly independent family) tells us that the space is infinite-dimensional. So far, there is no guarantee that there are no other functions in the space $L^2(\mathbb{T})$ orthogonal to the trigonometric functions, or in other words that *the system of trigonometric functions is complete.* We will see 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, and also to Parseval's Identity.

DEFINITION 5.21. Let $\{f_n\}_{n\in\mathbb{N}}$ be an orthonormal family in $L^2(\mathbb{T})$. We say that the family is *complete*, or that $\{f_n\}_{n\in\mathbb{N}}$ is a *complete orthonormal system in* $L^2(\mathbb{T})$, or that the vectors in the family 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 exists a sequence of complex numbers $\{a_n\}_{n\in\mathbb{N}}$ such that

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

Equivalently, $f = \displaystyle\sum_{n=1}^{\infty} a_n f_n$ where equality holds in the L^2-sense. $\Diamond$

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

LEMMA 5.22. (Uniqueness of the coefficients). *If $\{f_n\}_{n\in\mathbb{N}}$ is an orthonormal basis in $L^2(\mathbb{T})$, and $f \in L^2(\mathbb{T})$, then there exists complex numbers $\{a_n\}_{n\in\mathbb{N}}$ such that*

$$f = \sum_{n=1}^{\infty} a_n f_n,$$

and the coefficients have no other choice than to be $a_n = \langle f, f_n\rangle$ for all $n \in \mathbb{N}$.

Proof. Take the expansion of f in the basis elements, and pair it with f_k, interchange inner-product and summation, and use the orthonormality of the system to get,

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

The interchange of the inner-product and the summation is justified because the inner-product is a *continuous map*. See Corollary 5.17. □

In the case of the trigonometric system, which is an orthonormal system, if the system is complete, the coefficients must be the Fourier coefficients. It will take us some work to verify exactly that, which we now state as a theorem, still to be proven.

Theorem 5.23. *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.

5.3. Completeness of the trigonometric system

The trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are an orthonormal set in $L^2(\mathbb{T})$. Our goal is to show that the system is complete, that is Theorem 5.23. We will show 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,

$$\|f - S_N f\|_2 \to 0 \qquad \text{as } N \to \infty, \text{ for } f \in L^2(\mathbb{T}).$$

Aside 5.24. *To simplify notation, we use $\|f\|_2$ instead of $\|f\|_{L^2(\mathbb{T})}$.*

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 it is clear, by Definition 5.21, that to verify completeness of the trigonometric system suffices to check the mean-square convergence of the partial Fourier sums,

$$\lim_{N\to\infty} \Big\| f - \sum_{|n|\le N} \langle f, e_n\rangle\, e_n \Big\|_2 = \lim_{N\to\infty} \|f - S_N f\|_2 = 0.$$

To prove Theorem 5.23 and hence Theorem 5.1, we will 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 will show that the partial Fourier sums converge to f in $L^2(\mathbb{T})$ for continuous functions f. Finally, an approximation argument will allow us to conclude the same for square-integrable functions.

5.3.1. $S_N f$ is the closest $N^{\mathbf{th}}$- 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 5.25. *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 $\quad e_j \perp (f - S_N f) \quad$ for each j such that $|j| \leq N$, because

$$\begin{aligned}
\langle f - S_N f, e_j\rangle &= \langle f - \sum_{|n|\leq N} \widehat{f}(n) e_n, e_j\rangle \\
&= \langle f, e_j\rangle - \sum_{|n|\leq N} \widehat{f}(n)\langle e_n, e_j\rangle \qquad \text{(linearity of the inner product)} \\
&= \langle f, e_j\rangle - \widehat{f}(j) \qquad \text{(orthonormality of the trigonometric system)} \\
&= 0.
\end{aligned}$$

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, as claimed. $\square$

Since the partial sum $S_N f = \sum_{|n|\leq N} \widehat{f}(n) e_n$ is a trigonometric polynomial of degree N, then by Lemma 5.25, it is orthogonal to $(f - S_N f)$, and by the Pythagorean Theorem 5.11,

$$\|f\|_2^2 = \|f - S_N f\|_2^2 + \|S_N f\|_2^2.$$

LEMMA 5.26. *If $f \in L^2(\mathbb{T})$, then*

$$\|f\|_2^2 = \|f - S_N f\|_2^2 + \sum_{|n|\leq N} |\widehat{f}(n)|^2.$$

PROOF. All we need to verify is that $\|S_N f\|_2^2 = \sum_{|n|\leq N} |\widehat{f}(n)|^2$. In fact, 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,

$$\begin{aligned}\|S_N f\|_2^2 &= \Big\| \sum_{|n|\le N} \widehat{f}(n) e_n \Big\|_2^2 \\ &= \langle \sum_{|n|\le N} \widehat{f}(n) e_n, \sum_{|j|\le N} \widehat{f}(j) e_j \rangle \\ &= \sum_{|n|\le N} \sum_{|j|\le N} \widehat{f}(n) \overline{\widehat{f}(j)} \langle e_n, e_j \rangle \\ &= \sum_{|n|\le N} \widehat{f}(n) \overline{\widehat{f}(n)} \\ &= \sum_{|n|\le N} |\widehat{f}(n)|^2,\end{aligned}$$

as required. Notice that this is nothing more than the Pythagorean theorem for $2N+1$ orthogonal summands. □

As a corollary we obtain Bessel's well-known inequality[4] for trigonometric functions.

LEMMA 5.27 (Bessel's Inequality). *For all $f \in L^2(\mathbb{T})$,*

$$\sum_{n\in\mathbb{Z}} |\langle f, e_n \rangle|^2 \le \|f\|_2^2. \tag{5.4}$$

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

It turns out that 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.

> *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 the above statement.

LEMMA 5.28 (Best Approximation Lemma). *Take $f \in L^2(\mathbb{T})$. Then for each $N \ge 0$ and for all trigonometric polynomials $P \in \mathcal{P}_N(\mathbb{T})$, we have*

$$\|f - S_N f\|_2 \le \|f - P\|_2.$$

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

PROOF. Let $P \in \mathcal{P}_N(\mathbb{T})$. Then there exist complex numbers $c_{-N}, \dots, c_N$ such that $P = \sum_{|n|\le N} c_n e_n$. Then

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

---

[4]This inequality is named after the German mathematician Friedrich Wilhelm Bessel (1784–1846).

Note that the underbraced terms are orthogonal to each other by Lemma 5.25. Hence, by the Pythagorean theorem,

$$\|f - P\|_2^2 = \|f - S_N f\|_2^2 + \underbrace{\Big\| \sum_{|n|\leq N} \big(\widehat{f}(n) - c_n\big) e_n \Big\|_2^2}_{\geq 0}.$$

Therefore

$$\|f - P\|_2^2 \geq \|f - S_N f\|_2.$$

Clearly equality holds if and only if the underbraced term is equal to zero, but by the argument in the proof of Lemma 5.26,

$$\Big\| \sum_{|n|\leq N} (\widehat{f}(n) - c_n) e_n \Big\|_2^2 = \sum_{|n|\leq N} \big|\widehat{f}(n) - c_n\big|^2.$$

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 5.29. *Every lemma in this section 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 theorems are valid for any inner-product vector space and for any orthonormal system.

Exercise 5.30. State and prove the analogues of Lemma 5.25, Lemma 5.26, Bessel's Inequality 5.27 and Best Approximation Lemma 5.28, for X any orthonormal system in $L^2(\mathbb{T})$, not necessarily the trigonometric system. ◊

5.3.2. Mean-square convergence for continuous functions.

Proof of Theorem 5.1 for Continuous Functions. We are ready to prove that

$$\|f - S_N f\|_2 \to 0 \qquad \text{as } N \to \infty,$$

for *continuous* functions f.

Fix $\varepsilon > 0$.

By Fejér's Theorem, the Cesàro sums $\sigma_N f$ converge *uniformly* to f on $\mathbb{T}$. So there is a trigonometric polynomial of degree M say, namely $P(\theta) = \sigma_M f(\theta)$, such that

$$|f(\theta) - P(\theta)| < \varepsilon \qquad \text{for all } \theta \in \mathbb{T}.$$

This implies that

$$\|f - P\|_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta) - P(\theta)|^2 \, d\theta < \varepsilon^2,$$

and so $\|f - P\|_2 \leq \varepsilon$.

Notice that if $M \leq N$, then $P \in \mathcal{P}_N(\mathbb{T})$. We can now use the Best Approximation Lemma (Lemma 5.28) to conclude that

$$\|f - S_N f\|_2 \leq \|f - P\|_2 < \varepsilon \qquad \text{for all } N \geq M.$$

Therefore, for continuous functions f,

$$\lim_{N\to\infty} \|f - S_N f\|_2 = 0$$

as claimed. □

5.3.3. Mean-square convergence for L^2-functions. We are now ready to prove the mean-square convergence for any function $f \in L^2(\mathbb{T})$, and hence to verify the completeness of the trigonometric system in $L^2(\mathbb{T})$.

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

$$\lim_{N\to\infty} \|f - S_N f\|_2 = 0.$$

The continuous functions on $\mathbb{T}$ are *dense* in $L^2(\mathbb{T})$, see Theorem 2.67. 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

$$\|f - g\|_2 \leq \varepsilon.$$

Since g is continuous, we can find a trigonometric polynomial P such that

$$\|g - P\|_2 \leq \varepsilon.$$

Therefore, by the triangle inequality,

$$\|f - P\|_2 \leq \|f - g\|_2 + \|g - P\|_2 \leq 2\varepsilon.$$

Let M be the degree of P. By Lemma 5.28, for all $N \geq M$ we have

$$\|f - S_N f\|_2 \leq \|f - P\|_2 \leq 2\varepsilon.$$

Lo and behold, for all $f \in L^2(\mathbb{T})$,

$$\lim_{N\to\infty} \|f - S_N f\|_2 = 0,$$

as claimed. □

5.3.4. Parseval's Identity. Given the completeness of the trigonometric system, we can now prove Parseval's Identity, namely

$$\|f\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta)|^2\, d\theta = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2.$$

Proof of Parseval's Identity 5.2. Recall Lemma 5.26,

$$\|f\|_2^2 = \|f - S_N f\|_2^2 + \sum_{|n|\leq N} |\widehat{f}(n)|^2. \tag{5.5}$$

Therefore the series on the right-hand side converges (all terms are nonnegative and the sum is bounded above by $\|f\|_2^2$),

$$\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2 \leq \|f\|_2^2.$$

We have just shown that if $f \in L^2(\mathbb{T})$ then $\|f - S_N f\|_2 \to 0$ as $N \to \infty$. In fact,

$$\|f\|_2^2 = \lim_{N\to\infty} \sum_{|n|\leq N} |\widehat{f}(n)|^2 =: \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2. \qquad \square$$

5.4. Equivalent conditions for completeness in $L^2(\mathbb{T})$

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 orthogonal to all the functions in $L^2(\mathbb{T})$ is the zero function. Another is that Plancherel's Identity, which is an infinite-dimensional Pythagorean Theorem, must hold. These ideas are summarized in Theorem 5.31, which holds for all inner-product vector spaces V over $\mathbb{C}$, see Appendix B. 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 $X = \{f_n\}_{n\in\mathbb{N}}$ an orthonormal family in $L^2(\mathbb{T})$. Then the following are equivalent:*

(i) *X is a complete system, hence an orthonormal basis of $L^2(\mathbb{T})$.*
(ii) *If $f \perp X$ then $f = 0$ in L^2.*
(iii) (Plancherel's Identity) *For all $f \in L^2(\mathbb{T})$,*

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

PROOF FOR THE TRIGONOMETRIC SYSTEM. When X is the trigonometric system, we have shown that completeness implies Parseval's Identity, which in this context coincides with Plancherel's Identity. That is, we have shown that (i) $\Rightarrow$ (iii).

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

$$\|f\|_2 = \lim_{N\to\infty} \|f - S_N f\|_2 = 0.$$

Therefore $f = 0$. That is the completeness of the trigonometric system implies that the only L^2-function orthogonal to all trigonometric functions is the zero function, that is (i) $\Rightarrow$ (ii) for the particular case of X being the trigonometric system.

If we had established Parseval's Identity by independent means, and knowing that $\{e_n(\theta) = e^{in\theta}\}_{n\in\mathbb{Z}}$ is an orthonormal system, then we can deduce the completeness of the trigonometric system. In fact, equation (5.5) holds for all $N > 0$, and so we can take the limit as $N \to \infty$ on both sides to get

$$\|f\|_2^2 = \lim_{N\to\infty} \|f - S_N f\|_2^2 + \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2.$$

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

$$\lim_{N\to\infty} \|f - S_N f\|_2 = 0.$$

To recapitulate: for X the trigonometric system we have shown the equivalence of conditions (i) and (iii) in Theorem 5.31. We also know that (i) implies (ii) in this case. To close the loop it suffices to argue that (ii) implies (iii).

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

$$\sum_{n\in\mathbb{Z}} |\langle g, e_n\rangle|^2 < \|g\|_2^2 < \infty.$$

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 Corollary 5.17), and

$$\|h\|_2^2 = \sum_{n\in\mathbb{Z}} |\langle g, e_n\rangle|^2,$$

by Exercise 5.19.

Now, let $f = g - h \in L^2(\mathbb{T})$. The first observation is that $f \neq 0$ in the L^2-sense, otherwise Parseval will hold for g. The second observation is that f is orthogonal to e_n for all $n \in \mathbb{Z}$, this is just 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 does not hold there is a non-zero function f orthogonal to all trigonometric functions. We have shown then that in the case of the trigonometric system, (ii) $\Rightarrow$ (iii). $\square$

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

5.4.1. Orthogonal projection onto a closed subspace. Lemma 5.28 says 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 actually true. We first recall that a subspace $\mathcal{W}$ of $L^2(\mathbb{T})$ is said to be *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, for all $g \in \mathcal{W}$*

$$\|f - g\|_2 \geq \|f - P_{\mathcal{W}}f\|_2 \qquad \text{for all } g \in \mathcal{W}.$$

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 any 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 subspaces $\mathcal{P}_N(\mathbb{T})$ of trigonometric polynomials of degree less than or equal to N. They are finite dimensional (of dimension $2N+1$) and therefore automatically closed.

The theorem holds for every complete inner-product vector space V and every closed subspace. Theorem 5.33 says that we can draw pictures as we do in Euclidean space of a vector and its orthogonal projection to a closed subspace (represented

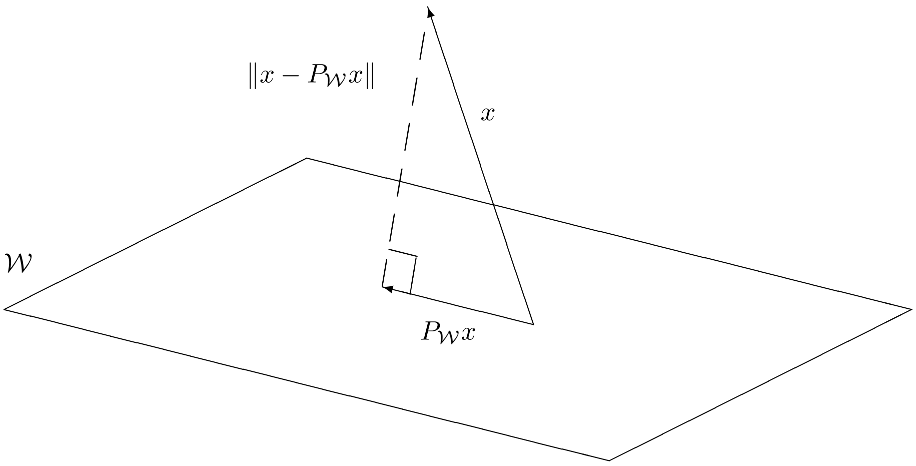


FIGURE 5.1. Orthogonal projection of a vector x onto a closed subspace $\mathcal{W}$ of a vector space V. The vector space V is represented here by three-dimensional space. The subspace $\mathcal{W}$ is represented by a plane. The vector x is pictured, and its orthogonal projection onto $\mathcal{W}$ 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 $x - P_{\mathcal{W}}x$.

by a plane), and the difference of these two vectors minimizes the distance to the subspace; see Figure 5.1.

We use this orthogonal projection theorem for the Hilbert space $L^2(\mathbb{R})$ instead of $L^2(\mathbb{T})$ in Section 9.3 when we discuss the Haar basis on $\mathbb{R}$, and in Chapter 10 when we talk about wavelets and multiresolution analyses.

EXERCISE 5.34. Prove Theorem 5.33 for a closed subspace $\mathcal{W}$ of $L^2(\mathbb{T})$. Knowing that the subspace has an orthonormal basis $\mathcal{B} = \{g_n\}_{n\in\mathbb{N}}$, hence

$$\mathcal{W} = \{f \in L^2(\mathbb{T}) : f = \sum_{n\in\mathbb{N}} a_n g_n,\ \sum_{n\in\mathbb{N}} |a_n|^2 < \infty\}.$$

$\diamondsuit$

CHAPTER 6

A whirlwind 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 collection of N orthonormal vectors, hence an orthonormal basis. The *discrete Fourier transform* is the linear transformation that gives us the coefficients in the trigonometric basis. 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.

For practical purposes, this finite theory is what is needed. Computers deal with finite vectors. The Discrete Fourier Transform (DFT) is calculated by applying an invertible $N \times N$ matrix to a given vector in $\mathbb{C}^N$, and the Discrete Inverse Fourier Transform by applying the inverse matrix to the transformed vector. The surprising fact is that one can perform these matrix multiplications in order $N \log_2 N$ operations, as opposed to the expected N^2 operations, using the celebrated *Fast Fourier Transform* (FFT) and algorithm which we describe in Section 6.5. This gain in the number of operations is of invaluable importance for numerical applications, especially when dealing with large problems.

In this chapter we introduce another orthonormal basis, the *discrete Haar basis*, in preparation for later chapters where we introduce the Haar basis for $L^2(\mathbb{R})$ and for $L^2[0,1]$). We highlight the similarities and the differences with the discrete trigonometric basis. There is a *Fast Haar Transform* (FHT), of order N operations, and we describe it in Section 6.8. The fast Haar transform is a precursor of the *Fast Wavelet Transform* (FWT) that we will study in more depth in later chapters.

6.1. Fourier Series: a summary

Let us begin by summarizing the ingredients of the theory of Fourier series, as we have developed it in Chapters 1–5.

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

(2) The Fourier coefficients are given by the formula

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

(3) The *building blocks* of the Fourier series are the *trigonometric functions* $e_n : \mathbb{T} \to \mathbb{C}$ given by

$$e_n(\theta) := e^{in\theta}, \qquad \text{where } n \in \mathbb{Z} \text{ and } \theta \in \mathbb{T}.$$

(They are called *trigonometric* rather than *exponential* to emphasize the connection with cosines and sines given by Euler's 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 L^2 inner product as

$$a_n = \widehat{f}\,(n) = \langle f, e_n \rangle.$$

(4) The Fourier series associated to f is the doubly infinite series formed by summing the products of the n^{th} Fourier coefficients a_n with the n^{th} trigonometric functions $e^{in\theta}$:

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

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 may not converge at all.

(5) 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 the partial Fourier sums converge in $L^2(\mathbb{T})$,

$$\|f - \sum_{|n|\leq N} \widehat{f}(n) e^{in\theta}\|_{L^2(\mathbb{T})} \to 0, \quad N \to \infty,$$

and the energy is preserved, that is Parseval's identity holds,

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

In the so-called *continuous setting* we have just described, the signal is a *function*, which takes as input the continuously varying quantity $\theta \in \mathbb{T}$. By contrast, in the *discrete setting* of finite Fourier analysis, the signal is a *vector* in $\mathbb{C}^N$, which can be thought of as a function taking as input the values in the discrete list $0, 1, \dots, N-1$. In the continuous setting the Fourier transform produces the doubly-infinite sequence of Fourier coefficients. In the discrete setting the discrete Fourier transform produces another vector in $\mathbb{C}^N$.

6.2. The discrete Fourier basis

We present the analogues in the discrete setting of the five ingredients of Fourier analysis listed above. Fix a positive integer N, and let $\omega := e^{2\pi i/N}$. Recall that $\mathbb{Z}/(N\mathbb{Z}) := \{0, 1, \dots, N-1\}$.

(1) To each *vector* $v = (z_0, \dots, z_{N-1}) \in \mathbb{C}^N$ we associate a second *vector* $\widehat{v} = (a_0, \dots, a_{N-1}) \in \mathbb{C}^N$. The vector $\widehat{v}$ is known as the *discrete Fourier transform* of v.

(2) The entries of the discrete Fourier transform are given by the formula

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

(3) The *building blocks* of the discrete Fourier transform are the N discrete trigonometric functions $e_l : \mathbb{Z}/(N\mathbb{Z}) \to \mathbb{Z}/(N\mathbb{Z})$ given by

$$e_l(k) := \frac{1}{\sqrt{N}} \omega^{kl}, \qquad k \in \{0, 1, \ldots, N-1\},$$

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

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

(4) The original vector v can be exactly determined from its discrete Fourier transform. The k^{th} entry of v is given by

$$v(k) := z_k = \sum_{k=0}^{N-1} a_n e_n(k).$$

(5) The set $\{e_0, \ldots, e_{N-1}\}$ forms an *orthonormal basis* for the vector space $\mathbb{C}^N$. Here *orthonormal* means with respect to the standard inner product or *dot product* in $\mathbb{C}^N$.

EXERCISE 6.1. Verify that the set $\{e_0, e_1, \ldots, e_{N-1}\}$ is an orthonormal set in $\mathbb{C}^N$. $\Diamond$

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 continuous 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 $\{a_n\} \in l^2(\mathbb{Z})$. In Chapter 7 we will develop the 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 $v = (z_0, \ldots, z_{N-1}) \in \mathbb{C}^N$ in the discrete setting can be thought of as functions $v : \mathbb{Z}/(N\mathbb{Z}) \to \mathbb{C}$, in the sense that for each number $n \in \mathbb{Z}/(N\mathbb{Z}) := \{0, 1, \ldots, N-1\}$, the function v picks out a complex number $v(n) = z_n$. So

$$v = (v(0), v(1), \ldots, v(N-1)).$$

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

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

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. We denote $\mathbb{C}^N$ with the standard inner-product as the space $\ell^2(\mathbb{Z}/N\mathbb{Z})$, to emphasize the similarities with the Hilbert space $L^2(\mathbb{T})$.

Fourth, in the discrete setting we can establish *Parseval's Identity*, also known as *Plancherel's Formula*:

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

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 continuous setting, Parseval's Identity reads

$$\|v\|_{\ell^2(\mathbb{Z}/N\mathbb{Z})} = \|\widehat{v}\|_{\ell^2(\mathbb{Z}/N\mathbb{Z})},$$

Here the norm, known as the ℓ^2-norm, is the Euclidean norm induced by the inner product; explicitly, it is given by the square root of the sum of the squares of the entries of the vector v, and thus by the square root of the inner product of v with itself.

As in the continuous setting, Parseval's Identity says that the energy of the transformed signal is the same as that of the original signal. In the continuous 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 6.2. *Historically people have preferred to work with the orthogonal basis* $\{f_0, f_1, \dots, f_{N-1}\}$ *in* $\mathbb{C}^N$ *defined by*

$$f_l = \sqrt{N}\, e_l,$$

instead of the orthonormal basis $\{e_l\}$, *the prize being the appearance of a factor* $1/N$ *in the definition of the coefficients and also in Parseval's formula (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* $v, w \in \mathbb{C}^N$*:*

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

We made an analogous convention in $L^2(\mathbb{T})$ *when we chose to include the factor* $1/(2\pi)$ *in the definition of the inner product. However, 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 6.3. Show that formula (6.2) defines an inner product in $\mathbb{C}^N$. $\Diamond$

EXERCISE 6.4. Consider the vectors $\{f_0, f_1, \dots, f_{N-1}\}$. Verify that they are orthonormal with respect to the normalized inner product in $\mathbb{C}^N$. $\Diamond$

Exercise 6.5. Define the normalized discrete Fourier coefficients of $v \in \mathbb{C}^N$ to be

$$\widehat{v}^{nor}(n) = \langle v, f_n \rangle_{nor}.$$

Use the theory of orthonormal bases to justify the reconstruction formula

$$v = \sum_{n=0}^{N-1} \widehat{v}^{nor}(n) f_n.$$

Verify Parseval's identity in this setting, namely

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

$\Diamond$

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

Example 6.6. (*Fourier Orthonormal Basis for* $\mathbb{C}^5$) Let $N = 5$, so that

$$\omega = e^{2\pi i/5}.$$

Let

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

Notice the interesting arrangement of the powers of ω in each vector f_l.

The vectors f_0, f_1, f_2, f_3, f_4 are pairwise orthogonal. Dividing by the length $\sqrt{5}$ of each f_l, we obtain an orthonormal basis for $\mathbb{C}^5$, namely $e_o = \{(1/\sqrt{5})f_0,\ e_1 = (1/\sqrt{5})f_1,\ e_2 = (1/\sqrt{5})f_2,\ e_3 = (1/\sqrt{5})f_3,\ e_4 = (1/\sqrt{5})f_4\}$. $\Diamond$

Example 6.7. (*Fourier Orthonormal Basis for* $\mathbb{C}^8$) Let $N = 8$, so that

$$\omega = e^{2\pi i/8} = e^{\pi i/4} = \frac{(1+i)}{\sqrt{2}}.$$

We have pairwise orthogonal vectors f_0, f_1, $\ldots$, f_7 given by, respectively,

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

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 $e_0 = (1/\sqrt{8})f_0$, $e_1 = (1/\sqrt{8})f_1$, $\ldots$, $e_7 = (1/\sqrt{8})f_7$. $\Diamond$

6.3. A parenthesis on bases and dual bases in $\mathbb{C}^N$

Before discussing in more detail the discrete Fourier transform, 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 linear combination of the elements of the basis: there are complex numbers $a_1, a_2, \dots, a_N$ such that

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

The complex numbers $a_1, a_2, \dots, a_N$ are uniquely determined by the vector v. $\Diamond$

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

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

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):

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

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

THEOREM 6.9. *The following statements are equivalent.*

(1) *A set of N vectors $\{v_1, v_2, \dots, v_N\}$ is a basis for $\mathbb{C}^N$.*
(2) *The vectors $\{v_1, v_2, \dots, v_N\}$ are linearly independent. That is, if $0 = a_1 v_1 + a_2 v_2 + \cdots + a_N v_N$, 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 non-singular. Therefore we can find the coefficients $a_1, a_2, \dots, a_N$ by applying the inverse matrix B^{-1} to v:*

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

EXERCISE 6.10. Prove the equivalences in Theorem 6.9. $\Diamond$

Given N linearly independent vectors $\{v_1, v_2, \dots, v_N\}$ 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_j the complex conjugate of the j^{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 performing the matrix multiplication that

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

EXERCISE 6.11. Verify that the j^{th} entry of the vector $B^{-1}v$ is the (complex!) inner product between the vector v and the vector w_j. $\Diamond$

The coefficients of v in the basis $\{v_1, v_2, \ldots, v_N\}$ are given exactly by $B^{-1}v$. 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}. In other words, there is a set of *dual vectors* forming a *dual basis* $\{w_1, w_2, \ldots, w_N\}$ to the basis $\{v_1, v_2, \ldots, v_N\}$, 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 6.12. Verify that given a basis $\{v_1, v_2, \ldots, v_N\}$ and its dual vectors $\{w_1, w_2, \ldots, w_N\}$ satisfy the following *orthonormality* condition:

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

Moreover, given a basis $\{v_1, v_2, \ldots, v_N\}$, if you can locate vectors $\{w_1, w_2, \ldots, w_N\}$ with the orthonormality property (6.4), then the matrix whose rows are the conjugates of the vectors $\{w_1, w_2, \ldots, w_N\}$ must be the inverse matrix to the matrix whose columns are the vectors in the basis. That is, the vectors $\{w_1, w_2, \ldots, w_N\}$ are the dual basis for $\{v_1, v_2, \ldots, v_N\}$. ◊

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

In particular if the basis $\{v_1, v_2, \ldots, v_N\}$ is orthonormal, the dual vectors are the vectors themselves. On the matrix side, this is equivalent to observing that

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

DEFINITION 6.13. 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 calculated by taking the conjugate of the transpose of U.

6.4. The discrete Fourier transform and its inverse

It is not hard to see that the transformation that maps the vector $v \in \mathbb{C}^N$ into the vector $\widehat{v} \in \mathbb{C}^N$ is a linear transformation. We call this transformation the *discrete Fourier transform*, or the *Fourier transform* in $\ell_2\big(\mathbb{Z}/(N\mathbb{Z})\big)$. The latter notation means that we are viewing $\mathbb{C}^N$ as an inner-product vector space whose elements are functions defined on the discrete set $\mathbb{Z}/(N\mathbb{Z}) = \{0, 1, \ldots, N-1\}$, with the standard Euclidean inner product.

The finite Fourier transform is completely determined by the $N \times N$ complex-valued symmetric matrix $(1/\sqrt{N})\overline{F_N}$, where F_N is the matrix whose columns are exactly given by the orthogonal Fourier vectors f_l. We call F_N the *Fourier matrix.* In other words, the following matrix formula holds for the discrete Fourier transform:

$$\widehat{v} = \frac{1}{\sqrt{N}}\overline{F}_N v. \tag{6.5}$$

EXAMPLE 6.14. (*Fourier Matrix for* $\mathbb{C}^8$) We display the Fourier matrix in the case when $N = 8$:

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

$\Diamond$

Notice that the matrix $\frac{1}{\sqrt{N}}F_N$ is a symmetric *unitary matrix*, because its columns are orthonormal vectors. In particular its inverse is its complex conjugate $\frac{1}{\sqrt{N}}\overline{F_N}$. We conclude that

$$\left(\frac{1}{\sqrt{N}}\overline{F_N}\right)^{-1} = \frac{1}{\sqrt{N}}F_N.$$

Therefore the *discrete inverse Fourier transform in* $\ell_2\big(\mathbb{Z}_N/(N\mathbb{Z})\big)$ becomes

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

In this finite-dimensional context, all we are doing is an orthogonal change of basis, from the standard basis to a representation in the Fourier basis.

EXERCISE 6.15. Verify that

$$\frac{1}{N}F_N\,\overline{F_N} = \frac{1}{N}\overline{F_N}\,F_N = I_N,$$

where I_N denotes the $N \times N$ identity matrix. $\Diamond$

Both F_N and $\overline{F_N}$ are full matrices, meaning that all entries are nonzero. Thus it seems that computing the discrete Fourier transform or inverse Fourier transform involves N^2 multiplications. Since additions take up an insignificant amount of computer time compared with multiplications, the Fourier transform has an operation count of the order of N^2 operations. However there is a lot of hidden structure in these matrices, and one can dramatically improve on this operation count. The celebrated *Fast Fourier Transform* (FFT) is the algorithm that allows one to perform the finite Fourier transform, and the inverse Fourier transform, in $N\log_2 N$ operations.

6.5. The Fast Fourier Transform (FFT)

A key to the practical success of the discrete Fourier Transform (6.5) is a fast algorithm to accomplish it. The $N \times N$ matrix F_N is full, with entries $F_N(n,m) = e^{-2\pi inm/N}$. Applying it therefore involves N^2 complex multiplications. In the 1960's the American applied mathematician James Cooley and the American statistician John Wilder Tukey [1915-2000] rediscovered a faster algorithm, which the German mathematician Johan Carl Friedrich Gauss [1777-1855] had found in the early 1800s. This famous *Fast Fourier Transform* (FFT) algorithm reduces the number of multiplications to order $N \log_2 N$. This improvement revolutionized digital signal processing (and of course at Gauss's time it was completely unforeseen!).

Mathematically, the FFT is based on a factorization of the Fourier matrix into a collection of sparse matrices. The process is easiest to explain, and implement, when $N = 2^n$. It begins with a factorization that reduces F_N to two copies of $F_{N/2}$. This reduction continues for $n = \log_2 N$ steps, until the original matrix is written as a product of $2 \log_2 N$ sparse matrices, half of which can be collapsed to a single permutation matrix. The total number of multiplications required to apply F_N becomes $N \log_2 N$ instead of the brute force N^2. To understand why this works, we rearrange (6.1) to express $\widehat{v}(m)$ as a linear combination of two terms A_m and B_m, by

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

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

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

By clever ordering of operations we can reuse A_m and B_m and gain on the total number of operations. Furthermore, the sums A_m and B_m have the same structure as the sum in $\widehat{v}(m)$, so we can do this process recursively.

In the language of matrices, what we are doing in the above calculations, is factorizing the conjugate Fourier matrix $\overline{F_N}$, for N even, as a product of three $N \times N$ matrices,

$$\overline{F_N} = \begin{bmatrix} I_{N/2} & D_{N/2} \\ I_{N/2} & -D_{N/2} \end{bmatrix} \begin{bmatrix} \overline{F_{N/2}} & 0_{N/2} \\ 0_{N/2} & \overline{F_{N/2}} \end{bmatrix} \begin{bmatrix} \mathrm{Even}_{N/2} \\ \mathrm{Odd}_{N/2} \end{bmatrix}, \tag{6.7}$$

here I_M means the $M \times M$ identity matrix, F_M is the $M \times M$ Fourier matrix, 0_M denotes the $M \times M$ zero matrix. We will now describe precisely the diagonal $M \times M$ matrix D_M, and the $M \times 2M$ matrices Even_M and Odd_M.

The diagonal matrix D_M has diagonal entries $e^{-\pi i m/M}$ for $m = 0, 1, \ldots, M-1$. Here is D_4,

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

The $M \times 2M$ matrices Even_M and Odd_M grab in order the even entries and the odd entries of a given $2M$-vector respectively. Here are Even_4 and Odd_4 acting on a generic vector in $\mathbb{C}^8$, $v = [v(0), v(1), \ldots, v(7)]$,

$$\begin{bmatrix} \mathrm{Even}_4 \\ \mathrm{Odd}_4 \end{bmatrix} v = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v(0) \\ v(1) \\ v(2) \\ v(3) \\ v(4) \\ v(5) \\ v(6) \\ v(7) \end{bmatrix} = \begin{bmatrix} v(0) \\ v(2) \\ v(4) \\ v(6) \\ v(1) \\ v(3) \\ v(5) \\ v(7) \end{bmatrix}.$$

Notice that the first matrix in the decomposition (6.7) requires only N multiplications. The second requires $2(N/2)^2 = N^2/2$ multiplications, but it has the Fourier matrix $F_{N/2}$ in the diagonal, and we can use the recursive equation (6.7), to replace the $(N/2) \times (N/2)$ matrix $\overline{F_{N/2}}$ by

$$\overline{F_{N/2}} = \begin{bmatrix} I_{N/4} & D_{N/4} \\ I_{N/4} & -D_{N/4} \end{bmatrix} \begin{bmatrix} \overline{F_{N/4}} & 0_{N/4} \\ 0_{N/4} & \overline{F_{N/4}} \end{bmatrix} \begin{bmatrix} \mathrm{Even}_{N/4} \\ \mathrm{Odd}_{N/4} \end{bmatrix},$$

and keep replacing until one cannot subdivide further.

Exercise 6.16. Verify that after two steps in the recursion the middle matrix in (6.7) is decomposed as the product of three $N \times N$ matrices as follows,

$$\begin{bmatrix} I_{N/4} & D_{N/4} & & \\ I_{N/4} & -D_{N/4} & & \\ & & I_{N/4} & D_{N/4} \\ & & I_{N/4} & -D_{N/4} \end{bmatrix} \begin{bmatrix} F_{N/4} & & & \\ & F_{N/4} & & \\ & & F_{N/4} & \\ & & & F_{N/4} \end{bmatrix} \begin{bmatrix} \mathrm{Even}_{N/4} \\ \mathrm{Odd}_{N/4} \\ \mathrm{Even}_{N/4} \\ \mathrm{Odd}_{N/4} \end{bmatrix}.$$

◊

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 $j - 1 = \log_2 N$ times (until the block size is 2×2). After $j - 1$ steps, the matrix $\overline{F_N}$ will be decomposed into the product of $log_2 N - 1$ matrices each one requiring only N multiplications for a total of $N(\log_2 N - 1)$ multiplications, followed by the product of $j - 1$ "scrambling" matrices (these are all the Odd/Even matrices). A miracle then occurs, the $j-1$-scrambling matrices, when applied in the right order, become a very simple operation which basically costs nothing in terms of operation counts. The miracle is observed when describing the numbers $0, 1, \ldots, N-1$ in binary notation: "the succesive 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". This is called in computer sciences *bit reversal*.

Let us see the miracle in action with an example when $N = 8 = 2^3$. Denote the succesive scrambling matrices by

$$S_4 = \begin{bmatrix} \text{Even}_4 \\ \text{Odd}_4 \end{bmatrix}, \quad S_2 = \begin{bmatrix} \text{Even}_2 \\ \text{Odd}_2 \\ \text{Even}_2 \\ \text{Odd}_2 \end{bmatrix}.$$

We start with the numbers $0, 1, \ldots, 7$ in binary notation and apply the matrix S_4, to the outcome we apply the matrix S_2,

$$\begin{bmatrix} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{bmatrix} \overset{S_4}{\longrightarrow} \begin{bmatrix} 000 \\ 010 \\ 100 \\ 110 \\ 001 \\ 011 \\ 101 \\ 111 \end{bmatrix} \overset{S_2}{\longrightarrow} \begin{bmatrix} 000 \\ 100 \\ 010 \\ 110 \\ 001 \\ 101 \\ 011 \\ 111 \end{bmatrix}.$$

Notice that the binary numbers in the right-hand-side column are obtained from the binary numbers in the left-hand-side column by reading backwards the binary digits.

Exercise 6.17. Can you justify the miracle for all $N = 2^j$? Try an argument by induction. ◊

Even without a miracle, the scrambling $N \times N$ matrices $S_{N/2^k}$ are very sparse, only N entries are non-zero, in fact their non-zero 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 based in this decomposition is no more than $2N(\log_2 N - 1)$ operations. This is still an algorithm of order $N \log_2 N$.

Exercise 6.18. Given vectors $v, w \in \mathbb{C}^N$, define their *discrete convolution* by

$$v * w(n) := \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} v(k)w(n-k), \quad n = 0, 1, \ldots, N-1.$$

Notice that if $n < k$ then $-N \leq n - k < 0$, in that case define $w(n-k) := w(N - (n-k))$ (or think as w as extended periodically over $\mathbb{Z}$ so that $w(n) = w(N+n)$ for all $n \in \mathbb{Z}$). Notice that in matrix notation, the linear transformation T_w that maps the vector v into the vector $v * w$ is given by the *circulant matrix*

$$\begin{bmatrix} w(0) & w(N-1) & w(N-2) & \ldots & w(1) \\ w(1) & w(0) & w(N-1) & \ldots & w(2) \\ w(2) & w(1) & w(0) & \ldots & w(3) \\ \vdots & \vdots & \vdots & & \vdots \\ w(N-1) & w(N-2) & w(N-3) & \ldots & w(0) \end{bmatrix},$$

This is a priori a full matrix and requires N^2 operations to apply it.

(a) Show that

$$\widehat{v * w}(m) = \widehat{v}(m)\,\widehat{w}(m).$$

(b) Describe a fast algorithm to compute convolution in order $N \log_2 N$ operations. $\Diamond$

Convolution with a fixed vector w is a linear transformation given by a circulant matrix as described in Exercise 6.18. These are also the only *translation invariant* or *shift invariant* linear transformations. This means 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 $v \in \mathbb{C}^N$ k units, that is $S_k v(n) = v(n-k)$ (we are viewing the vector as extended periodically so that $v(n-k)$ is defined for all integers), then $S_k T_w = T_w S_k$.

EXERCISE 6.19. Show that if $T_w v = v * w$ then $S_k T_w = T_w S_k$. Furthermore, show that if a linear transformation T is shift invariant, meaning $S_k T = T S_k$, then there exists a vector $w \in \mathbb{C}^N$ such that $Tv = v * w$. $\Diamond$

The book by Frazier [**Fra**] takes a linear algebra approach and discusses these and much more in the first three chapters, before moving on to infinite dimensional space. Chapter 1 is a thorough review of linear algebra with emphasis on changes of basis and translation invariant linear transformations. Chapter 2 discusses the discrete Fourier theory. Chapter 3 discusses the discrete wavelet theory.

Another book that has a strong linear algebra flavor is that by Strang and Nguyen [**SN**].

6.6. The discrete Haar basis

In this section we discuss the discrete Haar basis. Unlike the discrete Fourier basis, the Haar basis is more "localized", a concept that we now describe.

In this context by localized we mean, in a not very precise way, the vector is zero except for a few entries. The discrete trigonometric vectors have all entries non-zero, in fact each entry has absolute value $1/\sqrt{N}$. Certainly the discrete trigonometric vectors are not localized. In particular when computing Fourier coefficients, the inner product of a vector v against a trigonometric vector e_n, all entries of v are involved in the computation.

An example of a basis which is as localized as it can be, is the *standard basis* in $\mathbb{C}^N$. These vectors have all but one entry equal to zero; the non-zero entry is 1. Denote by s_j the vector in the standard basis whose j^{th}-entry is 1:

$$s_1 = [1, 0, \dots, 0], s_2 = [0, 1, 0, \dots, 0], \dots, s_N = [0, \dots, 0, 1].$$

EXERCISE 6.20. Verify that the discrete Fourier transform of e_j is s_j, that is $\widehat{e_j} = s_j$. $\Diamond$

Although the Fourier basis is not localized at all, its Fourier transform is as localized as it can be. We say the Fourier basis is *localized in frequency*, but not in space or time.

EXERCISE 6.21. Verify that the discrete Fourier transform of s_j is the complex conjugate of e_j, that is $\widehat{s_j} = \overline{e_j}$. $\Diamond$

Although the standard basis is as localized as it can be, its Fourier transform is not localized at all. We say the standard basis is *localized in space or time* but not in frequency.

EXERCISE 6.22. Let $\{v_1, v_2, \ldots, v_N\}$ be an orthonormal basis for $\mathbb{C}^N$. Show that their discrete Fourier transforms $\{\widehat{v_1}, \widehat{v_2}, \ldots, \widehat{v_N}\}$ is an orthonormal basis for $\mathbb{C}^N$. $\Diamond$

These 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 intermediate between the discrete Fourier basis and the standard basis, in terms of its space and frequency localization.

Let us first see an example.

EXAMPLE 6.23. Let $N = 2^3 = 8$. Define the following vectors $\tilde{h}_n$ for $n = 0, 1, \ldots, 7$:

$$\begin{aligned}
\tilde{h}_0 &:= [1,1,1,1,1,1,1,1],\\
\tilde{h}_1 &:= [-1,-1,-1,-1,1,1,1,1],\\
\tilde{h}_2 &:= [-1,-1,1,1,0,0,0,0],\\
\tilde{h}_3 &:= [0,0,0,0,-1,-1,1,1],\\
\tilde{h}_4 &:= [-1,1,0,0,0,0,0,0],\\
\tilde{h}_5 &:= [0,0,-1,1,0,0,0,0],\\
\tilde{h}_6 &:= [0,0,0,0,-1,1,0,0],\\
\tilde{h}_7 &:= [0,0,0,0,0,0,-1,1].
\end{aligned}$$

$\Diamond$

EXERCISE 6.24. Verify that the vectors $\{\tilde{h}_0, \tilde{h}_1, \tilde{h}_2, \tilde{h}_3, \tilde{h}_4, \tilde{h}_5, \tilde{h}_6, \tilde{h}_7\}$ are orthogonal vectors in $\mathbb{C}^8$. Verify that the vectors

$$h_0 = \frac{\tilde{h}_0}{\sqrt{8}}, \quad h_1 = \frac{\tilde{h}_1}{\sqrt{8}}, \quad h_2 = \frac{\tilde{h}_2}{\sqrt{4}}, \quad h_3 = \frac{\tilde{h}_3}{\sqrt{4}}, \tag{6.8}$$
$$h_4 = \frac{\tilde{h}_4}{\sqrt{2}}, \quad h_5 = \frac{\tilde{h}_5}{\sqrt{2}}, \quad h_6 = \frac{\tilde{h}_6}{\sqrt{2}}, \quad h_7 = \frac{\tilde{h}_7}{\sqrt{2}},$$

are orthonormal. They form the discrete Haar basis for $\mathbb{C}^8$. $\Diamond$

EXERCISE 6.25. Compute the discrete Fourier transform of each of the vectors in the Haar basis for $\mathbb{C}^8$, displayed in (6.8). Either analytically or using MATLAB, in the latter case, sketch the graphs of h_n and $\widehat{h}_n$ and notice that the more localized h_n is the less localized $\widehat{h}_n$ would be. $\Diamond$

When $N = 2^j$ then it should be clear how to define the N Haar vectors. The first not normalized vector ($\tilde{h}_0$) is a string of ones; the second one ($\tilde{h}_1$) has the first half of the entries equal to -1, the second half equal to 1; the third one ($\tilde{h}_2$) has the first quarter of the entries equal to -1, the second quarter 1, and the rest of the entries are zero; etc. The last not-normalized vector ($\tilde{h}_{N-1}$) consists of $N-2$ zeros, and the last entries are -1 and 1 respectively.

Notice that half of the vectors are very localized, having exactly two consecutive entries non-zero, 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 except four consecutive entries non-zero. 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 except

2^k consecutive entries non-zero. 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 non-zero. Furthermore the non-zero entries consists of a string of consecutive -1's followed by the same number of 1's, except for $\tilde{h}_0$.

We now give a precise definition of the N^{th} Haar basis. We should really tag each of the N vectors with an N to specify at what dimension we are. However, to make the notation less cumbersome, we will not tag them, and we expect the reader to be aware of the underlying dimension.

Given $0 \le n < N = 2^j$ there exist a unique $0 \le k < j$ such that

$$2^k \le n < 2^{k+1}, \quad \text{moreover} \quad n = 2^k + m, \;\; 0 \le m < 2^k.$$

We now define $\tilde{h}_n$ in terms of those unique numbers k and m determined by n:

$$\tilde{h}_n(l) = \begin{cases} -1 & \text{if } \; mN2^{-k} \le l < (2m+1)N2^{-k-1}, \\ 1 & \text{if } \; (2m+1)N2^{-k-1} \le l < (m+1)N2^{-k}, \\ 0 & \text{otherwise} \end{cases}$$

Exercise 6.26. Verify that using this formula when $N = 8$ we get the same vectors as in Example 6.23. ◊

Notice that $\tilde{h}_n(l)$ is non-zero if and only if $m2^{j-k} \le l < (m+1)2^{j-k}$. In fact those entries $2^{j-k} = N/2^k$ are ± 1. The Haar vectors are defined by

$$h_n = \frac{2^{k/2}}{\sqrt{N}} \tilde{h}_n.$$

Exercise 6.27. Show that the vectors $\{h_n\}_{n=0}^{N-1}$ are orthonormal. ◊

Notice that except for h_0, all the other vectors are rescaled contractions of h_1 and translates of those.

The parameters k, m, uniquely determined by n so that $n = 2^k + m$, are *scaling* and *translation* parameters respectively. The scaling parameter k, $0 \le k < j$, tells us that we will decompose the set $\{0, 1, 2, \dots, 2^{j-1}\}$ into 2^k disjoint subsets of $2^{j-k} = N/2^k$ consecutive numbers. The translation parameter m, $0 \le m < 2^{j-k}$, indicates which subinterval we are considering.

It would make sense to index the n^{th} Haar vector with the pair (k, m), and this is exactly what we will do when we study the Haar basis, and the 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) that are found by dilating and translating one function, ψ, the *wavelet*. More precisely,

$$\psi_{k,m}(x) = 2^{k/2}\psi(2^k x - k).$$

The functions $\psi_{k,m}$ retain the shape of the wavelet ψ. If we assume that ψ is concentrated around zero on the interval $[-1, 1]$, then the functions $\psi_{k,m}$ 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), as $k \to -\infty$ the localization gets worst and worst.

6.7. 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 its coefficients in the 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 following is the inverse discrete Haar transform,

$$v = H_N\, c, \quad \text{where } c = [c_0, c_1, \ldots, c_{N-1}]^t,$$

and the Haar matrix H_N is the $N \times N$ matrix whose j^{th}-column is the Haar vector h_{j-1}. That is,

$$H_N = \begin{bmatrix} | & & | \\ h_0 & \ldots & h_N \\ | & & | \end{bmatrix}.$$

The Haar matrix H_N is a unitary matrix (orthonormal columns), hence its inverse is its conjugate transpose. H is also a real matrix (the entries are the real numbers $0, \pm 1$), 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. That is the discrete Haar transform,

$$c = H_N^t\, v.$$

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 6.28. Here is the matrix H_8^t:

$$H_8^t = \begin{bmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

$\diamondsuit$

Note that the Haar matrix H_8^t is *sparse* i.e it has lots of zero entries. If we count 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 rowss as in H_8^t and the 9th to 16th 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 $N(log_2 N + 1)$ multiplications. This implies that

Discrete Haar Transform can be performed in order $N log_2 N$ operations. The brute force operation count is of the same order as the FFT, because of the sparcity of the Haar matrix. Here is an argument for seeking localized basis vectors. 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. Such algorithm we will call the *Fast Haar Transform*, and it is the first example of the *Fast Wavelet Transform*.

6.8. The Fast Haar Transform

We will try to illustrate with some examples how one could 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 explicitely 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 consist 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 2 multiplications.*

Consider now the case $N = 4$. Let us explicitely compute the action of the matrix H_4^t on a vector $v = [v(0), v(1), v(2), v(3)]^t \in \mathbb{C}^4$.

$$H_4^t v = \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} v(0) \\ v(1) \\ v(2) \\ v(3) \end{bmatrix} = \begin{bmatrix} \frac{v(3)+v(4)+v(1)+v(0)}{\sqrt{4}} \\ \frac{\left(v(3)+v(2)\right)-\left(v(1)+v(0)\right)}{\sqrt{4}} \\ \frac{v(1)-v(0)}{\sqrt{2}} \\ \frac{v(3)-v(2)}{\sqrt{2}} \end{bmatrix}.$$

Denote the scaled averages of pairs of consecutive entries a_m^1, and their differences by d_m^1, in our case $j = 2$.

$$a_0^1 := \frac{v(1) + v(0)}{\sqrt{2}}, \quad d_0^1 := \frac{v(1) - v(0)}{\sqrt{2}}.$$

$$a_1^1 := \frac{v(3) + v(2)}{\sqrt{2}}, \quad d_1^1 := \frac{v(3) - v(2)}{\sqrt{2}}.$$

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 6.29. *The superscript 1 is at this stage unnecessary. However we are computing averages and differences of the averages. We can introduce* $a^0 = a_0^0 := \frac{a_1^1 + a_0^1}{\sqrt{2}}$, $d^0 := d_0^0 := \frac{a_1^1 - a_0^1}{\sqrt{2}}$. *With this notation we have that*

$$H_4^t v = \begin{bmatrix} a^0 & d^0 & d^1 \end{bmatrix}^t .$$

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.

It should now be clear that we can decompose H_4^t as the product of two 4×4 matrices,

$$H_4^t = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} H_2^t & 0_2 \\ 0_2 & H_2^t \end{bmatrix}.$$

The second matrix requires 2×2 multiplications (involves applying H_2^t twice), the first matrix involves only 2 multiplications (involves applying H_2^t once and the 2×2 identity matrix which involves no multiplications). The total multiplication count for H_4^t is then $2 \times 2 + 2 = 6 = 2(4 - 1)$.

Notice that the frist matrix can be decomposed as the product of a block diagonal matrix with H_2^t and I_2 (2×2 identity matrix) in the diagonal times a permutation matrix, more precisely,

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} H_2^t & 0_2 \\ 0_2 & I_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, .$$

EXERCISE 6.30. Verify that H_8^t can be decomposed as the product of three 8×8 matrices, as follows,

$$H_8^t = \begin{bmatrix} H_2^t & 0_2 \\ 0_2 & I_6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} H_4^t & 0_4 \\ 0_4 & H_4^t \end{bmatrix}.$$

Notice that the last matrix requires $2 \times 6 = 12$ multiplications (involves applying H_4^t twice), the first matrix involves only 2 multiplications (involves applying H_2^t once and I_6 the 6×6 identity matrix which involves no multiplications). The middle matrix, is a permutation matrix that we denote P_8, and it involves no multiplications. The total multiplication count for H_8^t is then $2 \times 6 + 2 = 14 = 2(8 - 1)$. ♢

We can see a pattern emerging. The matrix H_N^t, $N = 2^j$ can be decomposed as the product of three $N \times N$ matrices, the first and the last are block diagonal matrices, the middle matrix is a permutation matrix.

$$H_N^t = \begin{bmatrix} H_2^t & \\ & I_{N-2} \end{bmatrix} P_N \begin{bmatrix} H_{N/2}^t & 0_{N/2} \\ 0_{N/2} & H_{N/2}^t \end{bmatrix}.$$

Where I_{N-2} is the $(N-2)\times(N-2)$ identity, and P_N is an $N \times N$ permutation matrix (that is a permutation of the columns of I_N)

Notice that the last matrix in the decomposition requires $2 \times k_{N/2}$ multiplications (involves applying $H_{N/2}^t$ twice, and $k_{N/2}$ is the number of multiplications required to compute $H_{N/2}^t$), the first matrix involves only 2 multiplications (involves applying H_2^t once and 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 is then $2 \times k_{N/2} + 2$.

Exercise 6.31. Verify by induction that the if the above decomposition holds, the number of multiplications required to apply H_N^t is $2(N-1)$. That is, $k_N = 2(N-1)$. ◊

Returning to the averages and the differences but now in dimension $N = 2^j$. We start with the vector $v \in \mathbb{C}^N$ and we create two vectors a^{j-1} and d^{j-1} of dimension $N/2 = 2^{j-1}$ each by taking averages and diferences 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, and keep repeating until we reach dimension $1 = 2^0$. This scheme can be repeated j times, and can be represented by the following tree or *pyramid scheme*,

$$\begin{array}{ccccccccc} v := a^j & \rightarrow & a^{j-1} & \rightarrow & a^{j-2} & \rightarrow & \cdots & a^1 & a^0 \\ & \searrow & & \searrow & & \searrow & & \searrow & \\ & & d^{j-1} & & d^{j-2} & \cdots & d^1 & d^0 & \end{array}.$$

With this notation it can be seen that,

$$H_N^t v = \begin{bmatrix} a^0 & d^0 & d^1 & \dots & d^{j-1} \end{bmatrix}^t.$$

Notice that the vectors a^{j-k}, d^{j-k} have dimension $2^{j-k} = N2^{-k}$, so the vector $\begin{bmatrix} a^0 & d^0 & d^1 & \dots & d^{j-1} \end{bmatrix}$ 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.

Exercise 6.32. Can you be more precise in the description of the permutation matrix P_N in the decomposition of H_N^t? ◊

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. As we have seen, traditional Fourier series represent a periodic function as a sum of pure harmonics (sines and cosines)[1]:

$$f(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n) e^{2\pi inx}. \tag{7.1}$$

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

$$\widehat{f}(n) := \int_0^1 f(x)e^{-2\pi inx}\,dx, \qquad \text{for } n \in \mathbb{Z}. \tag{7.2}$$

It took almost 150 years to resolve exactly what this 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 was unknown until then).

We have already discussed these facts, and more, in Chapters 1–5. Among other things, we established in Chapter 5 the following fact.

> *The exponential functions $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ form an orthonormal basis for $L^2([0,1])$.*

This immediately implies that equation (7.1), the inverse Fourier transform, holds in the L^2-sense.

We also discussed in Chapter 6 the discrete Fourier transform in the finite dimensional vector space $\mathbb{C}^N$. In this setting the trigonometric vectors $\{e_l\}_{l=0}^{N-1}$, with k^{th}-entries given by $e_l(k) = \frac{1}{\sqrt{N}}e^{2\pi kl/N}$, form an orthonormal basis for $\mathbb{C}^N$. Therefore the discrete inverse Fourier transform is given by

$$v(k) = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} \widehat{v}(l)\, e^{2\pi kl/N},$$

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,

$$\widehat{v}(l) = \langle v, e_l\rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} v(k)\, e^{-2\pi kl/N}.$$

[1] A technical point here is that we are now using functions of period 1 instead of 2π; see Section 1.3.2. 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 non-periodic setting, the *Fourier transform* $\widehat{f}$ of a 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 a function $g : \mathbb{R} \to \mathbb{R}$ is defined by

$$(g)^\vee(x) = \int_{\mathbb{R}} g(\xi) e^{2\pi i \xi x}\, d\xi. \tag{7.4}$$

There is much to be said about for which functions these formulae make sense, and for which of them, and in what sense, the following *Fourier inversion formula* holds:

$$f(x) = \int_{\mathbb{R}} \widehat{f}\,(\xi) e^{2\pi i \xi x}\, d\xi,$$

or more concisely

$$(\widehat{f}\,)^\vee = f.$$

We discuss these issues and more in this chapter and the following one. We will see that the above formulas work for functions in the *Schwartz class*: infinitely often differentiable functions such that they and all their derivatives decay faster than any polynomial. In the Schwartz class the Fourier theory is perfect: we are in paradise. We will see in Chapter 8 that there is life beyond paradise, in the sense that we also have a very nice Fourier theory in 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

Heuristically, we could arrive at the integral formulae (7.3) and (7.4) by calculating the Fourier series on larger and larger intervals, until we cover the whole line, so to speak. This was Fourier's approach: first calculate the Fourier series of a "nice" function f on the symmetric interval $[-L/2, L/2)$ (see Section 1.3.2). For nice enough functions the Fourier series converges uniformly to the function f:

$$f(x) = \sum_{n\in\mathbb{Z}} a_L(n) e^{2\pi i n x/L} \qquad \text{for all } x \in [-L/2, L/2), \tag{7.5}$$

where the coefficients are given by the formula

$$a_L(n) = \widehat{f}^{\,L}(n) := \frac{1}{L}\int_{-L/2}^{L/2} f(x) e^{-2\pi i n x/L}\, dx.$$

Let

$$\xi_n = n/L,$$

so that

$$\Delta\xi := \xi_{n+1} - \xi_n = 1/L.$$

Then the Fourier coefficient $a_L(n)$ can be rewritten as

$$F_L(\xi) := \int_{-L/2}^{L/2} f(x) e^{-2\pi i \xi x}\, dx.$$

Thus we can now rewrite the Fourier series in equation (7.5) as

$$f(x) = \sum_{n\in\mathbb{Z}} F_L(\xi_n) e^{2\pi i \xi_n x} \Delta\xi.$$

This expression resembles a Riemann sum for the integral $\int_{-L/2}^{L/2} F_L(\xi)e^{-2\pi i\xi x}\,d\xi$, except that the parameter L appears (twice) in the function to be integrated, and in the limits of integration, as well as in the partition of the interval. Let us be careless about that. Let $\Delta\xi \to 0$ (that is, $L \to \infty$). After observing that $\lim_{L\to\infty} F_L(\xi) = \widehat{f}(\xi)$, we could perhaps accept that we need *all* frequencies to recover a function in $\mathbb{R}$, and that the following integrals will play the same role that equations (7.1) and (7.2) play in the Fourier series theory:

$$f(x) = \int_{\mathbb{R}} \widehat{f}(\xi)e^{2\pi i\xi x}\,d\xi, \qquad \text{where} \qquad \widehat{f}(\xi) := \int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}\,dx. \tag{7.6}$$

To make this argument rigorous, we have to specify what *nice* means, so that we have uniform convergence or some other 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 Fourier transform on the Schwartz class $\mathcal{S}(\mathbb{R})$

It turns out that formulae (7.6) 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 often 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'', $\dots$, $f^{(\ell)}$, $\dots$ are *rapidly decreasing*, meaning that as $|x| \to \infty$ they decrease faster than any polynomial increases. Rephrasing, we obtain the following definition.

DEFINITION 7.1. The *Schwartz space* $\mathcal{S}(\mathbb{R})$ is defined by

$$\mathcal{S}(\mathbb{R}) := \Big\{ f \in C^\infty(\mathbb{R}) \,:\, \sup_{x\in\mathbb{R}} |x|^k |f^{(\ell)}(x)| < \infty \text{ for all integers } k,\, \ell \geq 0 \Big\}. \tag{7.7}$$

$\Diamond$

The Schwartz space is a vector space over the complex numbers. We invite you to verify that it is closed under the operations of multiplication by polynomials, multiplications by trigonometric functions, and differentiation.

EXERCISE 7.2. Show that if $f \in \mathcal{S}(\mathbb{R})$, then the products $x^k f(x) \in \mathcal{S}(\mathbb{R})$ for all $k \geq 0$. Show that the product $e^{-2\pi i x\xi} f(x) \in \mathcal{S}(\mathbb{R})$ for all $\xi \in \mathbb{R}$. Finally show that the derivatives $f^{(\ell)} \in \mathcal{S}(\mathbb{R})$ for all $\ell \geq 0$. $\Diamond$

Our first examples of Schwartz functions are the compactly supported functions in $C^\infty(\mathbb{R})$.

DEFINITION 7.3. 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* $[a,b]$. $\Diamond$

EXAMPLE 7.4. (*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 decay faster as $|x| \to \infty$ than the reciprocals of all polynomials, and so $f \in \mathcal{S}(\mathbb{R})$. $\Diamond$

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?

EXAMPLE 7.5. (*A Bump Function*) Here is an example of a *compactly supported Schwartz function* that is supported on the interval $[a, b]$ and non-zero on the interval (a, b):

$$B(x) = \begin{cases} \exp\left(\dfrac{-1}{x-a}\right)\exp\left(\dfrac{-1}{b-x}\right), & \text{if } a \leq x \leq b; \\ 0, & \text{otherwise.} \end{cases}$$

$\Diamond$

EXERCISE 7.6. Show that the bump function in Example 7.5 is in the Schwartz class. Notice that it suffices to verify 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 7.7. (*The Gaussian Function*) The canonical example of a function in $\mathcal{S}(\mathbb{R})$ is the *Gaussian function* $G(x)$, defined by

$$G(x) := e^{-\pi x^2}.$$

$\Diamond$

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 in the Riemann or in the Lebesgue sense, and they coincide):

$$\int_{\mathbb{R}} f(x)\,dx := \lim_{T\to\infty}\int_{-T}^{T} f(x)\,dx. \tag{7.8}$$

EXERCISE 7.8. Justify the existence of the limit of $\int_{-T}^{T} f(x)\,dx$ as $T \to \infty$, for functions $f \in \mathcal{S}(\mathbb{R})$. $\Diamond$

We then define the Fourier transform $\widehat{f} : \mathbb{R} \to \mathbb{C}$ of a Schwartz function f as follows:

$$\widehat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}\,dx := \lim_{T\to\infty}\int_{-T}^{T} f(x)e^{-2\pi i\xi x}\,dx. \tag{7.9}$$

Notice that the integrand is in $\mathcal{S}(\mathbb{R})$ (see Exercise 7.2).

The Gaussian, introduced in Example 7.7, plays a very important role in Fourier analysis, probability theory and physics. It has the unusual property of being equal to its own Fourier transform:

$$\widehat{G}(\xi) = e^{-\pi\xi^2}.$$

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 7.9. (*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 a direct calculation. $\Diamond$

A very important operation in harmonic analysis is the *convolution*. In the context of Fourier series, we have already seen how to convolve two periodic functions on $\mathbb{T}$. Periodic convolution was extensively discussed in Section 4.3, and used informally before that when convolving with the Dirichlet kernel. Section 4.3.2 described the smoothing properties of convolution. In Chapter 6 we touched on circular convolution for vectors. On the line, the convolution of two functions f, $g \in \mathcal{S}(\mathbb{R})$ is defined as follows:

$$f * g(x) := \int_{\mathbb{R}} f(x-y)g(y)\,dy. \tag{7.10}$$

This integral is well defined for each $x \in \mathbb{R}$ and for f, $g \in \mathcal{S}(\mathbb{R})$. As we will see, the Schwartz class is closed under convolutions. In Chapter 8 we discuss convolution of a Schwartz function with a distribution, and in Chapter 11 convolution of an L^p function with an L^q function.

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 (certainly in the linear case). Also, *convolutions are transformed into products* and vice versa, which explains the success of Fourier transform techniques in signal processing, since *filtering* (= convolution) is one of the most important signal-processing tools. In Section 7.3, we give a time–frequency dictionary that lists all these useful interactions.

We show in Section 7.4 that as a consequence of the time–frequency dictionary, the Schwartz class is closed under the Fourier transform. Furthermore we show in Section 7.7 that the Fourier inversion formula holds in $\mathcal{S}(\mathbb{R})$, and as a consequence, the Fourier transform is a bijection from $\mathcal{S}(\mathbb{R})$ into $\mathcal{S}(\mathbb{R})$. Moreover, the Fourier transform on $\mathcal{S}(\mathbb{R})$ preserves energy or L^2-norms (Plancherel).

7.3. The time–frequency dictionary for $\mathcal{S}(\mathbb{R})$

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:

$$[af + b\,g]^\wedge = a\widehat{f} + b\,\widehat{g}.$$

We have already mentioned that the Fourier transform converts differentiation into polynomial multiplication, and convolution into multiplication. It 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, less precisely, *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

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 $\widehat{\tau_h f}(\xi) = e^{-2\pi i h\xi}\widehat{f}(\xi) = M_{-h}\widehat{f}(\xi)$
(c)	modulation $M_h f(x) := e^{2\pi i h x} f(x)$	translation $\widehat{M_h f}(\xi) = \widehat{f}(\xi - h) = \tau_h \widehat{f}(\xi)$
(d)	dilation $D_s f(x) := s f(sx)$	inverse dilation $\widehat{D_s f}(\xi) = \widehat{f}(s^{-1}\xi) = s D_{s^{-1}} \widehat{f}(\xi)$
(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 $\widehat{\overline{f}}(\xi) = \overline{\widehat{f}(-\xi)} = \overline{\widetilde{\widehat{f}}}(\xi)$
(g)	differentiation $f'(x)$	multiplication by polynomial $\widehat{f'}(\xi) = 2\pi i \xi \widehat{f}(\xi)$
(h)	multiplication by polynomial $-2\pi i x f(x)$	differentiation $[-2\pi i x f(x)]^\wedge(\xi) = \frac{d}{d\xi}\widehat{f}(\xi)$
(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 $\widehat{fg}(\xi) = \widehat{f} * \widehat{g}(\xi)$

the signal $f(x)$. To give some terminology, *modulation* of a function $f(x)$ means multiplying the function by a term of the form $e^{2\pi i h x}$ for some real number h, as shown in the term $e^{2\pi i h x} f(x)$ on the left-hand side of property (c) in Table 7.1. As usual, horizontal *translation* of a function means adding a constant, say $-h$, to its argument, as shown in the term $\widehat{f}(\xi - h)$ on the right-hand side of property (c).

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 equivalent to 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 like linear operations, translations, dilations, and conjugations. Hence we are entitled to compute the Fourier transforms of the functions in the time column of the table.

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. For us this domain is either $\mathbb{R}$ or $\mathbb{T}$; both are groups. Properties (f)–(j) explain how the Fourier transform interacts with conjugation, differentiation, multiplication by polynomials, products and convolution of functions f.

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 in the Fourier-series setting, and the proofs in the present setting are almost identical.

EXERCISE 7.10. Verify properties (a)–(g) of the time–frequency dictionary. $\Diamond$

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 $-2\pi ix$ 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, although we did prove in Chapter 3 the analogue of its companion property (g), about the Fourier coefficients of the derivative of a given periodic and continuously differentiable function.

PROOF OF PROPERTY (h). Differentiating (formally) under the integral sign leads us immediately to the right formula. It only remains to justify the interchange of the limit and the integral.

We will check that $\widehat{f}$ is differentiable (as a bonus we get continuity of f), and that its derivative coincides with the Fourier transform of $-2\pi ixf(x)$, both in one stroke. Consider the difference

(7.11)
$$\frac{\widehat{f}(\xi+h)-\widehat{f}(\xi)}{h}-[-2\pi ixf(x)]^{\wedge}(\xi)=\int_{\mathbb{R}}f(x)e^{-2\pi ix\xi}\left(\frac{e^{-2\pi ixh}-1}{h}+2\pi ix\right)dx.$$

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 (7.11) smaller than a constant multiple of ε. We bound separately the contributions for large x and for small x. For all $N>0$,

$$|(7.11)|\le\underbrace{\int_{|x|\le N}|f(x)|\left|\frac{e^{-2\pi ixh}-1}{h}+2\pi ix\right|dx}_{(A)}+\underbrace{2\pi\int_{|x|>N}|x||f(x)|\left|\frac{e^{-2\pi ixh}-1}{2\pi xh}+i\right|dx}_{(B)}.$$

Since $f\in\mathcal{S}(\mathbb{R})$, f is bounded, say by $M>0$, and 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 $\lim_{N\to 0}\int_{|x|>N}x^{-2}\,dx=0$. Therefore, given $\varepsilon>0$, there is an $N>0$ such that

$$\int_{|x|\ge N}|xf(x)|\,dx\le\int_{|x|>N}\frac{C}{x^2}\,dx\le\varepsilon.$$

Moreover, for $|x| \leq N$, there is some h_0 such that for all $h < h_0$,

$$\left| \frac{e^{-2\pi i x h} - 1}{h} + 2\pi i x \right| \leq \frac{\varepsilon}{N},$$

because $\frac{d}{dx} e^{-2\pi i x y}\big|_{y=0} = -2\pi i x$, and $e^{-2\pi i x y}\big|_{y=0} = 1$, and on the compact interval $[-N, N]$, the continuous function $\frac{d}{dx} e^{-2\pi i x y}$ is necessarily uniformly continuous.

Also remember that for all real θ, $|e^{i\theta} - 1| \leq |\theta|$, and so $\left|\frac{e^{i\theta}-1}{\theta} + i\right| \leq 2$. All these facts together imply that $(A) \leq 2M\varepsilon$, and $(B) \leq 4\pi\varepsilon$. Hence for each $\varepsilon > 0$ there exists $h_0 > 0$ such that for all $h < h_0$

$$\left| \frac{\widehat{f}(\xi + h) - \widehat{f}(\xi)}{h} - [-2\pi i x f(x)]^\wedge(\xi) \right| \leq (2M + 4\pi)\varepsilon.$$

But this is exactly the definition of limit as $h \to 0$, we have shown

$$\lim_{h \to 0} \frac{\widehat{f}(\xi + h) - \widehat{f}(\xi)}{h} = [-2\pi i x f(x)]^\wedge(\xi).$$

We conclude that $\dfrac{d}{d\xi}\widehat{f}(\xi) = [-2\pi i x f(x)]^\wedge(\xi)$; this completes the proof of property (h). $\square$

Notice that in the proof of property (h) we only used decay properties of f. More precisely we used the fact that $\int |x||f(x)| < \infty$, which is the minimum necessary to ensure that we can calculate the Fourier transform of $-2\pi i x f(x)$.

Exercise 7.11. Use an induction argument to show that if

$$\int |x|^k |f(x)| < \infty,$$

then $\widehat{f}$ is k times differentiable. Moreover,

$$\frac{d^k}{d\xi^k}\widehat{f}(\xi) = [(-2\pi i x)^k f(x)]^\wedge(\xi).$$

$\lozenge$

7.4. The Schwartz class is closed under the Fourier transform

In this section we verify that the Fourier transform of a Schwartz function f is a Schwartz function. We must check that $\widehat{f}$ is infinitely differentiable, and that $\widehat{f}$ and all its derivatives are rapidly decaying.

Exercise 7.11 shows that $f \in \mathcal{S}(\mathbb{R})$ implies that $\widehat{f} \in C^\infty$, and gives a formula for the k^{th} derivative of $\widehat{f}$.

As a second step toward the proof that the Schwartz class is closed under the Fourier transform, we prove a version of the Riemann–Lebesgue Lemma in $\mathcal{S}(\mathbb{R})$, showing that the Fourier transform $\widehat{f}$ of a Schwartz function tends to zero as $|\xi| \to \infty$. Of course, this result must hold if $\widehat{f}$ is to be in the Schwartz class $\mathcal{S}(\mathbb{R})$.

Lemma 7.12 (Riemann–Lebesgue Lemma in $\mathcal{S}(\mathbb{R})$). *If $f \in \mathcal{S}(\mathbb{R})$, then*

$$\lim_{|\xi| \to \infty} \widehat{f}(\xi) = 0.$$

PROOF. Using a symmetrization trick that we have already encountered when verifying the Riemann-Lebesgue's Lemma for periodic functions, we see that

$$\begin{aligned}\widehat{f}(\xi) &= \int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}\,dx \\ &= -\int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}e^{\pi i}\,dx \\ &= -\int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi\left(x-\frac{1}{2\xi}\right)}\,dx \\ &= -\int_{-\infty}^{\infty} f\left(y+\frac{1}{2\xi}\right)e^{-2\pi i\xi y}\,dy.\end{aligned}$$

The last equality uses the change of variable $y = x - \frac{1}{2\xi}$. It follows that

$$\widehat{f}(\xi) = \frac{1}{2}\int_{-\infty}^{\infty}\left[f(x) - f\left(x+\frac{1}{2\xi}\right)\right]e^{-2\pi i\xi x}\,dx.$$

Hence

$$|\widehat{f}(\xi)| \leq \frac{1}{2}\int_{-\infty}^{\infty}\left|f(x) - f\left(x+\frac{1}{2\xi}\right)\right|\,dx.$$

Because $f \in \mathcal{S}(\mathbb{R})$, we know there exists $M > 0$ such that $|f(x)| \leq M/x^2$. Also, given $\epsilon > 0$, there exists $N > 0$ large enough such that for $|\xi| > 1$

$$\int_{|x|>N}\left|f(x) - f\left(x+\frac{1}{2\xi}\right)\right|\,dx \leq \int_{|x|>N}\left[\frac{M}{x^2} + \frac{M}{(x+\frac{1}{2\xi})^2}\right]dx \leq \epsilon.$$

As for the integral over $|x| \leq N$, since $g_\xi(x) = f(x) - f\left(x+\frac{1}{2\xi}\right)$ converges uniformly to zero as $|\xi| \to \infty$ on $|x| \leq N$ (because f is uniformly continuous on $[-N, N]$), we can make it as small as we wish. More precisely, given $\varepsilon > 0$ there exists $K > 0$ such that for all $|\xi| > K$, and for all $|x| \leq N$, $|g_\xi(x)| < \frac{\varepsilon}{4N}$. Therefore

$$|\widehat{f}(\xi)| \leq \frac{\varepsilon}{2} + \int_{|x|\leq N}|g_\xi(x)|\,dx \leq \frac{\varepsilon}{2} + 2N\frac{\varepsilon}{4N} = \epsilon.$$

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 $|x| > N$, continuity of the integral for $|x| \leq N$. A density argument shows that if $f \in L^1(\mathbb{R})$ then the same result holds. More is actually true. If $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is continuous (we showed under the condition that $f \in \mathcal{S}(\mathbb{R})$, or just that $\int |x||f(x)| < \infty$, that $\widehat{f}$ was actually differentiable, and hence continuous). The statement of what is called the *Riemann-Lebesgue Lemma* in the literature is,

LEMMA 7.13 (Riemann-Lebesgue Lemma). *If $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is continuous and $\lim_{|\xi|\to\infty}|\widehat{f}(\xi)| = 0$. That is, $\widehat{f} \in C_0(\mathbb{R})$.*

The final ingredient needed to prove that the Fourier transform takes $\mathcal{S}(\mathbb{R})$ to itself (Theorem 7.15) is that the Fourier transforms of Schwartz functions decrease rapidly.

LEMMA 7.14. *If $f \in \mathcal{S}(\mathbb{R})$ then $\widehat{f}$ is rapidly decreasing.*

PROOF. To prove the Lemma, we must check that for $f \in \mathcal{S}(\mathbb{R})$,

$$\sup_{\xi \in \mathbb{R}} |\xi|^n |\widehat{f}(\xi)| < \infty \qquad \text{for all integers } n \geq 0. \tag{7.12}$$

We already know that $\widehat{f}$ is continuous, and in fact that $\widehat{f}$ is C^∞. Moreover, the Riemann–Lebesgue Lemma (Lemma 7.12) 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. We proved the analogue of property (g) in the Fourier series case, and (Exercise 7.10) that property (g) itself holds: $\widehat{f'}(\xi) = 2\pi i \xi \widehat{f}(\xi)$. Iterating, we obtain the formula

$$\widehat{f^{(n)}}(\xi) = (2\pi i \xi)^n \widehat{f}(\xi) \qquad \text{for } f \in \mathcal{S}(\mathbb{R}),\ n \geq 0. \tag{7.13}$$

Therefore, to verify that $\widehat{f}$ is rapidly decreasing it suffices to check that for each $n \geq 0$,

$$\sup_{\xi \in \mathbb{R}} |\widehat{f^{(n)}}(\xi)| < \infty. \tag{7.14}$$

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

$$\lim_{|\xi| \to \infty} (2\pi i \xi)^n \widehat{f}(\xi) = 0. \tag{7.15}$$

Since we have already shown that $\widehat{f}$ is a C^∞ function, and in particular that $\widehat{f}$ is continuous and hence necessarily bounded on every closed interval, the decay of $\widehat{f}$ at infinity shown in equation (7.15) implies inequality (7.12) for all $n \geq 0$. Hence $\widehat{f}$ is rapidly decreasing, as required. □

After much preparation, we can now prove that the Fourier transform of any Schwartz function is also a Schwartz function (Theorem 7.15).

THEOREM 7.15. *(*The Schwartz Space Is Closed Under Fourier Transform*) If $f \in \mathcal{S}(\mathbb{R})$ then $\widehat{f} \in \mathcal{S}(\mathbb{R})$.*

PROOF OF THEOREM 7.15. Suppose $f \in \mathcal{S}(\mathbb{R})$. We want to show that $\widehat{f} \in \mathcal{S}(\mathbb{R})$. We already argüed that $\widehat{f} \in C^\infty(\mathbb{R})$, and that

$$\frac{d^\ell}{d\xi^\ell} \widehat{f}(\xi) = [(-2\pi i x)^\ell f(x)]^\wedge(\xi).$$

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 7.14 to each $g_\ell \in \mathcal{S}(\mathbb{R})$, we find that $\frac{d^\ell}{d\xi^\ell} \widehat{f} = \widehat{g_\ell}$ is rapidly decreasing for each $\ell \geq 0$. We conclude that $\widehat{f} \in \mathcal{S}(\mathbb{R})$, as required. □

Equations (7.13) and (7.15) illustrate an important principle in Fourier analysis that we have already discussed in the context of Fourier series in Chapter 3.

Smoothness of a function is correlated with 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 $\xi \to \pm\infty$. This principle does not seem all that interesting when we are in paradise. After all, functions in the Schwartz class and their Fourier transforms are infinitely differentiable and rapidly decaying.

A *band-limited function* is one whose Fourier transform is compactly supported. Thus the Fourier transform of a band-limited function shows the most dramatic form possible of decay at infinity. In accordance with the previous principle, it is not surprising that band-limited functions turn out to be analytic, which is an even stronger property than being infinitely often differentiable. Likewise, smoothness of the Fourier transform is correlated with fast decay of the function transformed at infinity. If the function has compact support, then its Fourier transform will be analytic. Not all analytic functions can be reached since they must satisfy some at most exponential growth conditions (the code word *analytic functions of exponential type*).

Much is known about decay properties of a function at infinity, and the corresponding analyticity properties of its Fourier transform. Such results go under the generic name of Paley–Wiener Theorems. To learn more about them, see for example [**Str2**, Section 7.2] and [**RS**, Chapter IX.3].

7.5. Convolution and approximations of the identity in $\mathbb{R}$

The convolution of two functions f, $g \in \mathcal{S}(\mathbb{R})$ is defined as follows:

$$f * g(x) := \int_{\mathbb{R}} f(x-y)g(y)\,dy. \tag{7.16}$$

This integral is well defined for each $x \in \mathbb{R}$ and for f, $g \in \mathcal{S}(\mathbb{R})$. Notice that by change of variables, $f * g = g * f$: convolution is commutative.

The new function $f * g$ defined by the integral in equation (7.16) 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

$$\frac{d}{dx}\big(f * g\big)(x) = f' * g(x) = f * g'(x), \qquad \text{for } f,\ g \in \mathcal{S}(\mathbb{R}).$$

The first identity above is true because in this setting, we are entitled to interchange differentiation and integration. The formal justification of this fact involves breaking the integral into two pieces, one where the variable of integration is bounded ($|y| \leq M$), the other where is not ($|y| > M$): on the bounded part Theorem 2.48 about interchange of limits and integration can be used, on the unbounded part the decay of the function can be used to guarantee that the integral goes to zero as $M \to \infty$. This is exactly the argument we used when proving Property (h) in the time–frequency dictionary.

The second identity holds because convolution is commutative; it can also be verified by integration by parts.

EXERCISE 7.16. (*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 clear when thinking about convolution of Schwartz functions, because we are already in paradise. One can convolve much less regular functions (e.g. just an integrable function, 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 will 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 Chapter 6.

EXERCISE 7.17. Verify properties (i) and (j) of the time–frequency dictionary: if f, $g \in \mathcal{S}(\mathbb{R})$ then

$$\widehat{f * g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi), \quad \widehat{fg}(\xi) = \widehat{f} * \widehat{g}(\xi).$$

$\Diamond$

Convolution can be viewed as a binary operation on the Schwartz class $\mathcal{S}(\mathbb{R})$. It is commutative. However, the Schwartz class with convolution does not form a group, since there is no *identity element* in $\mathcal{S}(\mathbb{R})$. This is very simple to verify once we have established property (i) in the time–frequency dictionary. Suppose there exists $e \in \mathcal{S}(\mathbb{R})$ such that for all $f \in \mathcal{S}(\mathbb{R})$ we have that $f * e = f$, Fourier transform on both sides and obtain that $\widehat{f}(\xi)\widehat{e}(\xi) = \widehat{f}(\xi)$. In particular you can set $f(x) = G(x) = e^{-\pi x^2}$, the Gaussian, which is its own Fourier transform and never vanishes. Hence we can cancel G from the identity $G(\xi)\widehat{e}(\xi) = G(\xi)$, and conclude that $\widehat{e} = 1$. But we have just shown that if $e \in \mathcal{S}(\mathbb{R})$, then $\widehat{e} \in \mathcal{S}(\mathbb{R})$, therefore it must decrease rapidly, but the function identically equal to one does not decrease. We have reached a contradiction. There cannot be an identity in $\mathcal{S}(\mathbb{R})$ for the convolution product.

There are two ways to address this lack of an identity element. The first is to introduce the *delta function*, to be defined in Section 8.4 below. 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 rather a generalized kind of function called a *distribution*. It acts as an identity under convolution (see Exercise 8.24; 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 discussed them in the context of Fourier series; see Section 4.4 and especially Definition 4.16. The functions in the family may actually be Schwartz functions, or they may be less smooth.

DEFINITION 7.18. An *approximation of the identity* in $\mathbb{R}$ is a family $\{K_t\}_{t\in\Lambda}$ of real-valued functions on $\mathbb{R}$, where Λ is an index set, together with a point t_0 that is an accumulation[2] point of Λ, with the following three properties.

[2]That is we can find points $t \neq t_0$ from the index set Λ that are arbitrarily close to t_0.

(i) The functions K_t have mean value one:
$$\int_{\mathbb{R}} K_t(x)\,dx = 1 \quad \text{for all } t \in \Lambda.$$
(ii) The functions K_t are uniformly integrable in t: there is a constant $C > 0$ such that
$$\int_{\mathbb{R}} |K_t(x)|\,dx \le C \quad \text{for all } t \in \Lambda.$$
(iii) Stated informally, the mass[3] of K_t becomes increasingly concentrated at $x = 0$ as $t \to t_0$:
$$\lim_{t\to t_0} \int_{|x|>\delta} |K_t(x)|\,dx = 0 \quad \text{for each } \delta > 0.$$

$\Diamond$

In the examples it will become clear 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 $t \to 1$. In the periodic case the kernels were indexed by $N \in \mathbb{N}$ (Fejér kernel), or by $0 < r < 1$ (Poisson kernel), and we considered $N \to \infty$ and $r \to 1^-$ respectively.

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 8.25. This is the origin of the term *approximation of the identity*, or in a language we already met, because of the property $f * K_t \to f$ as $t \to t_0$, which is another aspect of the same idea, see Theorem 7.21.

Exercise 7.19. (*Generating an Approximation of the Identity from a Kernel*) An easy way to produce an approximation of the identity is to start with a non-negative function (or 'kernel') $K(x) \ge 0$ with mean value one, $\int_{\mathbb{R}} K(x)\,dx = 1$, and define the family of dilations
$$K_t(x) := t^{-1}K(t^{-1}x), \qquad \text{for all } t > 0. \tag{7.17}$$
Verify that $\{K_t\}$ is an approximation of the identity, with $t_0 = 0$. $\Diamond$

Example 7.20. (*Gaussian Kernels Form an Approximation of the Identity*) Gaussians are good kernels, and by scaling them we obtain an approximation of the identity. More precisely, given $G(x) = e^{-\pi x^2}$, the *Gauss kernels* $G_t(x)$ are defined by
$$G_t(x) := t^{-1}e^{-\pi(t^{-1}x)^2}.$$
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 $t \to 0$, the graphs of $G_t(x)$ look more and more like the delta function; see Figure 7.1. $\Diamond$

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 in Theorem 4.20.

[3]The mass is the integral of the absolute value of K_t.

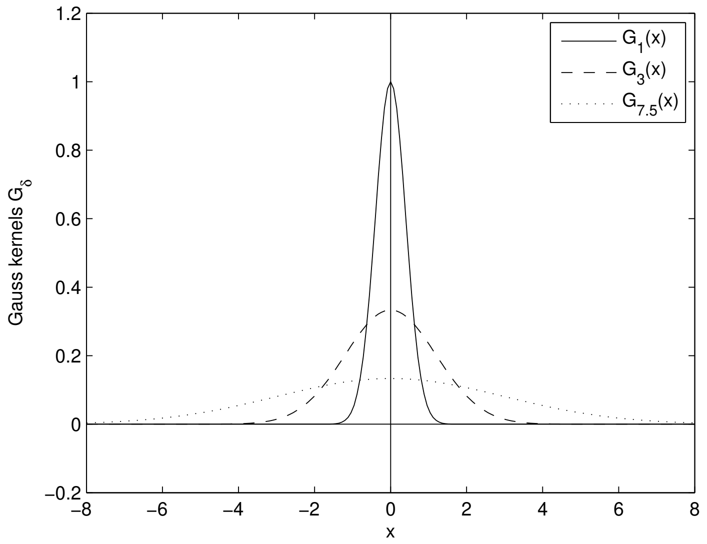


FIGURE 7.1. Gauss kernels.

THEOREM 7.21. *Let $\{K_t\}_{t\in\Lambda}$ be an approximation of the identity as $t \to t_0$. Let $K_t, f \in \mathcal{S}(\mathbb{R})$. Then the convolutions $f * K_t$ converge uniformly to f as $t \to t_0$.*

See also Theorem 7.38 and its proof, for convergence in L^p.

EXERCISE 7.22. Prove Theorem 7.21. $\Diamond$

Note that because of the smoothing effect of convolution, the approximating functions $f * K_t$ may be much smoother than the original function f. For instance, this idea gives examples illustrating that the uniform limit of continuous functions need not be continuous.

7.6. An interlude on kernels

When we studied summability methods in Sections 4.5 and 4.6, we used the Abel and Cesàro means to regularize 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 $\mathbb{T}$. We now define analogous Fejér and Poisson kernels on the real line $\mathbb{R}$; they can be used to regularize the convergence of the Fourier integral. In this subsection we also define the heat kernel on $\mathbb{R}$, the Dirichlet kernel on $\mathbb{R}$, and the conjugate Poisson kernel on $\mathbb{R}$.

The Fejér kernel on $\mathbb{R}$:

$$F_R(x) := R\left(\frac{\sin(\pi R x)}{R\pi x}\right)^2, \qquad \text{for } R > 0. \tag{7.18}$$

The Poisson kernel on $\mathbb{R}$:

$$P_y(x) := \frac{y}{\pi(x^2+y^2)}, \qquad \text{for } y > 0. \tag{7.19}$$

Exercise 7.23. Show that the Poisson kernel is a solution of *Laplace's equation*,

$$\Delta P := \frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} = 0,$$

for (x, y) in the upper half-plane. $\Diamond$

Neither the Fejér kernel nor the Poisson kernel are functions in the Schwartz class. They do belong to the class of *functions of moderate decrease*, discussed in Chapter 8 below. For functions of moderate decrease we can reproduce essentially all the proofs presented so far concerning approximations of the identity. In Chapter 8 we also discuss a much larger class of objects, the *tempered distributions*, which includes all the kernels in this section and much more.

The heat kernel on $\mathbb{R}$ (a modified Gaussian):

$$H_t(x) := \frac{1}{\sqrt{4\pi t}}\, e^{-\frac{|x|^2}{4t}}, \qquad \text{for } t > 0. \tag{7.20}$$

Exercise 7.24. Show that the heat kernel is a solution of the *heat equation* on the line:

$$\frac{\partial H}{\partial t} = \frac{\partial^2 H}{\partial x^2}.$$

$\Diamond$

Exercise 7.25. Show that the Gaussian kernels $\{G_\delta\}$, the Fejér kernels $\{F_R\}$, the Poisson kernels $\{P_y\}$, and the heat kernels $\{H_t\}$ generate approximations of the identity on $\mathbb{R}$, as $\delta \to 0$, $R \to \infty$, $y \to 0$, and $t \to 0$ respectively. The identities $\int_{\mathbb{R}} \frac{1 - \cos x}{x^2}\, dx = \pi$ and $\int_{\mathbb{R}} \frac{1}{1 + x^2}\, dx = \pi$ may be helpful. $\Diamond$

The Dirichlet kernel on $\mathbb{R}$:

$$D_R(x) := \int_{-R}^{R} e^{2\pi i x\xi}\, d\xi = \frac{\sin(2\pi R x)}{\pi x}, \qquad \text{for } R > 0. \tag{7.21}$$

Exercise 7.26. Let $S_R f$ denote the *partial Fourier integral* of f, defined by

$$S_R f := \int_{-R}^{R} \widehat{f}\,(\xi) e^{2\pi i x\xi}\, d\xi.$$

Show that $S_R f = D_R * f$. $\Diamond$

The Dirichlet kernels are not uniformly integrable, and so the family $\{D_R\}$ does not generate an approximation of the identity. (This obstacle is the origin of many of the convergence problems for Fourier series and integrals.) However, we can get around the problem by averaging, as we did in the periodic context. The averages of the Dirichlet kernels are the Fejèr kernels, which are an approximation of the identity.

Exercise 7.27. Show that

$$F_R(x) = \frac{1}{R}\int_0^R D_t(x)\, dt.$$

$\Diamond$

The conjugate Poisson kernel on $\mathbb{R}$:

$$Q_y(x) := \frac{x}{\pi(x^2+y^2)}\,, \qquad \text{for } y>0. \tag{7.22}$$

EXERCISE 7.28. Show that the conjugate Poisson kernels are not integrable, but that they are square integrable. In other words, $Q_y \notin L^1(\mathbb{R})$, 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 $1/(\pi i z)$. $\lozenge$

7.7. The Fourier Inversion Formula, and Plancherel's Identity

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 7.15.

We turn to the *multiplication formula*, relating the inner products of two Schwartz functions with each other's Fourier transforms.

EXERCISE 7.29. (*The Multiplication Formula*) Verify that for functions f, $g \in \mathcal{S}(\mathbb{R})$,

$$\int_{\mathbb{R}} f(s)\widehat{g}\,(s)\,ds = \int_{\mathbb{R}} \widehat{f}\,(s)g(s)\,ds. \tag{7.23}$$

Hint: Use the Fubini–Tonelli Theorem (Appendix A) to interchange the order of the integrals in an appropriate double integral. $\lozenge$

We have deliberately used the variable name s in the terms $f(s)$, $\widehat{g}\,(s)$, $g(s)$, and $\widehat{f}\,(s)$, rather than the more usual variable names x for f and g and ξ for $\widehat{g}$ and $\widehat{f}$, since we have to use the same variable of integration in each of the integrals.

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 inner product of f and $\widehat{g}$ in equation (7.23)) would not make sense. However it will make sense in the discrete case, where both the input and the output of the discrete Fourier transform are vectors in $\mathbb{C}^N$.

EXERCISE 7.30. State and prove a multiplication formula, analogous to formula (7.23), that is valid in $\mathbb{C}^N$. $\lozenge$

Several of the preceding facts can be woven together to prove the important Fourier inversion formula in $\mathcal{S}(\mathbb{R})$.

THEOREM 7.31 (Fourier Inversion Formula in $\mathcal{S}(\mathbb{R})$). *If $f \in \mathcal{S}(\mathbb{R})$, then for all $x \in \mathbb{R}$,*

$$f(x) = \int_{\mathbb{R}} \widehat{f}\,(\xi)e^{2\pi i\xi x}\,d\xi = (\widehat{f}\,)^{\vee}(x). \tag{7.24}$$

Notice the symmetry between the definition in formula (7.9) of the Fourier transform $\widehat{f}(\xi)$ of the function $f(t)$, and the formula (7.24) for $f(t)$ 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 that appear in the exponent. Whatever conditions are required on f so that formula (7.9) makes sense must also be required on $\widehat{f}$ so that formula (7.24) makes sense. In other words, the functions f and $\widehat{f}$ must enjoy the same status, which is the case in the Schwartz space. This is why the Fourier theory on $\mathcal{S}(\mathbb{R})$ is so beautiful and self-contained.

Proof of Theorem 7.31. Fix $x \in \mathbb{R}$. The idea is to apply the multiplication formula (7.23) to the functions f and g, where f is an arbitrary Schwartz function and g is the specific Schwartz function given by

$$g(s) := e^{2\pi i s x} e^{-\pi |ts|^2} \in \mathcal{S}(\mathbb{R}).$$

Compute directly, or use properties (c) and (d) from the time–frequency dictionary in Section 7.3, to verify that

$$\widehat{g}(s) = t^{-1} e^{-\pi |t^{-1}(x-s)|^2} = G_t(x-s), \tag{7.25}$$

which is a translate of the approximation of the identity generated by the Gaussian. Now equation (7.23) can be rewritten as

$$\int_{\mathbb{R}} f(s) G_t(x-s)\, ds = \int_{\mathbb{R}} \widehat{f}(s) e^{2\pi i s x} e^{-\pi |ts|^2}\, ds.$$

Let $t \to 0$. The left-hand side is $f * G_t(x)$, which converges to $f(x)$ (uniformly in x) as $t \to 0$ by Exercise 7.22. As for the right-hand side, we can interchange the limit as $t \to 0$ and the integral since $\widehat{f} \in \mathcal{S}(\mathbb{R})$ by Theorem 7.15, and the whole integrand is also in $\mathcal{S}(\mathbb{R})$. Therefore

$$\begin{aligned} f(x) &= \lim_{t\to 0} \int_{\mathbb{R}} f(s) G_t(x-s)\, ds \\ &= \lim_{t\to 0} \int_{\mathbb{R}} \widehat{f}(s) e^{2\pi i s x} e^{-\pi |ts|^2}\, ds \\ &= \int_{\mathbb{R}} \widehat{f}(s) e^{2\pi i s x} \lim_{t\to 0} e^{-\pi |ts|^2}\, ds = \int_{\mathbb{R}} \widehat{f}(s) e^{2\pi i s x}\, ds =: (\widehat{f}\,)^{\vee}(x). \end{aligned}$$

Lo and behold, $f(x) = (\widehat{f}\,)^{\vee}(x)$, which is the inversion formula we were seeking. □

Exercise 7.32. Use the time–frequency dictionary to check the identity (7.25) used in the proof of the Fourier Inversion Formula. ♢

Exercise 7.33. Justify the following interchange of limit and integral for $f \in \mathcal{S}(\mathbb{R})$:

$$\lim_{t\to 0} \int_{\mathbb{R}} \widehat{f}(s) e^{2\pi i s x} e^{-\pi |ts|^2}\, ds = \int_{\mathbb{R}} \widehat{f}(s) e^{2\pi i s x} \lim_{t\to 0} e^{-\pi |ts|^2}\, ds.$$

Hint: Break the integral in two pieces: where $|s| \le M$ 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 ensure that integral can be made small for M large enough. ♢

The inversion formula guarantees that the Fourier transform is in fact a *bijection* from the Schwartz space $\mathcal{S}(\mathbb{R})$ onto itself. More can be said: 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 that the Fourier transform preserves the L^2-norm.

THEOREM 7.34 (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 (7.23). 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 7.15.

EXERCISE 7.35. (*Polarization Identity*) Prove the polarization identity for real-valued functions f and g in $\mathcal{S}(\mathbb{R})$. Namely,

$$\langle f, g\rangle = \langle \widehat{f}, \widehat{g}\,\rangle. \tag{7.26}$$

Hint: Use Plancherel's Identity for $f+g$ and $f-g$, then add. $\diamondsuit$

EXAMPLE 7.36. (*An Application to Solving Differential Equations*) Let us find a function f on $\mathbb{R}$ that satisfies the 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

$$\widehat{f}\,(\xi)\big[(2\pi i\xi)^3 + 3(2\pi i\xi)^2 - 2(2\pi i\xi) - 6\big] = e^{-\pi\xi^2}. \tag{7.27}$$

Let $P(t) = t^3 + 3t^2 - 2t - 6 = (t+3)(t^2-2)$. Then the polynomial inside the brackets in equation (7.27) is $Q(\xi) := P(2\pi i\xi)$. Solving for $\widehat{f}\,(\xi)$ (fortunately $Q(\xi)$ has no real zeros), and using the Fourier inversion formula in $\mathcal{S}(\mathbb{R})$, we obtain

$$\widehat{f}\,(\xi) = \frac{e^{-\pi\xi^2}}{Q(\xi)} \qquad \text{and so} \qquad f(x) = \left(\frac{e^{-\pi\xi^2}}{Q(\xi)}\right)^{\vee}(x).$$

We have expressed the solution $f(x)$ as the inverse Fourier transform of a known function. $\lozenge$

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

$$P(D)f = \sum_{k=0}^{n} a_k D^k f = u$$

has a solution given by

$$f = \big(\widehat{u}(\xi)Q(\xi)^{-1}\big)^{\vee}.$$

Note that since $Q(\xi)$ has no real zeros, the function $\widehat{u}(\xi)Q(\xi)^{-1}$ is in the Schwartz class. Therefore we may compute its inverse Fourier transform and, by Theorem 7.15, the result is a function in $\mathcal{S}(\mathbb{R})$. In many cases $\big(\widehat{u}(\xi)Q(\xi)^{-1}\big)^{\vee}$ can be computed explicitly.

7.8. 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 $1 \leq p \leq \infty$, and state Minkowski's inequality. The heart of the subsection is the proof of Theorem 7.38, 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

$$\|f\|_p := \left(\int_{\mathbb{R}} |f(x)|^p \, dx\right)^{1/p}, \qquad \text{for} \quad 1 \leq p < \infty,$$

and

$$\|f\|_\infty := \sup_{|x| \in \mathbb{R}} |f(x)|, \qquad \text{for } p = \infty.$$

EXERCISE 7.37. (*Schwartz Functions Have Finite L^p-norm*) Verify that if $f \in \mathcal{S}(\mathbb{R})$, then $\|f\|_p < \infty$ for each p with $1 \leq p \leq \infty$. $\lozenge$

The Schwartz space $\mathcal{S}(\mathbb{R})$ with the L^p-norm is a normed space. In particular, the *triangle inequality* holds:

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p, \qquad \text{for all } f,\, g \in \mathcal{S}(\mathbb{R}) \text{ and all } p \text{ with } 1 \leq p \leq \infty.$$

This triangle inequality for the L^p-norm is also known as *Minkowski's inequality*. We prove it in Chapter 11 (Lemma 11.31).

When $\|f\|_p < \infty$, we say that $f \in L^p(\mathbb{R})$. 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 7.37 shows that $\mathcal{S}(\mathbb{R}) \subset L^p(\mathbb{R})$. A short-cut, 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.* Both definitions coincide. The second has the advantage that it immediately implies that the Schwartz class is dense in $L^p(\mathbb{R})$ with respect to the corresponding L^p-norm.

We show now that 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 each L^p-norm.

THEOREM 7.38. *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*

$$\lim_{t\to 0} \|K_t * f - f\|_p = 0 \qquad \textit{for each } p \textit{ with } 1 \leq p \leq \infty.$$

PROOF. We prove the result for $1 \leq p < \infty$. The proof below can be modified for the case $p = \infty$, which is the content of Theorem 7.21, whose proof was relegated to Exercise 7.22.

Define the translation operator τ_y by $\tau_y f(x) := f(x-y)$, for $y \in \mathbb{R}$. Notice that *translations preserve L^p-norms*, in other words $\|\tau_y f\|_p = \|f\|_p$ for all $f \in \mathcal{S}(\mathbb{R})$ and for all $h \in \mathbb{R}$. Thus for all $h \in \mathbb{R}$ we have the upper bound $\|\tau_y f - f\|_p \leq 2\|f\|_p$, by the triangle inequality.

The following estimate holds:

$$\|K_t * f - f\|_p = \left\| \int_{\mathbb{R}} [\tau_y f(x) - f(x)] K_t(y)\, dy \right\|_p \leq \int_{\mathbb{R}} \|\tau_y f - f\|_p |K_t(y)|\, dy. \tag{7.28}$$

We will justify (7.28) for $p = 1$ now, and we defer the sketch of the proof of inequality (7.28) for $p > 1$ until the end of our proof of Theorem 7.38, so as not to distract from the main line of the argument.

The following inequalities follow from property (i) of the approximation of the identity and the triangle inequality for integrals:

$$\begin{aligned} |K_t * f(x) - f(x)| &= \left| \int_{\mathbb{R}} f(x-y)\, K_t(y)\, dy - f(x) \right| \\ &= \left| \int_{\mathbb{R}} [\tau_y f(x) - f(x)]\, K_t(y)\, dy \right| \\ &\leq \int_{\mathbb{R}} |\tau_y f(x) - f(x)|\, |K_t(y)|\, dy. \qquad (7.29) \end{aligned}$$

In the second equality we used property (i) of the kernels K_t that says that they have integral equal to one, therefore $f(x) = \int f(x) K_t(y)\, dy$. The inequality holds because the absolute value of an integral is smaller than the integral of the absolute value. This inequality is what we are calling the triangle inequality for integrals, and 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. If instead of absolute values we were calculating with some other norm, in our case the L^p-norm, for which the triangle inequality holds, we would expect the analogous estimate (7.28) to hold. For $p = 1$, all that remains to be done is to integrate the pointwise inequality just derived and apply Fubini to interchange the order of integration:

$$\|K_t * f - f\|_1 \leq \int_{\mathbb{R}} \int_{\mathbb{R}} |\tau_y f(x) - f(x)|\, |K_t(y)|\, dy\, dx = \int_{\mathbb{R}} \|\tau_y f - f\|_1 |K_t(y)|\, dy.$$

We estimate separately the contributions to the right-hand side of inequality (7.28) 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 that for all $|t - t_0| < h$,

$$\int_{|y|>\delta} |K_t(y)|\, dy \leq \frac{\varepsilon}{4\|f\|_p}.$$

Second, we can make $\|\tau_y f - f\|_p$ small provided y is small. More precisely, since $f \in \mathcal{S}(\mathbb{R})$, we know that f is continuous and $f \in L^p(\mathbb{R})$. For continuous functions in $L^p(\mathbb{R})$,

$$\lim_{y \to 0} \|\tau_y f - f\|_p = 0. \tag{7.30}$$

In other words, translation is a continuous operation at zero in L^p, at least for continuous functions f. See Exercise 7.40. A density argument shows that the statement is actually true for all functions in $L^p(\mathbb{R})$, not just the continuous ones.

Hence, given $f \in \mathcal{S}(\mathbb{R})$ and $\varepsilon > 0$, there exists $\delta > 0$ (depending on f) such that

$$\|\tau_y f - f\|_p \leq \frac{\varepsilon}{2C}, \qquad \text{for all } |y| < \delta. \tag{7.31}$$

For $|y| < \delta$ we can apply estimate (7.31), with C the constant that controls uniformly the L^1-norms of K_t, that is $\int |K_t(y)| dy \leq C$ for all $t \in \Lambda$. Altogether, we have

$$\begin{aligned}
\|K_t * f - f\|_p &\leq \int_{|y|<\delta} |K_t(y)| \|\tau_y f - f\|_p \, dy + \int_{|y|\geq\delta} |K_t(y)| \|\tau_y f - f\|_p \, dy \\
&\leq \frac{\varepsilon}{2C} \int_{|y|<\delta} |K_t(y)|\, dy + 2\|f\|_p \int_{|y|\geq\delta} |K_t(y)|\, dy \\
&\leq \frac{\varepsilon}{2} + 2\|f\|_p \frac{\varepsilon}{4\|f\|_p} = \varepsilon.
\end{aligned}$$

In the last inequality we have used properties (ii) and (iii) of the approximation of the identity.

This completes the proof of Theorem 7.38, modulo inequality (7.28). □

It remains to establish inequality (7.28). The inequality (7.28) is a consequence of an integral version of the *triangle inequality* that goes by the name *Minkowski's Integral Inequality*, which essentially says that the *norm of the integral is the integral of the norms.* At the end of the book you can find a precise statement and some applications (page 222). Since f is a Schwartz function, both integrals in inequality (7.28) are limits of Riemann integrals on larger and larger intervals $[-M, M]$. The Riemann integral $\int_{-M}^{M} |\tau_y f(x) - f(x)| K(y)\, dy$ is a limit of finite sums of the type

$$\sum_{n=1}^{N} |\tau_{y_n} f(x) - f(x)|\, K(y_n)\, \Delta y.$$

We can now apply the triangle inequality[4] for the L^p-norm to this sum:

$$\left\| \sum_{n=1}^{N} |\, \tau_{y_n} f(\cdot) - f(\cdot)|\, K(y_n)\, \Delta y \right\|_p \leq \sum_{n=1}^{N} \|\tau_{y_n} f - f\|_p\, K(y_n)\, \Delta y.$$

The right-hand side is a Riemann sum for the integral $\int_{-M}^{M} \|\tau_y f - f\|_p\, K(y)\, dy$.

Taking appropriate limits on the partitions defining the Riemann sums, we conclude that for all $M > 0$,

$$\left\| \int_{-M}^{M} |\tau_y f(x) - f(x)|\, K(y)\, dy \right\|_p \leq \int_{-M}^{M} \|\tau_y f - f\|_p\, K(y)\, dy.$$

Finally, let $M \to \infty$ to obtain inequality (7.28).

Exercise 7.39. Show that if $f \in \mathcal{S}(\mathbb{R})$, then $\tau_y f \in \mathcal{S}(\mathbb{R})$, and for all $1 \leq p < \infty$,

$$\|f\|_p = \|\tau_y f\|_p.$$

$\diamondsuit$

Exercise 7.40. Show that for $f \in \mathcal{S}(\mathbb{R})$, we can interchange the following limit and integral:

$$\lim_{y\to 0} \int_{\mathbb{R}} |f(x-y) - f(x)|^p\, dx = \int_{\mathbb{R}} \lim_{y\to 0} |f(x-y) - f(x)|^p\, dx = 0. \tag{7.32}$$

Thus equation (7.30) is valid. $\diamondsuit$

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 for f to be L^p-integrable. There are stronger versions of the theorem; one can weaken the hypotheses and still get the same conclusion.

We encountered this theorem earlier in the context of Fourier series. It implies that the Cesàro and Abel sums converge to f in $L^p(\mathbb{T})$, at least for continuous functions f. See Section 4.7.

[4]More precisely, the triangle inequality for N summands $f_n \in L^p(\mathbb{R})$, $a_n \in \mathbb{C}$,

$$\|\sum_{n=1}^{N} a_n f_n\|_p \leq \sum_{n=1}^{N} |a_n|\|f_n\|_p.$$

In our case, choose $f_n = \tau_{y_n} f - f$, and $a_n = k(y_n)\Delta y$.

CHAPTER 8

Beyond paradise

The Fourier transform and most of its properties can be established in a more general setting than the class $\mathcal{S}(\mathbb{R})$ of Schwartz functions. In this section we introduce first the class of *functions of moderate decrease*, which contains $\mathcal{S}(\mathbb{R})$, and then the still larger class 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.

Strictly speaking, the tempered distributions are not functions at all, as we can see from the example of the delta distribution. However, the distributions still have a Fourier theory, meaning that we can extend to them the fundamental operations discussed above such as translation, dilation, the Fourier transform, and convolution with Schwartz functions. We clarify these statements below.

8.1. Functions of moderate decrease

A function f is said to be *of moderate decrease* if f is continuous and if there is a constant $A > 0$ such that

$$|f(x)| \leq \frac{A}{1+x^2} \qquad \text{for all } x \in \mathbb{R}. \tag{8.1}$$

The 1 in the denominator means that $|f|$ cannot be too big for x near 0, while the x^2 in the denominator means that $|f|$ cannot be too big as $x \to \pm\infty$. We note in passing that there is nothing special about the exponent 2 in the defining inequality (8.1). Every continuous function f such that $|f(x)| \leq A/(1+|x|^{1+\varepsilon})$ for all $x \in \mathbb{R}$, where $\varepsilon > 0$ and $A > 0$ are constants independent of x, has all the properties discussed in this subsection; in particular such an f is integrable.

Example 8.1. (*Functions of Moderate Decrease Need Not Be Schwartz*) All Schwartz functions are functions of moderate decrease. However there are functions such as

$$\frac{1}{1+|x|^n} \quad \text{for } n \geq 2, \qquad \text{and} \qquad e^{-a|x|} \quad \text{for } a > 0,$$

that are of moderate decrease, although they are not in $\mathcal{S}(\mathbb{R})$. (Check!) $\Diamond$

Example 8.2. Neither the Poisson kernel (7.19) nor the Fejér kernel (7.18) belong to $\mathcal{S}(\mathbb{R})$. However, they are functions of moderate decrease. $\Diamond$

Exercise 8.3. Verify that for each R and each $y > 0$, the Poisson kernel P_y and the Fejér kernel F_R are functions of moderate decrease. Show that the conjugate Poisson kernel Q_y is not a function of moderate decrease. $\Diamond$

We summarize the integrability and boundedness properties of functions of moderate decrease. We define the improper integral over $\mathbb{R}$ of a function f of moderate decrease as the limit as $N \to \infty$ of the Riemann integrals of f over the compact intervals $[-N, N]$:

$$\int_{-\infty}^{\infty} f(x)\,dx := \lim_{N\to\infty} \int_{-N}^{N} f(x)\,dx < \infty. \tag{8.2}$$

EXERCISE 8.4. Verify that functions of moderate decrease are closed under multiplication by bounded and continuous functions. ◊

In particular, if the function f is of moderate decrease, then the function $e^{-2\pi i x\xi} f(x)$ is also of moderate decrease, and its integral over $\mathbb{R}$ is well defined and finite for each real number ξ. That is, the Fourier transform for functions of moderate decrease is well defined by the formula

$$\widehat{f}(\xi) := \int_{-\infty}^{\infty} f(x) 2^{-2\pi i x\xi}\,dx.$$

EXERCISE 8.5. Show that the Fourier transform of the Poisson kernel is $e^{-2\pi|\xi|y}$. Show that the Fourier transform of the Fejér kernel is $\left(1 - \frac{|\xi|}{R}\right)$ if $|\xi| \le R$, and zero otherwise. ◊

EXERCISE 8.6. Show that the convolution of two functions f and g of moderate decrease is also a function of moderate decrease. Show that

$$\widehat{f * g} = \widehat{f}\widehat{g}.$$

◊

Notice that functions of moderate decrease are always bounded, since $|f(x)| \le A$ for all $x \in \mathbb{R}$. It follows that they are in the space $L^\infty(\mathbb{R})$ of *essentially bounded* functions. (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 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 as the smallest positive number M such that $|f(x)| \le M$ for all $x \in \mathbb{R}$ except for a set of measure zero. When the function f is actually *bounded*, then $\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$. Think of $L^\infty(\mathbb{R})$ as the space of bounded functions on the line.) It also follows that functions of moderate decrease are *locally integrable*, written $f \in L^1_{\text{loc}}$, meaning that for each compact interval $[a, b]$ the integral $\int_a^b |f(x)|dx$ is finite. Finally, all functions of moderate decrease are in $L^1(\mathbb{R})$ and in $L^2(\mathbb{R})$. First, the decay condition (8.1) guarantees that every function f of moderate decrease is integrable ($f \in L^1(\mathbb{R})$). Since f is also bounded, the following norm estimate holds:

$$\|f\|_2^2 = \int_{-\infty}^{\infty} |f(x)|^2\,dx \le \left(\sup_{x\in\mathbb{R}} |f(x)|\right) \int_{-\infty}^{\infty} |f(x)|\,dx = \|f\|_\infty \|f\|_1.$$

Thus f is also square integrable ($f \in L^2(\mathbb{R})$).

EXERCISE 8.7. (*Functions of Moderate Decrease are in* $L^p(\mathbb{R})$) Check that functions of moderate decrease are in $L^1(\mathbb{R})$ and indeed in every space $L^p(\mathbb{R})$. In other words, check that

$$\int_{-\infty}^{\infty} |f(x)|^p\,dx := \lim_{N\to\infty} \int_{-N}^{N} |f(x)|^p\,dx < \infty, \qquad \text{for } 1 \le p < \infty,$$

and that $\|f\|_\infty < \infty$. ◊

The Fourier transform is well defined for functions of moderate decrease. The conclusion of the Approximation of the Identity Theorem (Theorem 7.38) holds if we assume that the kernels and the function are functions of moderate decrease. Namely, if f and the good kernels K_δ are 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 when f and $\widehat{f}$ are both of moderate decrease. See [**SS1**, Sections 5.1.1 and 5.1.7] for more details.

In fact, if one follows closely the proofs presented for the Schwartz class, one realizes that the full power of the Schwartz class is not being used. In the arguments what is really needed is the fast decay of the function (integrability) to ensure that integrals outside a large interval ($|x| > M$) can be made arbitrarily small for M sufficiently large, and inside the compact interval $|x| \leq M$, one could now use uniform convergence arguments to interchange limit and integral, and take advantage of the integrand being small to guarantee that the integral over the compact interval is arbitrarily small. We can repeat the arguments if we assume the functions involved are functions of moderate decrease instead of Schwartz functions. By similar arguments, one can also verify the inversion formula when f and $\widehat{f}$ are both of in $L^1(\mathbb{R})$. That requires understanding of the Lebesgue integration theory, and the theorems for interchanging limits and integrals; see for example [**SS2**, Section 2.4].

Exercise 8.8. Verify that the Approximation of the Identity Theorem (Theorem 7.38) holds when the kernels and the functions convolved are functions of moderate decrease. ♢

Exercise 8.9. Verify that the inversion formula and Plancherel's Identity hold when f and $\widehat{f}$ are both of moderate decrease. ♢

We mentioned in the interlude on kernels (Section 7.6) 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 $\mathbb{R}$. It was left as an exercise to verify that the Fejér kernel on $\mathbb{R}$ can be written as an integral mean of Dirichlet kernels,

$$F_R(x) = \frac{1}{R} \int_{-R}^{R} D_t(x)\, dt,$$

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

$$\sigma_R f(x) := \frac{1}{R} \int_{-R}^{R} S_t f(x)\, dt = F_R * f(x) \to f(x).$$

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 11, we give a proof that the partial Fourier sums of f converge in $L^p(\mathbb{T})$, this time as a consequence of the boundedness of the periodic Hilbert transform in $L^p(\mathbb{T})$. The same is true in $\mathbb{R}$. Here the case $p = 2$ is taken care of by Plancherel's Identity.

8.2. Tempered distributions

We begin with some definitions. Let V be a vector field over some scalar field F; F is usually the real numbers or the complex numbers. A *linear functional* on F, or more briefly a *functional* on F, is a linear mapping from V to the scalar field F. If the vector space V has a topology (more below on the subject of topology), then a functional on V may or may not be *continuous* with respect to that topology.

DEFINITION 8.10. 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_a and T_b are said to *coincide in the sense of distributions* if $T_a(\phi) = T_b(\phi)$ 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 distribution sense to $T \in \mathcal{S}'(\mathbb{R})$ if for all $\phi \in \mathcal{S}(\mathbb{R})$

$$\lim_{n\to\infty} T_n(\phi) = T(\phi).$$

$\Diamond$

The notation $\mathcal{S}'(\mathbb{R})$ is used because, more generally, the collection of all continuous linear functionals $T : V \to F$ on a topological vector space V is usually denoted by V'. This collection V' is called the *dual* of V. Thus the space of tempered distributions is the dual of the space of Schwartz functions.

To say that T is linear means that

$$T(a\phi + b\psi) = aT(\phi) + bT(\psi) \qquad \text{for all } a,\, b \in \mathbb{C} \text{ and all } \phi,\, \psi \in \mathcal{S}(\mathbb{R}).$$

Continuity requires a bit more explanation. Very schematically, we note that a *topology* on a set X is a collection of subsets of X; they are known as the *open* subsets of X. The collection has to satisfy certain conditions that we won't go into here. Convergence of a sequence of elements x_n of X can be defined in terms of the topology on X. A function from a set X that has a topology to another set Y that also has a topology is said to be *continuous* if the inverse image of every open subset of Y is an open subset of X. In the familiar cases where X and Y are $\mathbb{R}$ or $\mathbb{R}^n$, there is a standard topology that arises from the notion of Euclidean distance, and the above definition of continuity for functions from X to Y coincides with the usual ε–δ definition of continuity. If X is a normed vector space, with the natural topology that arises from the norm, it turns out that a linear functional on X is continuous if and only if it is bounded. The topology of $\mathcal{S}(\mathbb{R})$ is not induced by a norm but by a *family of seminorms*. In the following aside we describe very briefly what that means. See for example [**Mun**] for more about topology, and [**Fol**, Chapter 5] for more about continuous functionals.

ASIDE 8.11. *We say a little more about the topology on $\mathcal{S}(\mathbb{R})$. If $\phi \in \mathcal{S}(\mathbb{R})$, then the quantities $\rho_{k,\ell}(\phi) = \sup_{x\in\mathbb{R}} |x|^k |\phi^{(\ell)}(x)|$ are finite for each k, $\ell \in \mathbb{N}$. These quantities are* seminorms*, meaning that they are positive (although each one of them is not necessarily positive definite, that is $\rho_{k,\ell}(\phi) = 0$ does not imply $\phi = 0$, but if for all k, $\ell \in \mathbb{N}$ the corresponding seminorm is zero, then the function must be zero), they are homogeneous ($\rho_{k,\ell}(\lambda\phi) = |\lambda|\rho_{k,\ell}(\phi)$), and they satisfy the triangle inequality. These seminorms generate a topology on $\mathcal{S}(\mathbb{R})$. With respect to that topology, a sequence $\{\phi_n\}$ 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*

$k, \ell \in \mathbb{N}$. It turns out that a linear functional T on $\mathcal{S}(\mathbb{R})$ is continuous if and only if whenever $\{\phi_n\}$ is a sequence in $\mathcal{S}(\mathbb{R})$ that converges in the topology of $\mathcal{S}(\mathbb{R})$ to some $\phi \in \mathcal{S}(\mathbb{R})$, the sequence $\{T(\phi_n)\}$ of complex numbers converges to $T(\phi)$ in $\mathbb{C}$. We systematically check the continuity of only a few of the tempered distributions introduced in this chapter. In particular, we verify that the delta distribution is continuous. Other cases 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:

$$T_f(\phi) := \int_{\mathbb{R}} f(x)\phi(x)\,dx \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}). \tag{8.3}$$

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 would overwhelm the decay of the Schwartz functions ϕ as $x \to \pm\infty$. However, polynomials or continuous functions that increase no faster than a polynomial are acceptable. Functions f in the Schwartz class are as reasonable as they could possibly be since they are rapidly decreasing, and functions of moderate decrease are also reasonable. Both Schwartz functions and functions of moderate decrease are bounded.

We say for short that a distribution T is a function f, when we mean that T is the tempered distribution induced by f, in other words $T = T_f$.

EXAMPLE 8.12. (*Bounded Functions Induce Tempered Distributions*) As an example, we show that if f is a bounded function then the formula (8.3) 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 function of moderate decrease.

Linearity is a consequence of the linearity of the integral. To check continuity, consider $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, this means in particular that $\lim_{n\to\infty} \rho_{2,0}(\phi_n) = 0$, that is for all $n > N_0$, $|x|^2|\phi_n(x)| \le 1$. By choosing $M > 0$ large enough we can guarantee that for all $n > N_0$

$$\int_{|x|>M} |\,\phi_n(x)|\,dx \le \int_{|x|>M} \frac{1}{x^2}\,dx < \frac{\varepsilon}{2}.$$

On the compact interval $[-M, M]$ the functions ϕ_n converge uniformly to zero, and we can interchange the limit and the integral:

$$\lim_{n\to\infty} \int_{|x|\le M} |\phi_n(x)|\,dx = \int_{|x|\le M} \lim_{n\to\infty} |\phi_n(x)|\,dx = 0.$$

So given $\varepsilon > 0$ there exists $N_1 > 0$ so that for all $n > N_1$,

$$\int_{|x|\le M} |\phi_n(x)|\,dx < \frac{\varepsilon}{2}.$$

Thus for $n > N := \max\{N_0, N_1\}$ we can ensure that

$$\int |\phi_n(x)|\,dx \le \varepsilon.$$

Finally, T_f is continuous, since for $n > N$,

$$|T_f(\phi_n)| \le \int |f(x)||\phi_n(x)|\,dx \le \sup_{x\in\mathbb{R}} |f(x)| \int |\phi_n(x)|\,dx \le \sup_{x\in\mathbb{R}} |f(x)|\varepsilon,$$

and therefore $\lim_{n\to\infty} T_f(\phi_n) = 0$. $\Diamond$

In particular the distribution T_1 induced by the bounded function $f = 1$ is a well defined tempered distribution.

EXERCISE 8.13. (*Polynomials Induce Tempered Distributions*) Let $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_kx^k$ be a polynomial of degree k. Show that the linear functional defined by integration against the polynomial f is a tempered distribution. $\Diamond$

Example 8.12 tells us that we can view $\mathcal{S}(\mathbb{R})$ as sitting inside $\mathcal{S}'(\mathbb{R})$, via the 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 itself continuous, in the sense that if $f_n,\ f \in \mathcal{S}(\mathbb{R})$ and $f_n \to f$ in the topology of $\mathcal{S}(\mathbb{R})$, then $T_{f_n}(\phi) \to T_f(\phi)$ in $\mathbb{C}$ for all $\phi \in \mathcal{S}(\mathbb{R})$, that is, T_{f_n} converges to T_f in the sense of distributions. Such continuous mapping are sometimes called *embeddings*. This nomenclature will be clearer in the next section, where we will define some operations for distributions that generalize ordinary operations on functions, in the sense that if $\mathcal{O}f$ is the action of an operation on a function (for example differentiation, translation, etc.), then the corresponding operation for the distribution is $\mathcal{O}\big(\varepsilon(f)\big) := \varepsilon(\mathcal{O}f)$.

EXERCISE 8.14. 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 $\phi \in \mathcal{S}(\mathbb{R})$. $\Diamond$

Like bounded functions, unbounded functions that have some decay at infinity and 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, as we show in Section 8.4 for the example of the delta distribution, and in Section 8.6 for the example of the *principal value distribution* of $1/x$.

In Example 8.12 we used the fact that if $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, then $\|\phi_n\|_{L^1} \to 0$. A similar argument allows one to show that the L^p-norms will also go to zero.

EXERCISE 8.15. Verify that if $\phi_n \in \mathcal{S}(\mathbb{R})$ and $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, then for all $1 \le p \le \infty$,

$$\|\phi_n\|_p^p := \int |\phi_n(x)|^p\, dx \to 0.$$

$\Diamond$

The collection of continuous linear functionals that acts on the space of compactly supported C^∞ functions (a subspace of the Schwartz class) are the so-called *distributions*. If we denote by $\mathcal{D}(\mathbb{R})$ the space of compactly supported C^∞ functions, the space of distributions $\mathcal{D}'(\mathbb{R})$ is its dual. Tempered distributions are distributions, the two spaces are nested, that is

$$\mathcal{S}'(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R}).$$

This is no accident but a fundamental fact in functional analysis. Given V, W topological vector spaces, denote their duals by V', W', then if $V \subset W$ (the topology of V is inherited from the topology in W) then $W' \subset V'$. Although we sometimes write just distributions for short, in this textbook we always mean tempered distributions. One could also consider the larger class of infinitely differentiable functions and its dual which will be a subset of $\mathcal{S}'(\mathbb{R})$, the so-called *compactly supported distributions*. If we think about what functions will give a continuous linear functional on C^∞, you realize the functions have to be compactly supported, hence the name (there is also a notion of support for distributions, and of compact support for distributions). The book by Robert Strichartz [**Str2**] is a delightful introduction to distribution theory at a level comparable to that of this book, but with a lot more information and applications.

8.3. The time–frequency dictionary for $\mathcal{S}'(\mathbb{R})$

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 will take it for granted that these extensions of the fundamental operations, and the Fourier transform, are nicely behaved with respect to the corresponding topology.

Given a tempered distribution $T : \mathcal{S}(\mathbb{R}) \to \mathbb{C}$, we define nine new tempered distributions as follows.

(i) The *translate* $\tau_h T$ of T by $h \in \mathbb{R}$, defined by
$$(\tau_h T)(\phi) := T(\tau_{-h}\phi) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(ii) The *dilation* $D_s T$ the *dilation* of T by $s > 0$, defined by
$$(D_s T)(\phi) := T\big(sD_{1/s}(\phi)\big) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}),$$
where $D_s\phi(x) = s\phi(sx)$.

(iii) The *derivative* T' of T, defined by
$$(T')(\phi) := -T(\phi') \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(iv) The *Fourier transform* $\widehat{T}$ of T, defined by
$$(\widehat{T}\,)(\phi) := T(\widehat{\phi}\,) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(v) The *inverse Fourier transform* $(T)^\vee$ of T, defined by
$$(T)^\vee(\phi) := T(\phi^\vee) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(vi) The *product* $M_g T$ of T with a function g such that if $\phi \in \mathcal{S}(\mathbb{R})$ then $g\phi \in \mathcal{S}(\mathbb{R})$ (g could be a function on the Schwartz class, a bounded C^∞ function like $e^{2\pi i x\xi}$, or an unbounded C^∞ that does not increase too fast, such as a polynomial).
$$(M_g T)(\phi) := T(g\phi) = T(M_g\phi) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(vii) The *reflection* $\widetilde{T}$ of T defined by,
$$(\widetilde{T})(\phi) := T(\widetilde{\phi}) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(vii) The *conjugate* $\overline{T}$ of T defined by,
$$(\overline{T})(\phi) := T(\overline{\phi}) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

(ix) The *convolution* $\psi * T$ of T with $\psi \in \mathcal{S}(\mathbb{R})$, defined by
$$(\psi * T)(\phi) := T(\widetilde{\psi} * \phi) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

The new functionals defined in (i)–(ix) are linear. It can be shown that they are also continuous. So we have indeed defined new tempered distributions, starting from a given T.

EXERCISE 8.16. Verify that if $T \in \mathcal{S}'(\mathbb{R})$ then the nine linear functionals defined in (i)-(ix) are in $\mathcal{S}'(\mathbb{R})$. $\Diamond$

It takes some time to get used to working with these definitions. 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 a modification of a tempered distribution by pushing the modification across to the Schwartz function. But we should be careful to which Schwartz function we pass it on, the argument ϕ? NO.

Let us try to understand with an example, differentiation. Since $\mathcal{S}(\mathbb{R})$ is embedded in $\mathcal{S}'(\mathbb{R})$ via the mapping $f \to T_f$, $f \in \mathcal{S}(\mathbb{R})$ we would like to be able to identify a new distribution, T_f', the *derivative* of T_f, with $T_{f'}$, so that the differentiation defined for distributions reduces to good old differentiation of functions when the distribution under consideration is given by integration by the function f. With that in mind, we can attempt to give a meaning to T' for general tempered distributions, not just the ones given by integration. For instance, to understand the derivative T' of a tempered distribution T, we need to know what T' does to each $\phi \in \mathcal{S}(\mathbb{R})$, how are we going to figure it out? We compute $T_{f'}(\phi)$, and attempt to write the output as T_f acting on some modification of ϕ involving, hopefully, ordinary differentiation. In this case, this is nothing more than integration by parts, where the boundary terms disappear because the functions decay at infinity:

$$T_{f'}(\phi) = \int f'(x)\phi(x)\,dx = -\int f(x)\phi'(x)\,dx = -T_f(\phi').$$

Now we keep the left and right ends of these string of inequalities and use them to define T' for any tempered distribution T:

$$T'(\phi) := -T(\phi').$$

The process of moving from a function f to its induced tempered distribution 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 the tempered distribution T_f induced by 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, as for differentiation, $\tau_h T(\phi) \neq T(\tau_h \phi)$. In a few instances, like Fourier transform and multiplication by functions, the operation "migrates" unchanged to the argument. This is only becuase those operations are "self-adjoint', meaning that $\int (\mathcal{O}f)\,g = \int f\,(\mathcal{O}g)$.

Example 8.17 and Exercise 8.18 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 8.17. (*Forming the Tempered Distribution of a Function Respects Convolution.*) We verify that for Schwartz functions f and ψ, the convolution of ψ with the tempered distribution T_f induced by 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 $\phi \in \mathcal{S}(\mathbb{R})$,

$$\begin{aligned} T_{\psi * f}(\phi) &= \int_{\mathbb{R}} \phi(y)(\psi * f)(y)\, dy \\ &= \int_{\mathbb{R}} \phi(y) \left(\int_{\mathbb{R}} f(x)\psi(y-x)\, dx \right) dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x)\psi(y-x)\phi(y)\, dy\, dx \\ &= \int_{\mathbb{R}} f(x) \left(\int_{\mathbb{R}} \widetilde{\psi}\,(x-y)\phi(y)\, dy \right) dx \\ &= \int_{\mathbb{R}} f(x)(\widetilde{\psi} * \phi)(x)\, dx \\ &= T_f(\widetilde{\psi} * \phi) \;=\; (\psi * T_f)(\phi), \end{aligned}$$

as required. $\Diamond$

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 $\psi \in \mathcal{S}(\mathbb{R})$ and $T \in \mathcal{S}'(\mathbb{R})$ can be identified with an infinitely differentiable function that grows slower than some polynomial. Specifically, there exists a function $f : \mathbb{R} \to \mathbb{C}$, $f \in C^{\infty}(\mathbb{R})$, and a polynomial $P(x)$ such that $|f(x)| \le P(|x|)$ for all $x \in \mathbb{R}$, and such that $\psi * T = T_f$, where f depends of course on ψ and T. In fact, f is the function defined by $f(y) = T(\tau_y \widetilde{\psi})$. See [**Graf**, pp.116–117] for details.

Exercise 8.18. (*The Tempered Distribution of a Function Respects the Fundamental Operations*) Check that for all $f \in \mathcal{S}(\mathbb{R})$, $h \in \mathbb{R}$, $s > 0$, and $g \in C^{\infty}$ such that $gf \in \mathcal{S}(\mathbb{R})$ for all $f \in \mathcal{S}(\mathbb{R})$, we have

$$\tau_h(T_f) = T_{\tau_h f}, \quad D_s(T_f) = T_{D_s f}, \quad \widehat{T_f} = T_{\widehat{f}}, \quad (T_f\,)^{\vee} = T_{(f)^{\vee}},$$
$$\widetilde{T_f} = T_{\widetilde{f}}, \quad \overline{T_f} = T_{\overline{f}}, \quad \text{and } (M_g T_f) = T_{M_g f}.$$

$\Diamond$

Definitions (iv) and (v) show that the Fourier transform can be extended to be a bijection on $\mathcal{S}'(\mathbb{R})$, because it is a bijection on $\mathcal{S}(\mathbb{R})$. Just observe that

$$[\widehat{T}]^{\vee}(\phi) = T([\widehat{\phi}]^{\vee}) = T(\phi).$$

Similarly $\widehat{[T]^{\vee}} = T$.

We can now build a time–frequency dictionary on $\mathcal{S}'(\mathbb{R})$, displayed in Table 8.1 that is in a 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, how they interact with the Fourier transform.

Example 8.19. As an example of how to derive the time–frequency dictionary encoded in Table 8.1, we derive the multiplication formula for the convolution. Consider $\psi, \phi \in \mathcal{S}(\mathbb{R})$, apply the definitions of Fourier transform of a distribution, and of convolution with a distribution,

$$\widehat{\psi * T}(\phi) = \psi * T(\widehat{\psi}) = T(\widetilde{\psi} * \widehat{\phi}). \tag{8.4}$$

TABLE 8.1. The time–frequency dictionary in $\mathcal{S}'(\mathbb{R})$.

	Time	**Frequency**
(a)	linear properties $aT_1 + bT_2$	linear properties $a\widehat{T_1} + b\widehat{T_2}$
(b)	translation $\tau_h T(\phi) := T(\tau_{-h}\phi)$	modulation $\widehat{\tau_h T} = M_{-h}\widehat{T}$
(c)	modulation $M_h T(\phi) := T(M_h\phi)$	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 $\widehat{T'} = [2\pi i\xi]\widehat{T}$
(h)	multiply by polynomial $M_{-2\pi ix}T(\phi) := T(-2\pi ix\phi(x))$	derivative $[M_{-2\pi ix}T]^\wedge = \widehat{T}'$
(i)	convolution $\psi * T(\phi) := T(\widetilde{\psi} * \phi)$	product $\widehat{\psi * T} = M_{\widehat{\psi}}\widehat{T}$

Let us look more closely at the argument on the right hand side of (8.4), and we will see that it is the Fourier transform of $\widehat{\psi}\phi$,

$$\begin{aligned}
\widetilde{\psi} * \widehat{\phi}(\xi) &= \int \widetilde{\psi}(s)\widehat{\phi}(\xi - s)\,ds \\
&= \int \widetilde{\psi}(s)\left(\int \phi(x)e^{-2\pi ix(\xi-s)}\,dx\right)ds \\
&= \int\left(\int \psi(-s)e^{-2\pi ix(-s)}\,ds\right)\phi(x)e^{-2\pi ix\xi}\,dx \\
&= \int \widehat{\psi}(x)\phi(x)e^{-2\pi ix\xi}\,dx \\
&= \widehat{\widehat{\psi}\phi}.
\end{aligned}$$

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 get that for all $\psi \in \mathcal{S}(\mathbb{R})$,

$$\widehat{\psi * T}(\phi) = T(\widehat{\widehat{\psi}\phi}) = \widehat{T}(\widehat{\psi}\phi) = M_{\widehat{\psi}}\widehat{T}(\phi).$$

This is exactly the multiplication formula we were seeking:

$$\widehat{\psi * T} = M_{\widehat{\psi}}\widehat{T}.$$

$\Diamond$

EXERCISE 8.20. Derive the formulas in the frequency column of Table 8.1. It will be useful to consult Table 7.1, the corresponding time–frequency dictionary for the Schwartz class. $\Diamond$

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 Exercise 8.12, 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 defines a new tempered distribution $\widehat{T_1}$ 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 of the Fourier Inversion Formula for Schwartz functions (Theorem 7.31).

DEFINITION 8.21. The *delta distribution*, $\delta : \mathcal{S}(\mathbb{R}) \to \mathbb{C}$, is defined by

$$\delta(\phi) := \widehat{T_1}(\phi\,) = \phi(0) \qquad \text{for } \phi \in \mathcal{S}(\mathbb{R}).$$

$\Diamond$

It is clear from the definition that δ is linear. It can be shown that δ is also continuous, and so δ is indeed a tempered distribution. In fact, suppose $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$, then we want to verify that the complex numbers $\delta(\phi_n) = \phi_n(0) \to 0$ as $n \to \infty$. But $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$ implies that in particular $\rho_{0,0}(\phi_n) = \sup_{x\in\mathbb{R}} |\phi_n(x)| \to 0$, clearly $|\phi_n(0)| \le \rho_{0,0}(\phi_n)$ hence it must go to zero as desired.

What is its Fourier transform $\widehat{\delta}$? We compute that

$$\widehat{\delta}\,(\phi) = \delta(\widehat{\phi}\,) = \widehat{\phi}\,(0) = \int_{\mathbb{R}} \phi(x)\, dx = T_1(\phi) \qquad \text{for } \phi \in \mathcal{S}(\mathbb{R}).$$

That is, the Fourier transform $\widehat{\delta}$ of δ is the tempered distribution T_1 induced by the function $f(x) \equiv 1$. We write $\widehat{\delta} = T_1$, or sometimes $\widehat{\delta} = 1$ for short.

EXERCISE 8.22. Show that the *inverse* Fourier transform of δ is also given by T_1,

$$[\delta]^{\vee}(\phi) = T_1(\phi) = \widehat{\delta}\,(\phi) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}).$$

$\Diamond$

Why isn't the delta distribution a true function? For a moment let's try to think of the delta distribution as a function, written $\delta(x)$ say, and called the *delta function*. Then the distribution induced by $\delta(x)$ should give us back the delta *distribution* again, and so we would have

$$\int_{\mathbb{R}} \delta(x)\phi(x)\,dx = \phi(0) \qquad \text{for all } \phi \in \mathcal{S}(\mathbb{R}). \tag{8.5}$$

In particular the integral on the left must be zero 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 (8.5). One can use measure theory to extend this idea, concluding that $\delta(x)$ can only be non-zero at a small collection of points x, specifically on a set E of measure zero, and further that we can assume $\delta(x) = 0$ at all points except $x = 0$, since even if δ is non-zero on E, the set is so small that the values of δ on E do not contribute anything to the integral. (To see the idea, think of E as being just a single point x_0.) Thus $\delta(x)$ would have to be a function whose mass is concentrated at zero. In other words, the support[1] of the delta function $\delta(x)$ should be the singleton set $\{0\}$. To reiterate, the delta function should take the value zero at all $x \neq 0$. But what value can we assign to it at the special point $x = 0$ so that integration against $\delta(x)$ will pick out the value of ϕ at 0? If we assign a finite value there, then $\int_{\mathbb{R}} \delta\phi = 0$ rather than 1. Therefore $\delta(0)$ cannot be a finite number, and so $\delta(0)$ must be infinity. Thus the delta function is not a function in the usual sense.

Fortunately, all is not lost: one can interpret δ as a distribution as we do in this section, or as a *point-mass measure* in the language of measure theory [**Fol**].

It is interesting to note 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. This observation illustrates a time–frequency principle.

> *The smaller the support of the function, the larger the support of its Fourier transform, and vice versa.*

This idea will be made precise by the *Heisenberg Uncertainty Principle* in Section 8.5.3.

EXERCISE 8.23. (*The Derivatives of the Delta Distribution.*) Find the derivatives $D^k\delta$, $k \geq 1$, of the delta distribution. Check that their Fourier transforms $\widehat{D^k\delta}$ 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 $x > 0$. ♢

We see that although the delta distribution is not a function, it can be viewed as the derivative of a function. The delta distribution is not an anomaly in this respect. It turns out that for *each tempered distribution*, no matter how wild, there is a function such that the given tempered distribution is the derivative of some

[1]One can define a notion of support of a distribution [**Str2**, Section 6.1] that extends the usual notion of the support of a function (Definition 7.3). With that notion the support of the delta distribution is the singleton set $\{0\}$. In fact, all distributions supported on $\{0\}$ are finite linear combinations of the delta distribution and its derivatives. See [**Str2**, Sections 6.2, 6.3] and also [**Graf**, p.122].

order (in the sense of distributions) of the function. For example, consider the distribution given by the k^{th} derivative of the delta distribution. By Exercise 8.23, $D^k\delta$ coincides with the $(k+1)^{\text{th}}$-derivative of the Heaviside function. See [**Str2**, Section 6.2].

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 role.

Exercise 8.24. (*The Delta Distribution Acts as an Identity Under Convolution.*) Verify that the delta distribution acts as an identity for the operation of convolution. More precisely, show that for all $\phi \in \mathcal{S}(\mathbb{R})$,

$$\phi * \delta = \phi \qquad \text{(as usual, we mean } = T_\phi\text{)}.$$

$\Diamond$

Exercise 8.25. (*Approximations of the Identity Converge to the Delta Distribution.*) Let $\{K_t(x)\}_{x\in\Lambda}$ with some t_0 be an approximation to the identity in $\mathbb{R}$. Prove that the distribution induced by K_t converges to the delta distribution as $t \to t_0$, in the sense of distributions, that is, for all $\phi \in \mathcal{S}(\mathbb{R})$, $T_{k_t}(\phi) \to \delta(\phi) = \phi(0)$ as $t \to t_0$. $\Diamond$

For those who have seen some measure theory, we note that every Borel measure μ on $\mathbb{R}$ defines a tempered distribution T_μ, via integration, similarly to the way in which equation (8.3) defines the tempered distribution induced by a function:

$$T_\mu(\phi) := \int_{\mathbb{R}} \phi(x)\, d\mu \qquad \text{for } \phi \in \mathcal{S}(\mathbb{R}).$$

In the next section we discuss some beautiful and very classical consequences of the Fourier theory. To close the chapter, we return to tempered distributions, Fourier transform and its interplay with the Lebesgue spaces $L^p(\mathbb{R})$. We also work out the details of a tempered distribution unlike the examples that we have discussed so far. This tempered distribution is not induced by a function, nor is it a linear combination of delta distributions or their derivatives. It is the so-called principal value distribution of $1/x$, the building block for the Hilbert transform.

8.5. Some applications of the Fourier transform

Among many applications of the Fourier transform we have chosen to present some of the most classical applications. The formulae 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 Shannon Sampling Formula), to quantum mechanics (Heisenberg's Uncertainty Principle). The proofs are beautiful and elegant. For many more applications to physics, partial differential equations, probability, and more, see the books by Strichartz [**Str2**], by Dym and McKean [**DM**], Stein and Shakarchi [**SS1**], and Reed and Simon [**RS**].

8.5.1. The Poisson Summation Formula. Given a function on the line, how can we construct a periodic function of period 1?

There are two methods. One is called *periodization*, and the other uses a Fourier series whose Fourier coefficients are given by the Fourier transform evaluated at the integers. The Poisson Summation Formula gives conditions under which these two methods coincide.

The periodization of a function f on the line is constructed by summing over its integer translates,

$$P_1\phi(x) = \sum_{n\in\mathbb{Z}} f(x+n) = \lim_{N\to\infty} \sum_{|n|\le N} \phi(x+n),$$

where the series is supposed to converge in the sense of symmetric partial sums. Notice that when ϕ is in $\mathcal{S}(\mathbb{R})$ or is a function of moderate decrease, the sum in the periodization is absolutely convergent.

EXAMPLE 8.26. If H_t is the *heat kernel* on the line, defined by formula (7.20), then its periodization is the periodic heat kernel

$$P_1 H_t(x) = (4\pi t)^{-1/2} \sum_{n\in\mathbb{Z}} e^{-(x+n)^2/4t}.$$

$\Diamond$

The second method assignes to a given function $f \in \mathcal{S}(\mathbb{R})$ a periodic function defined by,

$$P_2\phi(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n) e^{2\pi i x n}.$$

The series is well defined and convergent because $\phi \in \mathcal{S}(\mathbb{R})$.

EXAMPLE 8.27. If $f \in L^1(0,1)$ and is defined to be zero outside $(0,1)$, then both periodizations are well defined, and they coincide with its periodic extension to the whole line (in particular, $P_1 f = P_2 f$). $\Diamond$

The phenomenon displayed in the example is not an accident. The Poisson Summation Formula states precisely that.

THEOREM 8.28 (Poisson Summation Formula). *If $f \in \mathcal{S}(\mathbb{R})$ then*

$$P_1\phi(x) = \sum_{n\in\mathbb{Z}} \phi(x+n) = \sum_{n\in\mathbb{Z}} \widehat{\phi}(n) e^{2\pi i n x} = P_2\phi(x). \tag{8.6}$$

In particular, setting $x = 0$,

$$\sum_{n\in\mathbb{Z}} \phi(n) = \sum_{n\in\mathbb{Z}} \widehat{\phi}(n). \tag{8.7}$$

PROOF. It suffices to check that the right- and left-hand sides of equation (8.6) have the same Fourier coefficients. The n^{th} Fourier coefficient on the right-hand side is $\widehat{\phi}(n)$. Computing the left-hand side we get

$$\begin{aligned}
\int_0^1 \left(\sum_{m\in\mathbb{Z}} \phi(x+m) \right) e^{-2\pi i n x}\, dx &= \sum_{m\in\mathbb{Z}} \int_0^1 \phi(x+m) e^{-2\pi i n x}\, dx \\
&= \sum_{m\in\mathbb{Z}} \int_m^{m+1} \phi(y) e^{-2\pi i n y}\, dy \\
&= \int_{-\infty}^{\infty} \phi(y) e^{-2\pi i n y}\, dy \\
&= \widehat{\phi}(n).
\end{aligned}$$

We may interchange the sum and the integral in the first step since f is rapidly decreasing. $\square$

EXAMPLE 8.29. If H_t is the heat kernel, then all 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:

$$(4\pi t)^{-1/2} \sum_{n\in\mathbb{Z}} e^{-(x+n)^2/4t} = \sum_{n\in\mathbb{Z}} e^{-4\pi^2 tn^2} e^{2\pi inx}.$$

◊

The above calculation can be justified for functions of moderate decrease, and so the Poisson Summation Formula also holds for that larger class of functions.

EXERCISE 8.30. Use the Poisson Formula to show that the periodizations of the Poisson and Fejér kernels on the line coincide with the periodic Poisson and Fejér kernels, introduced in Sections 4.5 and 4.6. ◊

We can give an interpretation in the language of distributions of the Poisson Summation Formula. Recall that $\tau_n\delta(\phi) = \phi(n)$ for all $\phi \in \mathcal{S}(\mathbb{R})$. With this in mind, the left hand side of (8.7) can be writen as,

$$\sum_{n\in\mathbb{Z}} \phi(n) = \sum_{n\in\mathbb{Z}} \tau_n\delta(\phi).$$

Similarly for the right hand side,

$$\begin{aligned} \sum_{n\in\mathbb{Z}} \widehat{\phi}(n) &= \sum_{n\in\mathbb{Z}} \tau_n\delta(\widehat{\phi}), \\ &= \sum_{n\in\mathbb{Z}} \widehat{\tau_n\delta}(\phi), \end{aligned}$$

where in the last equality we used the definition of the Fourier transform for distributions. What the Poisson Summation Formula is saying is that the Fourier transform of the distribution $\sum_{n\in\mathbb{Z}} \tau_n\delta$ is the same distribution,

$$\widehat{\sum_{n\in\mathbb{Z}} \tau_n\delta} = \sum_{n\in\mathbb{Z}} \tau_n\delta.$$

EXERCISE 8.31. Verify that $\sum_{n\in\mathbb{Z}} \tau_n\delta$ is a tempered distribution. ◊

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 $e^{2\pi in\xi}$, see Exercise 8.39, so the apparently very complicated distribution $\sum_{n\in\mathbb{Z}} e^{2\pi in\xi}$ is nothing more than the infinite comb $\sum_{n\in\mathbb{Z}} \tau_n\delta$.

See [**Str2**, Section 7.3] for a beautiful application of the Poisson Summation Formula in two dimensions to crystals and quasicrystals.

8.5.2. The Shannon Sampling Formula.

The *Shannon Sampling Formula* states that *band-limited functions* (functions whose Fourier transform is supported on a compact interval) can be recovered from appropriate samples. We will need the so-called *sine kernel* function sinc, defined by

$$\operatorname{sinc}(x) := \frac{\sin(\pi x)}{\pi x}.$$

THEOREM 8.32 (Shannon Sampling Formula). *Suppose f is a function of moderate decrease such that*

$$\operatorname{supp} \widehat{f} \subset [-1/2, 1/2].$$

Then

$$f(x) = \sum_{n\in\mathbb{Z}} f(n) \operatorname{sinc}(x-n).$$

PROOF. Since the support of $\widehat{f}$ is the interval $[-1/2, 1/2]$, we can consider the periodization of $\widehat{f}$ to $\mathbb{R}$. Viewed as a 1-periodic function, we can compute its Fourier coefficients (notice that $\widehat{f}$ is the Fourier transform in $\mathbb{R}$, and in the next formula we are abusing the hat notation to denote also the Fourier coefficients of a 1-periodic function, $(g)^\wedge(n)$):

$$\begin{aligned}(\widehat{f}\,)^\wedge(n) &= \int_{-1/2}^{1/2} \widehat{f}\,(x)e^{-2\pi i n x}\,dx \\ &= \int_{-\infty}^{\infty} \widehat{f}\,(x)e^{2\pi i(-n)x}\,dx && \big(\widehat{f} \text{ is supported on } [-1/2,1/2]\big) \\ &= f(-n). && (\text{Fourier Inversion Formula on } \mathbb{R})\end{aligned}$$

The Fourier inversion formula was justified for functions in $\mathcal{S}(\mathbb{R})$ and for functions of moderate decrease such that $\widehat{f}$ is of moderate decrease. This entails verifying that $\widehat{f}$ is continuous, which it is true for all functions in $L^1(\mathbb{R})$ as a consequence of the Riemann-Lebesgue Lemma 8.45 (see also Exercise 8.46), and that it decreases at a certain rate. Notice that the band-limited assumption forces $\widehat{f}$ to vanish outside the interval $[-1/2, 1/2]$, hence it decays as fast as it could be. We can safely apply the Fourier inversion formula for band-limited functions of moderate decrease.

Therefore, on the interval $[-1/2, 1/2]$, $\widehat{f}$ coincides with its Fourier series:

$$\widehat{f}\,(\xi) = \sum_{n\in\mathbb{Z}} f(-n)e^{2\pi i n\xi} = \sum_{n\in\mathbb{Z}} f(n)e^{-2\pi i n\xi}, \qquad \text{for } \xi \in [-1/2, 1/2]. \tag{8.8}$$

The Fourier inversion formula on $\mathbb{R}$ also tells us that

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}\,(\xi)e^{2\pi i\xi x}\,d\xi.$$

Using the compact support of $\widehat{f}$ again, substituting formula (8.8), and doing a little more calculation, we can obtain the desired formula. More precisely,

$$\begin{aligned} f(x) &= \int_{-1/2}^{1/2} \widehat{f}\,(\xi)e^{2\pi i\xi x}\,d\xi \;=\; \int_{-1/2}^{1/2} \sum_{n\in\mathbb{Z}} f(n)e^{2\pi i n\xi}e^{-2\pi i\xi x}\,d\xi \\ &= \sum_{n\in\mathbb{Z}} f(n) \int_{-1/2}^{1/2} e^{2\pi i n\xi}e^{-2\pi i\xi x}\,d\xi \;=\; \sum_{n\in\mathbb{Z}} f(n) \int_{-1/2}^{1/2} e^{2\pi i(x-n)\xi}\,d\xi \\ &= \sum_{n\in\mathbb{Z}} f(n) \operatorname{sinc}(x-n). \end{aligned}$$

The interchange of sum and integral is justified because there is uniform convergence. The last identity holds because

$$\operatorname{sinc}(x) = \int_{-1/2}^{1/2} e^{2\pi i x\xi}\,d\xi. \qquad \square$$

The assumption that the support of the Fourier transform lies in an interval of length 1 is "artificial". 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 8.33. Suppose that f is a function of moderate decrease and $\operatorname{supp} \widehat{f} \subset [-L/2, L/2]$. Show that

$$f(x) = \sum_{n\in\mathbb{Z}} f(n/L) \ \operatorname{sinc}\big((x-n)/L\big).$$

$\diamondsuit$

The larger the support of $\widehat{f}$, the more we will have to sample.

This statement is reasonable, for if only frequencies satisfying $|\xi| < L/2$ are present in f, then it is plausible that sampling at the $1/L$ rate will suffice. Sampling at a larger rate would not suffice.

This formula has many applications in signal processing. See [**Sha**].

8.5.3. Heisenberg's Uncertainty Principle. It is impossible to find a function that is simultaneously well localized in time and in frequency. This idea can be carried one step further: there is no function that is both compactly supported and band-limited.

THEOREM 8.34 (Heisenberg's Uncertainty Principle). *Suppose that $\psi \in \mathcal{S}(\mathbb{R})$, and that ψ is normalized in $L^2(\mathbb{R})$: $\|\psi\|_2 = 1$. Then*

$$\left(\int_{\mathbb{R}} x^2|\psi(x)|^2\,dx\right)\left(\int_{\mathbb{R}} \xi^2|\widehat{\psi}\,(\xi)|^2\,d\xi\right) \geq \frac{1}{16\pi^2}.$$

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

$$\frac{d}{dx}|\psi|^2 = \frac{d}{dx}\psi\overline{\psi} = \overline{\psi}\psi' + \overline{\overline{\psi}\psi'} = 2\mathrm{Re}(\overline{\psi}\psi').$$

Also remember that $\mathrm{Re} z \leq |z|$, and that $|\int f| \leq \int |f|$. Hence

$$1 = 2\int_{\mathbb{R}} x\mathrm{Re}\big[\overline{\psi(x)}\psi'(x)\big]\,dx \leq 2\int_{\mathbb{R}} |x||\psi(x)||\psi'(x)|\,dx.$$

Now apply the Cauchy–Schwarz inequality to the right-hand side obtaining

$$1 \leq 2\left(\int_{\mathbb{R}} |x\psi(x)|^2\,dx\right)^{1/2}\left(\int_{\mathbb{R}} |\psi'(x)|^2\,dx\right)^{1/2}.$$

But by Plancherel and property (g) of the time–frequency dictionary,

$$\int_{\mathbb{R}} |\psi'(x)|^2\,dx = 4\pi^2\int_{\mathbb{R}} \xi^2|\widehat{\psi}\,(\xi)|^2\,d\xi.$$

This completes the proof. $\square$

The Uncertainty Principle holds for the class of functions of moderate decrease, and for the even larger class $L^2(\mathbb{R})$ of square-integrable functions.

EXERCISE 8.35. Check that equality holds in Heisenberg's Uncertainty Principle if and only if

$$\psi(x) = \sqrt{2B/\pi}\, e^{-Bx^2} \qquad \text{for some } B > 0.$$

Hint: Remember that equality holds in the Cauchy–Schwarz inequality if and only if one of the functions is a constant multiple of the other. $\Diamond$

EXERCISE 8.36. Check that under the conditions of Theorem 8.34,

$$\left(\int_{\mathbb{R}} (x - x_0)^2 |\psi(x)|^2 \, dx\right) \left(\int_{\mathbb{R}} (\xi - \xi_0)^2 |\widehat{\psi}(\xi)|^2 \, d\xi\right) \geq \frac{1}{16\pi^2}$$

for all x_0, $\xi_0 \in \mathbb{R}$. **Hint:** Use Heisenberg's Uncertainty Principle and the time–frequency dictionary for $\psi_{x_0,\xi_0}(x) = e^{2\pi i x \xi_0}\psi(x + x_0) \in \mathcal{S}(\mathbb{R})$. $\Diamond$

EXERCISE 8.37. Prove the following inequality:

$$\int_{\mathbb{R}} x^2 |\psi(x)|^2 \, dx + \int_{\mathbb{R}} \xi^2 |\widehat{\psi}(\xi)|^2 \, d\xi \geq \frac{1}{2\pi}$$

for $\psi \in \mathcal{S}(\mathbb{R})$. $\Diamond$

EXERCISE 8.38. Show that there is no function that is compactly supported both in time and in frequency, as follows. Suppose f is continuous on $\mathbb{R}$. Show that f and $\widehat{f}$ cannot both be compactly supported, unless $f = 0$. **Hint:** Assume f is supported on $[0, 1/2]$, expand f in a Fourier series on $[0, 1]$, and observe that f must be a trigonometric polynomial, but trigonometric polynomials can not vanish on an interval. $\Diamond$

The trigonometric functions are not in $L^2(\mathbb{R})$. However we can view them as distributions, and we can compute their Fourier transforms.

EXERCISE 8.39. 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, δ_ξ, defined by $\delta_\xi(\phi) = \phi(\xi)$. $\Diamond$

The preceding exercise also illustrates the time–frequency localization principle.

EXERCISE 8.40. Prove the analogues of the previous exercise in the finite-dimensional context, and in the Fourier-series context. Which vectors/sequences play the role of the delta distribution? $\Diamond$

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}. This says that the support on the *phase plane* or *time–frequency plane* will be essentially a *rectangle* of area 1 and dimensions $d \times d^{-1}$.

EXERCISE 8.41. Draw the phase planes for the trigonometric basis, the standard basis, and the Haar basis in $\mathbb{C}^8$ introduced in Chapter 6. $\Diamond$

8.6. $L^p(\mathbb{R})$ as distributions

Functions in the classical *Lebesgue spaces* $L^p(\mathbb{R})$, for real numbers p with $1 \leq p < \infty$, also define tempered distributions. The space $L^p(\mathbb{R})$ consists of those functions such that $\int_{\mathbb{R}} |f(x)|^p \, dx < \infty$. Here the integral is in the sense of Lebesgue.

In Section 2.1 we discussed the analogous spaces $L^p(\mathbb{T})$ of L^p functions on the unit circle. All the $L^p(\mathbb{R})$ spaces are *normed*, with norm given by

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

The $L^p(\mathbb{R})$ spaces are also *complete*, meaning that every Cauchy sequence in the space converges to an element in the space. Complete normed spaces are also known as *Banach spaces*. The space $L^2(\mathbb{R})$ of square-integrable functions on the line is a *Hilbert space*, with *inner product* given by

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

Functions in $L^p(\mathbb{R})$ are always locally integrable. In symbols we write $L^p(\mathbb{R}) \subset L^1_{\mathrm{loc}}(\mathbb{R})$.

The Schwartz class is *dense* in each $L^p(\mathbb{R})$ (see Appendix A). 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,

$$\lim_{n\to\infty} \|f - \phi_n\|_p = 0.$$

Or equivalently, given $\epsilon > 0$ there exists a $\phi \in \mathcal{S}(\mathbb{R})$ such that

$$\|f - \phi\|_p < \epsilon.$$

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

The idea that the Schwartz functions are dense in $L^p(\mathbb{R})$ leads to another way to think of the space $L^p(\mathbb{R})$. We have already mentioned that we can equip $\mathcal{S}(\mathbb{R})$ with the L^p-norm, making it into a normed space, although this space is not complete. If we "completed" the space by adding equivalence classes of Cauchy sequences in $\mathcal{S}(\mathbb{R})$ with respect to the L^p-norm, in the same way as one constructs the real numbers from the rational numbers, then at the end of this completion process (and there is an end to it) we obtain the Lebesgue spaces $L^p(\mathbb{R})$.

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

$$\|f - g\|_p = 0, \qquad \text{or } \lim_{t\to t_0} \|f_t - f\|_p = 0.$$

Equality in the L^p-sense occurs if and only if $f = g$ except possibly on a set of measure zero, that is $f = g$ a.e.

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 (8.3) must be well defined for all $f \in L^p(\mathbb{R})$ and $\phi \in \mathcal{S}(\mathbb{R})$. In fact the integral $\int_{\mathbb{R}} f(x)g(x)\,dx$ 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 $\frac{1}{p} + \frac{1}{q} = 1$ (in particular if $\psi \in \mathcal{S}(\mathbb{R}) \subset L^q(\mathbb{R})$ for all $1 \le q \le \infty$). This is a consequence of the celebrated *Hölder's Inequality* (proved in Section 11.4). Hölder's Inequality says that if $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$ with $\frac{1}{p} + \frac{1}{q} = 1$, then $fg \in L^1(\mathbb{R})$. Furthermore,

$$\|fg\|_1 = \int_{\mathbb{R}} |f(x)g(x)|\,dx \le \|f\|_p \|g\|_q.$$

This estimate implies that $\int_{\mathbb{R}} f(x)g(x)\,dx$ is well defined and finite. Furthermore, 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, once one notices that the L^p-norms of a function $\phi \in \mathcal{S}(\mathbb{R})$ can be controlled by the Schwartz seminorms, see Exercise 8.15. Notice that when $p = q = 2$ Hölder's inequality is nothing more than the Cauchy-Schwarz inequality in $L^2(\mathbb{R})$.

EXERCISE 8.42. Verify that if $f \in L^p(\mathbb{R})$ then the linear functional T_f defined by integration against f, is continuous, that is if $\phi_n \to 0$ in $\mathcal{S}(\mathbb{R})$ then the sequence of complex numbers $T_f(\phi) \to 0$. ◊

With this discussion in mind, we can compute the Fourier transform of L^p-functions in the sense of distributions. Can we identify these Fourier transforms with reasonable functions? There are two cases we have already discussed, at least in the cases of Schwartz functions and functions of moderate decrease.

- $\boxed{p=1}$ If $f \in \mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R})$, then
$$|\widehat{f}\,(\xi)| = \left|\int_{\mathbb{R}} f(x)e^{2\pi i x\xi}\,dx\right| \leq \int_{\mathbb{R}} |f(x)|\;dx.$$
It follows that, in the language of norms, $\|\widehat{f}\,\|_\infty \leq \|f\|_1$.
- $\boxed{p=2}$ 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 we can identify Fourier transforms of L^1-functions with bounded functions (actually with the subset of continuous functions vanishing at infinity), and the Fourier transform is a bijection in $L^2(\mathbb{R})$.

EXERCISE 8.43. The conjugate Poisson kernel $Q_y(x)$ (see equation (7.22)) is in $L^2(\mathbb{R})$ for each $y > 0$, therefore it can be considered as a tempered distribution by integration against it, and we can compute its Fourier transform. Check that
$$\widehat{Q_y}(\xi) = -i\ \mathrm{sgn}(\xi)\, e^{-2\pi|y\xi|}.$$
Compare with the result of Exercise 8.52. ◊

The following example shows that even in $\mathcal{S}(\mathbb{R})$, if we hope for an inequality of the type
$$\|\widehat{f}\,\|_q \leq C\|f\|_p, \qquad \text{for all } f \in \mathcal{S}(\mathbb{R}), \tag{8.9}$$
then p and q must be conjugate exponents: $\frac{1}{p} + \frac{1}{q} = 1$. Although it is not obvious from the example, it is also true that such an inequality will only hold for $1 \leq p \leq 2$.

EXAMPLE 8.44. Consider the following one-parameter family of functions in $\mathcal{S}(\mathbb{R})$, together with their Fourier transforms,
$$g_t(x) = e^{-\pi t x^2}, \qquad \widehat{g_t}(\xi) = \frac{1}{\sqrt{t}}e^{-\pi x^2/t}.$$
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
$$\|g_t\|_p^p = 1/\sqrt{tp}, \qquad \|\widehat{g_t}\|_q^q = (\sqrt{t})^{1-q}/\sqrt{q}.$$
According to Plancherel, when $p = q = 2$ the previous quantities should coincide, and indeed, they do. If we are hoping to obtain an inequality like (8.9), then we

must be able to bound the ratios $\|\widehat{g_t}\|_q/\|g_t\|_p$ independently of t, and that is only possible when p and q are conjugate exponents. In fact,

$$\frac{\|\widehat{g_t}\|_q}{\|g_t\|_p} = (\sqrt{t})^{\frac{1}{p}+\frac{1}{q}-1}\frac{(\sqrt{p})^{1/p}}{(\sqrt{q})^{1/q}}.$$

This ratio is bounded independently of t if and only if the exponent of $\sqrt{t}$ is zero. $\Diamond$

Table 8.2 summarizes how the Fourier transform acts on the L^p-spaces.

TABLE 8.2. The 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)	$L^q(\mathbb{R})$ $\frac{1}{p}+\frac{1}{q}=1$
$L^2(\mathbb{R})$	isometry $\|\widehat{f}\|_2 = \|f\|_2$ (Plancherel)	$L^2(\mathbb{R})$
$\mathcal{S}'(\mathbb{R})$	bijection	$\mathcal{S}'(\mathbb{R})$

The space of continuous functions that vanish at infinity is denoted by $C_0(\mathbb{R})$. It is a subset of $L^\infty(\mathbb{R})$, the space of *essentially bounded* functions on the line. The Riemann–Lebesgue Lemma is a strong statement. It says that the Fourier transform of an integrable function ($f \in L^1(\mathbb{R})$) is a *continuous* function, such that $\lim_{|\xi|\to\infty} \widehat{f}(\xi) = 0$. However, the Fourier transform is not a surjective map from $L^1(\mathbb{R})$ to $C_0(\mathbb{R})$ (see [**Kra**, p. 112, Prop 2.3.15]).

LEMMA 8.45 (Riemann-Lebesgue Lemma). *Suppose $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is continuous and*

$$\lim_{|\xi|\to\infty} \widehat{f}(\xi) = 0.$$

EXERCISE 8.46. Prove the Riemann–Lebesgue Lemma for functions f of moderate decrease (remember that functions of moderate decrease are continuous). More precisely, prove that if f is of moderate decrease, then $\widehat{f}$ is continuous, and $\lim_{|\xi|\to\infty} \widehat{f}(\xi) = 0$. In particular $\widehat{f}$ is bounded. (We verified the inequality $\|\widehat{f}\|_\infty \leq \|f\|_1$ for $f \in \mathcal{S}(\mathbb{R})$. The same proof goes through for functions of moderate decrease, and also for functions in $L^1(\mathbb{R})$, using the theory of Lebesgue integration.) $\Diamond$

Warning: the class of functions of moderate decrease is *not* closed under the Fourier transform. In fact, there is a familiar function of moderate decrease whose Fourier transform is not integrable.

EXERCISE 8.47. Verify that the Fourier transform of $\chi_{[-1/2,1/2]}$ is not integrable, but that it is in $C_0(\mathbb{R})$. $\Diamond$

The Fourier transform is well defined on $L^1(\mathbb{R})$, but since $\widehat{f}$ is not necessarily integrable, 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 both $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ (see Appendix A), one can extend the Fourier transform to these spaces by continuity. In $L^2(\mathbb{R})$, Plancherel will be transferred by continuity as well. Are these extensions the same on $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$? The answer is yes, and they also coincide with the definition of the Fourier transform in those spaces in the sense of distributions. Notice that in $\mathbb{R}$, we do not have the nice ladder structure for L^p-spaces that we had on the circle; see Figure 2.4 in Section 2.1. In particular $L^2(\mathbb{R})$ is not a subset of $L^1(\mathbb{R})$, nor is $L^1(\mathbb{R})$ a subset of $L^2(\mathbb{R})$.

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 $\frac{1}{p} + \frac{1}{q} = 1$. This is a consequence of the Hausdorff–Young inequality, which we will prove later, in Corollary 11.27. The Fourier transform does not map $L^p(\mathbb{R})$ for $p > 2$ into a nice functional space. In particular, for each $p > 2$, there exists a function $f \in L^p(\mathbb{R})$ whose Fourier transform $\widehat{f}$ is not locally integrable, but functions in L^q are always locally integrable, hence $\widehat{f}$ is not in L^q for any $1 \le q \le \infty$. Fourier transforms of generic L^p-functions for $p > 2$ are tempered distributions, but unless more information is given about the function we cannot say more. What is so special about $L^p(\mathbb{R})$, $1 < p < 2$? It turns out that in that range it is true that $L^p(\mathbb{R}) \subset L^1(\mathbb{R}) + L^2(\mathbb{R})$, and one can then use the theory in $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$. We will come back to this issue when we discuss interpolation in Section 11.4.

EXERCISE 8.48. (*$L^p(\mathbb{R})$ and $L^q(\mathbb{R})$ are Not Nested*) Suppose $1 \le p < q \le \infty$. 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 is not in $L^q(\mathbb{R})$. $\Diamond$

To finish we will present an example of a distribution that is given neither by a function in L^p, nor by a point mass (like the delta distribution), nor by a measure. It is the *principal value* distribution of $1/x$, the building block of the Hilbert transform, discussed in more detail in Chapter 11.

EXAMPLE 8.49. (*The Principal Value Distribution* $1/x$.) Define a linear functional on $\mathcal{S}(\mathbb{R})$ by,

$$H_0(\phi) = \lim_{\epsilon \to 0} \int_{|y|>\epsilon} \frac{\phi(y)}{y}\, dy. \tag{8.10}$$

To verify the continuity of this functional, we need to show 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 $k \to \infty$. We need to take advantage of some "hidden cancellation". For each $\epsilon > 0$,

$$\int_{\epsilon \le |y| < 1} \frac{1}{y}\, dy = 0.$$

We can add zero to (8.10),

$$\begin{aligned}
\left|\int_{|y|>\epsilon} \frac{\phi_k(y)}{y}\,dy\right| &= \left|\int_{|x|>1} \frac{\phi_k(y)}{y}\,dy + \int_{\epsilon\le|y|<1} \frac{\phi_k(y)-\phi_k(0)}{y}\,dy\right| \\
&\le \int_{|y|>1} \frac{|\phi_k(y)|}{|y|}\,dy + \int_{\epsilon\le|y|<1} \frac{|\phi_k(y)-\phi_k(0)|}{|y|}\,dy \\
&\le \|y\phi_k\|_\infty \int_{|y|>1} y^{-2}dy + \|\phi_k'\|_\infty \int_{1\ge|y|>\epsilon} 1\,dy \\
&\le 2\big(\rho_{1,0}(\phi_k) + \rho_{0,1}(\phi_k')\big)
\end{aligned}$$

To verify that the right hand side can be made arbitrarily small for k large enough, we use the fact that $\rho_{0,1}(\phi_k) \to 0$ and $\rho_{1,0}(\phi_k) \to 0$ as $k \to \infty$ since $\phi_k \to 0$ in $\mathcal{S}(\mathbb{R})$. Therefore H_0 is a tempered distribution, the *principal value distribution*, $H_0 = \text{p.v.}\frac{1}{y}$. For each $x \in \mathbb{R}$ define a new tempered distribution by translation, $\tau_x H_0(\phi) = H_0(\tau_{-x}\phi)$. We can now define the so-called *Hilbert transform* of $\phi \in \mathcal{S}(\mathbb{R})$:

$$\begin{aligned}
H\phi(x) &:= \frac{1}{\pi} H_0(\tau_x\tilde{\phi}) \\
&= \frac{1}{\pi}\lim_{\epsilon\to 0}\int_{|y|>\epsilon} \frac{\tilde{\phi}(y-x)}{y}\,dy \\
&= \frac{1}{\pi}\lim_{\epsilon\to 0}\int_{|y|>\epsilon} \frac{\phi(x-y)}{y}\,dy \\
&= \frac{1}{\pi}\lim_{\epsilon\to 0}\int_{|x-t|>\epsilon} \frac{\phi(t)}{x-t}\,dt.
\end{aligned}$$

$\Diamond$

EXERCISE 8.50. The functions $x^{-1}\chi_{\{|x|>\epsilon\}}$ are bounded and define tempered distributions for each $\epsilon > 0$, that we will denote H_ϵ. Check that the limit in the sense of distributions as $\epsilon \to 0$ of H_ϵ is H_0. This means that for all $\phi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{\epsilon\to 0} H_\epsilon(\phi) = H_0(\phi).$$

(This is just a consequence of the definition of H_0.) $\Diamond$

Now we will compare the conjugate Poisson kernel to the principal value distribution $\frac{1}{x}$.

EXERCISE 8.51. Check that the limit as $y \to 0$ of the conjugate Poisson kernel $Q_y(x)$ in $\mathcal{S}'(\mathbb{R})$ is $\frac{1}{\pi}H_0$. This means that for all $\phi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{y\to 0}\int_{\mathbb{R}} Q_y(x)\phi(x)\,dx = \frac{1}{\pi}H_0(\phi).$$

Hint: Notice that by Exercise 8.50 it will suffice to show that $(\pi Q_t - H_t) \to 0$ in the sense of distributions as $t \to 0$, that is, for all $\phi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{t\to 0}\left(\int_{\mathbb{R}} Q_t(x)\phi(x)\,dx - \int_{|x|>t} \frac{\phi(x)}{x}\,dx\right) = 0.$$

$\Diamond$

EXERCISE 8.52. Using Exercises 8.51 and 8.43 calculate the Fourier transform of the tempered distribution $\frac{1}{\pi}H_0$. Use this calculation to justify the following formula for the Fourier transform of the Hilbert transform:

$$\widehat{H(\phi)}(\xi) = i\operatorname{sgn}\xi\widehat{\phi}(\xi).$$

$\Diamond$

CHAPTER 9

From Fourier to Haar

In this chapter we give a brief survey of the windowed Fourier transform, also known as the Gabor transform, and introduce the newest member of the family, wavelet analysis. We also discuss the Haar basis with great care, and compare Fourier and Haar analysis, as well as Fourier and Haar basis in L^p spaces.

9.1. The windowed Fourier transform, and Gabor bases

The continuous Fourier transform provides a tool for analyzing a function defined on the whole real line $\mathbb{R}$, but the exponentials cannot be viewed as a "countable basis" any more, since there is one for each $\xi \in \mathbb{R}$. Also, the trigonometric functions $\{e^{2\pi i\xi x}\}$ are not even in $L^2(\mathbb{R})$, although of course their restrictions to $\mathbb{T}$ were in $L^2(\mathbb{T})$. The windowed Fourier transform addresses this problem.

How can we obtain an *orthonormal basis* for $L^2(\mathbb{R})$? A simple solution would be 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. Let χ_A denote the characteristic function of a given set $A \subset \mathbb{R}$:

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

The functions

$$g_{n,k}(x) = e^{2\pi i n x}\chi_{[k,k+1)}(x), \qquad \text{for } n, k \in \mathbb{Z},$$

form an orthonormal basis for $L^2(\mathbb{R})$. They give us the so-called *windowed Fourier transform*, defined by

$$Gf(n,k) := \langle f, g_{n,k}\rangle = \int_k^{k+1} f(x)e^{-2\pi i n x}\, dx,$$

and the corresponding *reconstruction formula*, where equality holds in the L^2-sense,

$$f(x) = \sum_{j,k} Gf(n,k) g_{n,k}(x).$$

Notice that now 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.

EXERCISE 9.1. Verify that the family $\{g_{n,k}\}_{n,k\in\mathbb{Z}}$ is an orthonormal basis in $L^2(\mathbb{R})$. $\Diamond$

We can think of each function $\chi_{[k,k+1)}$ giving us a window through which to view the behavior of f on the interval $[k, k+1)$. Both the function $\chi_{[k,k+1)}$ and the associated interval $[k, k+1)$ are commonly called windows.

FIGURE 9.1. The figure shows $g_{n,k}$ for some values of n and k, and for $g = \chi_{[0,1)}$.

We could have used windows of varying sizes. More precisely, 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 use the corresponding L_k-Fourier basis. Then the functions

$$\frac{1}{\sqrt{L_k}} e^{-2\pi i n x / L_k} \chi_{[a_k, a_{k+1})}(x), \qquad \text{for } n, k \in \mathbb{Z},$$

form an orthonormal basis of $L^2(\mathbb{R})$. This generalization provides some adaptability of 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, however wherever the function does not fluctuate much, we may use wider windows. We can 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 might not be that accurate but hopefully it is only on a few small intervals. On the other hand, by adapting the windows to the function, we might lose the translation structure provided by having all the windows the same size.

Because of their discontinuities at k and $k+1$, the windows $\chi_{[k,k+1)}$ are called *sharp windows*. It turns out that in numerical calculations, these sharp windows produce at the edges the same *artifacts* that are seen when analyzing periodic functions at discontinuity points (Gibb's phenomenon, or "kinks" at the divisions between windows).

EXERCISE 9.2. Consider the hat function

$$f(x) = \begin{cases} 1 - |x| & -1 \le x < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Compute its windowed Fourier transform with windows the intervals $[k, k+1)$, and plot using MATLAB successive approximations. You should see "kinks" at $x = -1$, 0, 1. Do the same with windows on the intervals $[2k-1, 2k+1)$. Do you see any kinks? How about windows $[k/2, (k+1)/2)$? ◊

To avoid Gibb's phenomenon, smoother windows are desirable. The sharp windows, $\chi_{[0,1)}$ and its modulated integer translates $e^{2\pi i n x}\chi_{[k,k+1)}(x)$, can be replaced by a *smooth window* g and its modulated integer translates

$$g_{n,k}(x) = g(x-k)e^{2\pi i n x}, \qquad \text{for } n, k \in \mathbb{Z}. \tag{9.1}$$

When a function $g \in L^2(\mathbb{R})$ is such that $\{g_{n,k}\}_{n,k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$, we call the function a *Gabor function*, and the basis a *Gabor basis*. In 1946, Gabor considered systems of this type and proposed using them in communication theory [**Gab**].

FIGURE 9.2. The figure shows $g_{n,k}$ for some values of n and k, and for g a smooth window.

The polarization formula (7.26) for square integrable functions ($\langle f, g\rangle = \langle \widehat{f}, \widehat{g}\rangle$) shows that a family of functions is an orthonormal basis if and only if the family of their Fourier transforms is itself an orthonormal basis. In other words,

$$\{\psi_n\}_{n\in\mathbb{N}} \quad \text{is an o.n. basis in } L^2(\mathbb{R}) \quad \Longleftrightarrow \quad \{\widehat{\psi_n}\}_{n\in\mathbb{N}} \quad \text{is an o.n. basis in } L^2(\mathbb{R}).$$

Orthonormality holds on one side if and only if it holds on the other, because for all n, $m \in \mathbb{N}$, $\langle \psi_n, \psi_m\rangle = \langle \widehat{\psi_n}, \widehat{\psi_m}\rangle$. 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 $\{g_{n,k}\}_{n,k\in\mathbb{Z}}$, generated by $g \in L^2(\mathbb{R})$ according to the scheme in equation (9.1), the family of their Fourier transforms $\{\widehat{g_{n,k}}\}_{n,k\in\mathbb{Z}}$ forms an orthonormal basis, by the above observation. What is remarkable is that the basis given by the Fourier transforms of a Gabor basis is again a Gabor basis. With the time–frequency dictionary in mind, we see on closer examination that the precise form of the modulated integer translates in equation (9.1) is exactly what is needed to achieve this property, since the Fourier transform converts translation to modulation, and vice versa. More precisely, notice that the Fourier transforms of the Gabor basis elements can be calculated using the time–frequency dictionary:

$$\widehat{g_{n,k}}(\xi) = \widehat{g}\,(\xi - n)e^{-2\pi i k\xi} = (\widehat{g}\,)_{-k,n}(\xi). \tag{9.2}$$

EXERCISE 9.3. Verify (9.2). $\Diamond$

To sum up:

> *A function $g \in L^2(\mathbb{R})$ generates a Gabor basis, in other words $\{g_{n,k}\}$ forms an orthonormal basis in $L^2(\mathbb{R})$, if and only if $\widehat{g} \in L^2(\mathbb{R})$ generates a Gabor basis, in other words $\{(\widehat{g})_{n,k}\}$ forms an orthonormal basis in $L^2(\mathbb{R})$.*

EXAMPLE 9.4. Since $g = \chi_{[0,1)}$ generates a Gabor basis, so does its Fourier transform

$$\widehat{g}\,(\xi) = (\chi_{[0,1)})^\wedge(\xi) = e^{-i\pi\xi}\,\frac{\sin(\pi\xi)}{\pi\xi} = e^{-i\pi\xi}\,\operatorname{sinc}(\xi).$$

This is an example of a differentiable window $\widehat{g}$, in contrast with our first window which was not even continuous. However, unlike g, $\widehat{g}$ is not compactly supported. $\Diamond$

Which other functions $g \in L^2(\mathbb{R})$ generate a Gabor basis? Can we find a Gabor function that is simultaneously smooth and compactly supported? The limitations of the Gabor analysis are explained by the following result.

THEOREM 9.5 (Balian–Low Theorem). *For $g \in L^2(\mathbb{R})$, if $\{g_{n,k}\}_{n,k\in\mathbb{Z}}$ is an orthonormal basis, then either*

$$\int_{\mathbb{R}} x^2|g(x)|^2\,dx = \infty, \qquad \textit{or} \qquad \int_{\mathbb{R}} \xi^2|\widehat{g}\,(\xi)|^2\,d\xi = \int_{\mathbb{R}} |g'(x)|^2\,dx = \infty.$$

EXERCISE 9.6. (*Examples of Balian–Low*) Verify the Balian–Low Theorem for the two examples discussed so far: $g(x) = e^{2\pi inx}\chi_{[k,k+1)}(x)$ and $g(x) = e^{-inx}\operatorname{sinc} x$. $\Diamond$

A proof of Theorem 9.5 can be found in [**Dau92**, p. 108, Theorem 4.1.1]. The theorem implies that a Gabor window or bell cannot be simultaneously compactly supported and smooth. The first example, $g(x) = \chi_{[0,1)}(x)$, is perfectly localized in time but is not even continuous, and the 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 $\chi_{[0,1)}$. This phenomenon is an incarnation of *Heisenberg's Uncertainty Principle*. However, if the exponentials are replaced by appropriate sines and cosines one can obtain Gabor-type bases with smooth bell functions. These are the so-called *local sine and cosine bases*, described by Coifman and Meyer [**CM**], but first discovered by Malvar [**Malv**]. For a very good discussion see the text by Hernández and Weiss [**HW**].

EXERCISE 9.7. Show that the Balian–Low Theorem implies that a Gabor function cannot be both smooth and compactly supported. $\Diamond$

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, normalized so that $\|g\|_2 = 1$, let

$$g_{\xi,u}(x) = g(x-u)e^{2\pi i\xi x}, \qquad u, \xi \in \mathbb{R}.$$

The Gabor transform is then

$$Gf(\xi, u) = \int_{\mathbb{R}} f(x)g(x-u)e^{-2\pi i\xi x}\,dx = \langle f, g_{u,\xi}\rangle.$$

The multiplication by the translated window localizes the Fourier integral in a neighborhood of u. The following *inversion formula* holds for $f \in L^2(\mathbb{R})$:

$$f(x) = \int_{\mathbb{R}}\int_{\mathbb{R}} Gf(u,\xi)g(x-u)e^{2\pi ix\xi}\,d\xi\,du = \int_{\mathbb{R}^2} \langle f, g_{\xi,u}\rangle g_{\xi,u}(x)\,d\xi\,du. \tag{9.3}$$

These formulae are similar in spirit to the Fourier transform and the inverse Fourier transform integral formulae in $\mathbb{R}$.

EXERCISE 9.8. Verify the inversion formula (9.3) for f, $g \in \mathcal{S}(\mathbb{R})$, g real-valued and symmetric ($g(x) = g(-x)$) with L^2-norm 1. **Hint:** Verify that (9.3) is equivalent to

$$f(x) = \int g_{\xi,0} * g_{\xi,0} * f(x)\,d\xi,$$

and then that the Fourier transform of the right-hand-side is $\widehat{f}$, using the time–frequency dictionary. $\Diamond$

9.2. The wavelet transform

Gabor bases give partial answers to the localization issues. A problem is that the size of the windows is fixed (whether all the windows have the same size or not, once a partition of $\mathbb{R}$ has been chosen it cannot be changed). Variable widths are the new ingredient added by wavelet analysis.

The wavelet transform involves *translations* (as in the Gabor basis) and *scalings* (instead of modulations). The resulting zooming mechanism lies behind the *multiresolution* structure of these bases that we will explore in detail in Chapter 10.

The goal is to find functions $\psi \in L^2(\mathbb{R})$ so that the family

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \qquad j,k \in \mathbb{Z}, \tag{9.4}$$

forms an orthonormal basis of $L^2(\mathbb{R})$.

EXERCISE 9.9. (*Fourier Transform of a Wavelet*) Suppose $\psi \in L^2(\mathbb{R})$. Use the time–frequency dictionary to compute $\widehat{\psi_{j,k}}(\xi)$. $\Diamond$

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.

The *orthogonal wavelet transform* is given by

$$Wf(j,k) = \langle f, \psi_{j,k}\rangle = \int_{\mathbb{R}} f(x)\overline{\psi_{j,k}(x)}\,dx,$$

and the following *reconstruction formula* holds in the L^2-sense,

$$f(x) = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}(x).$$

The oldest example of a wavelet basis is the *Haar basis*, introduced by Alfréd Haar in 1910 [**Haa**]. He succeeded in finding an orthonormal basis on $L^2([0,1))$ that, unlike the trigonometric basis, provides uniform convergence of its partial sums for continuous functions.

EXAMPLE 9.10. (*The Haar Basis*) The *Haar function* $h(x)$ on the unit interval is given by

$$h(x) := -\chi_{[0,1/2)}(x) + \chi_{[1/2,1)}(x).$$

The family $\{h_{j,k}(x) = 2^{j/2}h(2^j x - k)\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$. We will study this example in depth in Section 9.3. $\Diamond$

EXERCISE 9.11. Show that $\{h_{j,k}\}$ is an orthonormal set:

$$\langle h_{j,k}, h_{j',k'}\rangle = \begin{cases} 1, & \text{if } j = j' \text{ and } k = k'; \\ 0, & \text{otherwise.}\end{cases}$$

Also show that the functions $h_{j,k}$ have zero integral: $\int h_{j,k} = 0$. $\Diamond$

An important part of wavelet theory is the search for smoother wavelets. The Haar function is discontinuous and perfectly localized in time, but not perfectly localized in frequency.

EXERCISE 9.12. Find the Fourier transform of the Haar function $h(x)$. $\Diamond$

EXAMPLE 9.13. (*The Shannon Basis*) At the other end of the spectrum, one finds the *Shannon basis*. Let ψ be given on 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 $L^2(\mathbb{R})$. ◊

EXERCISE 9.14. Show that the Shannon functions $\{\psi_{j,k}\}$ form an orthonormal set, and have zero integral. Furthermore, they are a basis. **Hint:** Work on Fourier side using the polarization formula (7.26). It can be seen that on Fourier side we are dealing with a windowed Fourier basis, with *double-paned windows* $\omega_j := [-2^{j+1}, -2^j) \cup [2^j, 2^{j+1})$ of combined length 2^j, for $j \in \mathbb{Z}$, that are congruent[1] modulo 2^j with the interval $[0, 2^j)$. The collection of double-paned windows ω_j, provide a partition of $\mathbb{R}\backslash\{0\}$. On each double-paned window, the trigonometric functions

$$\frac{1}{2^{j/2}}e^{2\pi ikx2^{-j}}\chi_{\omega_j}(x), \quad k \in \mathbb{Z},$$

form an orthonormal basis of $L^2(\omega_j)$. ◊

The Shannon wavelets are perfectly localized in frequency, but not in space.

The goal is to find functions that are simultaneously localized in time and in frequency (within the limits imposed by the Heisenberg Principle). Compactly supported wavelets with arbitrary smoothness were constructed by I. Daubechies[2] in a fundamental paper in wavelet theory [**Dau88**]. We can associate to most wavelets a sequence of numbers, known as a *filter*. The filters of compactly supported wavelets are zero except for finitely many entries. In the engineering community, such filters are called *finite impulse response (FIR) filters*. The connections to filter bank theory and the possibility of implementing FIR filters opened the door to widespread use of wavelets in applications. We will explore this connection in Section 10.4.

We can develop the theory of wavelets in $\mathbb{R}^N$ or $\mathbb{C}^N$ (linear algebra!), in the same way as we built a finite Fourier theory and introduced the discrete Haar basis in Chapter 6. This is done in full detail in the book by Frazier [**Fra**]. This is what ends up being implemented. Not only is there a *Fast Fourier Transform* (FFT), but it turns out that there is also a *Fast Wavelet Transform* (FWT) which has been instrumental in the success of wavelets in the "real world". Let us just mention two of the most popular applications: both the FBI fingerprint data base and retrieval system, and the JPEG 2000 Standard for image compression, are based on wavelets.

As in the Fourier and Gabor cases, there is a *continuous wavelet transform*. In this case we have continuous translation and scaling parameters, $s \in \mathbb{R}^+$ (that is, $s > 0$), $u \in \mathbb{R}$, and a family of *time–frequency atoms* that is obtained by rescaling by s and shifting by u a normalized wavelet $\psi \in L^2(\mathbb{R})$ with zero average ($\int \psi = 0$).

Let

$$\psi_{s,u}(x) = \frac{1}{\sqrt{s}}\psi\left(\frac{x-u}{s}\right), \qquad (\psi_{s,u})^{\wedge}(\xi) = \sqrt{s}e^{-2\pi iu\xi}\widehat{\psi}(s\xi).$$

The continuous wavelet transform is then defined by

$$Wf(s,u) = \langle f, \psi_{s,u}\rangle = \int_{\mathbb{R}} f(x)\overline{\psi_{s,u}(x)}\,dx.$$

[1]The intervals $[-1/2, 0) \cup [1/2, 1)$ are not congruent modulo 1 to the interval $[0, 1)$.

[2]Ingrid Daubechies, 1954–, Belgian mathematician.

If ψ is real valued, and it is 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 orthonormal case, $u = k2^{-j}$ and $s = 2^{-j}$). As the scale s goes to zero, or equivalently as j goes to infinity, the decay of the wavelet coefficients characterizes the regularity of f near u (in the discrete case $k2^{-j}$). On the other hand, 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 or *phase-plane portrait* of the wavelets are rectangles of area one and dimensions $s \times 1/s$.

Under very mild assumptions on the wavelet ψ, we obtain a reconstruction formula. For any $f \in L^2(\mathbb{R})$,

$$f(x) = \frac{1}{C_\psi} \int_0^\infty \int_{-\infty}^{+\infty} Wf(s,u)\psi_{s,u}(x) \frac{duds}{s^2}, \tag{9.5}$$

provided that ψ satisfies *Calderón's admissibility condition* [**Cal**]:

$$C_\psi := \int_0^\infty \frac{|\widehat{\psi}(\xi)|^2}{\xi} \, d\xi < \infty.$$

The reconstruction formula can be traced back to the famous *Calderón reproducing formula* given by,

$$f(x) = \frac{1}{C_\psi} \int_0^\infty \psi_{s,0} * \overline{\widetilde{\psi_{s,0}}} * f(x) \, \frac{ds}{s^2}, \tag{9.6}$$

where $\widetilde{\psi}(x) = \psi(-x)$, and $*$ denotes convolution in $\mathbb{R}$.

Exercise 9.15. Verify that Calderón's reproducing formula (9.6), and the reconstruction formula (9.5) are the same. Verify now that (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 $\widehat{f}$. ◊

9.3. Haar analysis

In this section we discuss the Haar basis in detail. In particular we do show that the Haar functions form a complete orthonormal system. Verifying the orthonormality of the system is reduced to understanding the geometry of the dyadic intervals. Verifying the completeness of the system is reduced to understanding that the limit in $L^2(\mathbb{R})$ of the averaging operators over intervals as the interval shrinks to a point $x \in \mathbb{R}$ is the identity operator, and the limit as the intervals grow to be infinitely long is the zero operator.

For these to make sense we have to first define the dyadic intervals, and describe their geometry. We then define the expectation (or averaging) and difference operators, and reduce the question of completeness of the Haar system to understanding the limiting behaviour of the averaging operators. We prove the required limit results for continuous and compactly supported functions, finally an approximation argument coupled with some uniform bounds give the desired result.

9.3.1. The dyadic intervals.

Intervals of the form

$$I_{j,k} = [k2^{-j}, (k+1)2^{-j}), \qquad \text{for integers } j,\, k,$$

are called *dyadic intervals*. For instance, $[5/8, 3/4)$ and $[-16, -12)$ are dyadic intervals, while $[3/8, 5/8)$ is not. Each dyadic interval is half-open: $I_{j,k}$ contains its left endpoint but not its right. The collection of all dyadic intervals is denoted by $\mathcal{D}$, and $\mathcal{D}_j$ denotes the collection of dyadic intervals I of length 2^{-j}, also called the j^{th} *generation*. It is clear that each $\mathcal{D}_j$ forms a partition of the real line, and that

$$\mathcal{D} = \bigcup_{j\in\mathbb{Z}} \mathcal{D}_j.$$

Given two distinct intervals I, $J \in \mathcal{D}$, then either I and J are disjoint or one is contained in the other. Each dyadic interval I is in a unique generation $\mathcal{D}_j$, and there are exactly two subintervals of I in the next generation $\mathcal{D}_{j+1}$. These are the *children* of I, denoted by I_l and I_r for the *left* and *right* child respectively. See Figure 9.3. Clearly, $I = I_l \cup I_r$. We denote by $|I|$ the length of the interval I.

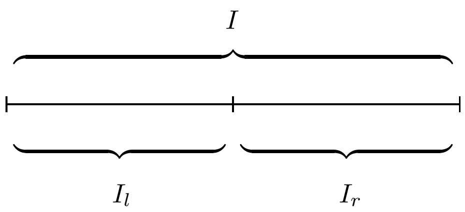


FIGURE 9.3. Parent dyadic interval I and its children I_l and I_r.

EXERCISE 9.16. (*Dyadic Intervals Are Nested Or Disjoint*) Show that if I, $J \in \mathcal{D}$, then exactly one of the following is true: $I \bigcap J = \emptyset$, or $I \subseteq J$, or $J \subset I$. $\Diamond$

Given $J \in \mathcal{D}$, we denote by $\mathcal{D}(J)$ the collection of dyadic intervals that are contained in J.

Given $x \in \mathbb{R}$ and $j \in \mathbb{Z}$, there is a unique interval $I \in \mathcal{D}_j$ such that $x \in I$. We denote this unique interval by $I_j(x)$.

9.3.2. The Haar basis. Associated to each dyadic interval I there is a *Haar function* h_I defined by

$$h_I(x) := \frac{1}{|I|^{1/2}}\big[\chi_{I_r}(x) - \chi_{I_l}(x)\big].$$

Figure 9.4 shows the graphs of two Haar functions.

EXERCISE 9.17. Let $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$. Show that

$$h_{I_{j,k}}(x) = 2^{j/2} h(2^j x - k) = h_{j,k}(x), \qquad \text{where } h = h_{[0,1)}.$$

$\Diamond$

In Exercise 9.11, you were asked to check that $\{h_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal set in $L^2(\mathbb{R})$. Exercise 9.17 shows that $\{h_I\}_{I\in\mathcal{D}}$ is the same set of functions; hence it is an orthonormal set as well. However we can prove directly now that the Haar functions indexed on the dyadic intervals form an orthonormal family using the fact that the dyadic intervals are nested or disjoint (Exercise 9.16). In fact, consider $I, J \in \mathcal{D}$, and $I \neq J$, then either they are disjoint or one is strictly contained in the other. If I and J are disjoint then clearly $\langle h_I, h_J\rangle = 0$ because the supports are disjoint. If, say I is strictly contained in J, then necessarily I is contained in one

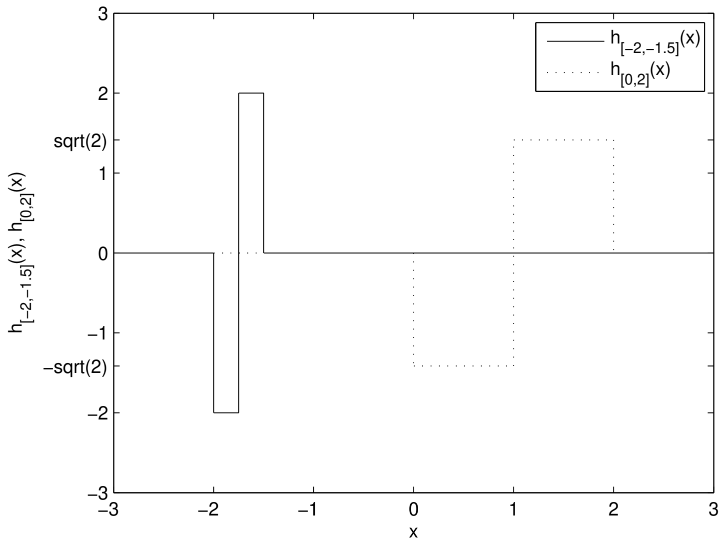


FIGURE 9.4. Graphs of the two Haar functions defined by $h_{[-2,-1.5)}(x) = 2\left[\chi_{[-1.75,-1.5)}(x) - \chi_{[-2,-1.75)}(x)\right]$ and $h_{[0,2)}(x) = \sqrt{2}\left[\chi_{[1,2)}(x) - \chi_{[0,1)}(x)\right]$.

of J^{th} children, say, to fix ideas the right child, but then, the inner product is the constant value of h_J on its right child times the integral of h_I which vanishes, i.e.

$$\langle h_I, h_J\rangle = \int_I h_I(x)h_J(x)\,dx = \frac{1}{|J|^{1/2}}\int h_I(x)\,dx = 0.$$

Not only the Haar system is orthonormal it is also a complete orthonormal system, hence a basis of $L^2(\mathbb{R})$.

THEOREM 9.18. *The Haar functions* $\{h_I\}_{I\in\mathcal{D}}$ *form an orthonormal basis in* $L^2(\mathbb{R})$.

In Chapter 6 we discussed the discrete Haar basis on $\mathbb{C}^N$. We had N Haar vectors that were proven to be orthonormal, and hence a basis. In the finite-dimensional case, we can count the elements in the orthonormal set, and if that number coincides with the dimension, we know we have a basis. In infinite-dimensional space we do not have that luxury. To prove Theorem 9.18, we must make sure the set is *complete*. This means that for all $f \in L^2(\mathbb{R})$, the following identity must hold in the L^2-sense[3]:

$$f = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I. \tag{9.7}$$

An alternative will be to show that the only square integrable function orthogonal to all Haar functions is the zero function (see Theorem B.24). We will show that both arguments boil down to checking some limit properties of the expectation operators defined in Section 9.3.3 below.

[3]Recall that equation (9.7) holds in the L^2-sense if $\|f - \sum_{I\in\mathcal{D}}\langle f, h_I\rangle h_I\|_2 = 0$.

Before we go on to prove the completeness of the Haar system, let us play the Devil's advocate.

- First, consider the function $f(x) = 1$. Then $\langle f, h_I\rangle = \int h_I = 0$ for all $I \in \mathcal{D}$. We have found a function that is orthogonal to all the Haar functions, so how can the system be complete? Are we contradicting the theorem? No, because the function that is identically equal to 1 on $\mathbb{R}$ is not in $L^2(\mathbb{R})$!
- Second, how can it be true that functions that have zero integral (the Haar functions) can reconstruct functions that do not have zero integral?[4]

If the Haar system is complete, then for any $f \in L^2(\mathbb{R})$, equation (9.7) holds. Integrating on both sides and interchanging the sum and the integral, we see that

$$\begin{aligned}\int_{\mathbb{R}} f(x)\,dx &= \int_{\mathbb{R}} \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I(x)\,dx\\ &= \sum_{I\in\mathcal{D}} \langle f, h_I\rangle \int_{\mathbb{R}} h_I(x)\,dx\\ &= 0,\end{aligned}$$

where the last equality holds because the Haar functions have integral zero. We seem to be implying that all functions in $L^2(\mathbb{R})$ must themselves have integral zero. But we know this is not true, since for example $\chi_{[0,1)} \in L^2(\mathbb{R})$ and has integral one. What's wrong? Perhaps the Haar system is not complete after all. Or is there something wrong in the above calculation? The Haar system *is* complete; it turns out that what is illegal above is the interchange of sum and integral.

For comfort, let us check that $\sum_{I\in\mathcal{D}} \langle \chi_{[0,1)}, h_I\rangle h_I$ coincides with $\chi_{[0,1)}$ pointwise, furthermore we will see that the partial sums converge uniformly to $\chi_{[0,1)}$.

Exercise 9.19. Verify that for $I \in \mathcal{D}$

$$\langle \chi_{[0,1)}, h_I\rangle = \begin{cases} \int_0^1 h_I(x)\,dx, & \text{if } [0,1) \subset I,\ I \neq [0,1);\\ 0, & \text{otherwise.}\end{cases}$$

◊

Which dyadic intervals I strictly contain the interval $[0, 1)$? Only those dyadic intervals of the form $I = [0, 2^n) =: I^n$, for $n \geq 1$. Notice that $|I^n| = 2^n$, and that the unit interval is contained on the left half of I^n for all $n \geq 1$, that is $[0,1) \subset I^n_l = [0, 2^{n-1})$. So if $x \in [0,1)$, then $h_{I^n}(x) = -1/\sqrt{2^n}$. This observation and Exercise 9.19 show that

$$\langle \chi_{[0,1]}, h_I\rangle = \begin{cases} -\frac{1}{\sqrt{2^n}}, & \text{if } I = I^n,\ n \geq 1;\\ 0, & \text{otherwise.}\end{cases}$$

[4] This question was posed by Lindsay Crowl, one of the participants in the 2004 Program for Women in Mathematics where we gave the lectures that lead to this book. We decided it was a perfectly natural concern, and furthermore it was very illustrative of the dangers of interchanging limiting operations!

Bearing in mind that $h_{I^n}(x) = \frac{1}{\sqrt{2^n}}\big[-\chi_{[0,2^{n-1})}(x) + \chi_{[2^{n-1},2^n)}(x)\big]$, we can now evaluate pointwise the series:

$$\begin{aligned}\sum_{I\in\mathcal{D}} \langle \chi_{[0,1]}, h_I\rangle h_I(x) &= \sum_{n=1}^{\infty} \langle \chi_{[0,1]}, h_{I^n}\rangle h_{I^n}(x)\\ &= \sum_{n=1}^{\infty} \frac{1}{2^n}\big[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\big].\end{aligned} \tag{9.8}$$

We claim that the last term in equation (9.8) is equal to $\chi_{[0,1)}(x)$ for each $x \in \mathbb{R}$. To justify this, let us evaluate the sum term by term for different values of x.

For $x < 0$**:** Clearly $h_{I^n}(x) = 0$ for all $n \geq 0$, and so

$$\sum_{n=1}^{\infty} \langle \chi_{[0,1]}, h_{I^n}\rangle h_{I^n}(x) = 0.$$

For $0 \leq x < 1$**:** We have already mentioned that $[0,1)$ sits on I_l^n, the left half of I^n, for all $n \geq 1$. Hence

$$\sum_{n=1}^{\infty} \frac{1}{2^n}\big[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\big] = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

For $1 \leq x < 2$**:** This time $[1,2) = I_r^1$, and $[1,2) \subset I_\ell^n$ for all $n \geq 2$. Therefore

$$\sum_{n=1}^{\infty} \frac{1}{2^n}\big[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\big] = -\frac{1}{2} + \sum_{n=2}^{\infty} \frac{1}{2^n} = 0.$$

For $2 \leq x < 2^2$**:** Here $[2,2^2) \cap I^1 = \emptyset$, $[2,2^2) = I_r^2$, and $[2,2^2) \subset I_\ell^n$ for all $n \geq 3$. Therefore

$$\sum_{n=1}^{\infty} \frac{1}{2^n}\big[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\big] = -\frac{1}{2^2} + \sum_{n=3}^{\infty} \frac{1}{2^n} = 0.$$

The pattern is now clear.

For $2^k \leq x < 2^{k+1}$**:** We have $[2^k,2^{k+1}) \cap I^n = \emptyset$ for $n < k$, $[2^k,2^{k+1}) = I_r^k$, and $[2^k,2^{k+1}) \subset I_\ell^n$ for all $n \geq k+1$. Therefore

$$\sum_{n=1}^{\infty} \frac{1}{2^n}\big[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\big] = -\frac{1}{2^k} + \sum_{n=k+1}^{\infty} \frac{1}{2^n} = 0.$$

Hence we obtain that, pointwise,

$$\sum_{I\in\mathcal{D}} \langle \chi_{[0,1]}, h_I\rangle h_I(x) = \chi_{[0,1)}(x).$$

Furthermore, we have uniform convergence on $\mathbb{R}$ of the functions

$$f_N(x) := \sum_{n=1}^{N} \frac{1}{2^n}\big[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\big],$$

to $\chi_{[0,1)}(x)$.

EXERCISE 9.20. Verify that

$$f_N(x) = \begin{cases} 1 - 2^{-N} & x \in [0,1) \\ -2^{-N} & x \in [1, 2^N) \\ 0 & \text{otherwise} \end{cases} .$$

Check that $\int_{\mathbb{R}} f_N(x)\,dx = 0$. ◊

EXERCISE 9.21. Verify that $f_N \to \chi_{[0,1)}$ uniformly on $\mathbb{R}$. ◊

It is clear that we cannot interchange limit and integral despite having the uniform convergence of f_N,

$$\lim_{N\to\infty} \int_{\mathbb{R}} f_N(x)\,dx = 0 \neq 1 = \int_{\mathbb{R}} \chi_{[0,1)}(x)\,dx.$$

But wait, doesn't uniform convergence guarantee the interchange of the limit and the integral? NO, only when integrating on a compact set, and the line is not compact.

EXERCISE 9.22. (i) Show that f_N cannot converge to $\chi_{[0,1)}$ in $L^1(\mathbb{R})$.
(ii) Show that f_N converges to $\chi_{[0,1)}$ in $L^2(\mathbb{R})$, that is

$$\|f_N - \chi_{[0,1)}\|_{L^2(\mathbb{R})} \to 0.$$

More precisely, verify that

$$\|f_N - \chi_{[0,1)}\|_{L^2(\mathbb{R})} = 2^{-N/2}.$$

◊

There is a very important result in the Lebesgue integration theory on $\mathbb{R}$ that we stated for intervals in Chapter 2, the Lebesgue Dominated Convergence Theorem. This theorem states that if a sequence of measurable functions f_n converges pointwise almost everywhere to a function f (which must be measurable because pointwise a.e. limits preserve measurability), AND there is an integrable function g that dominates pointwise all the functions, namely, $|f_n(x)| \leq g(x)$, then the interchange of limit and integral is legal, i.e. $\lim_{n\to\infty} \int f_n = \int f$. We know in our case, that the interchange does not hold, so it better be that there is NO dominating function g. That is the content of the following exercise.

EXERCISE 9.23. Verify that if $g : \mathbb{R} \to [0,\infty)$ has the property that $|f_N(x)| \leq g(x)$ for all $x \in \mathbb{R}$ and for all $N > 1$, then $g(x) \geq g_0(x)$, where $g_0(x) = 1$ if $x \in [0,1)$, $g_0(x) = 2^{-j}$ if $x \in [2^{j-1}, 2^j)$ for all $j \geq 1$, and zero when $x < 0$. Now verify that g_0 is not integrable, hence no dominating function g can be integrable. ◊

9.3.3. The expectation and difference operators, P_j and Q_j. We introduce here two important *operators*[5] that will help us to understand the zooming properties of the Haar basis.

The *expectation* operators $P_j : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $j \in \mathbb{Z}$ are nothing more than averages over dyadic intervals at generation j:

$$P_j f(x) := \frac{1}{|I|} \int_I f(t)\,dt, \tag{9.9}$$

[5]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 operator, and we discussed in Chapter 8 its mapping properties for different functional spaces. In particular its action on L^p-spaces is summarized in the table on page 149.

where I is the unique interval of length 2^{-j} containing x.

Notice that the new function $P_j f$ is a step function with steps on dyadic intervals $I \in \mathcal{D}_j$. Figure 9.5 shows the graph of a specific function f together with the graph of $P_j f$. In this example $j = -1$.

EXERCISE 9.24. Verify that

$$P_j f(x) = \sum_{I \in \mathcal{D}_j} m_I f \, \chi_I(x),$$

where $m_I f$ denotes the average of f on the interval I. $\lozenge$

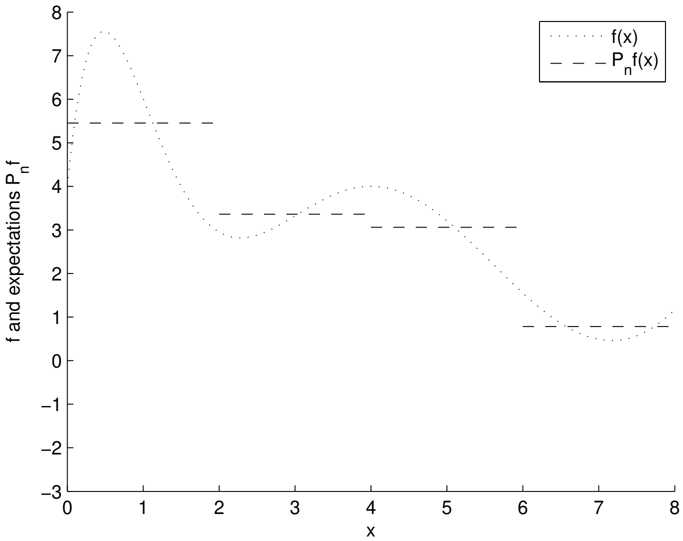


FIGURE 9.5. Graphs of f and $P_n f$. [***Add the formula for this f, also $n = j$.]

As $j \to \infty$, the size of the steps goes to zero, and we would expect $P_j f$ to be a better and better approximation of f. We will make this precise in Section 9.3.4 below.

The *difference operators* $Q_j : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $j \in \mathbb{Z}$ are given by

$$Q_j f(x) := P_{j+1} f(x) - P_j f(x). \tag{9.10}$$

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 9.6 shows the graph of the same function f as in Figure 9.5, together with the graph of $Q_j f$, again for $j = -1$. Figure 9.7 shows the graphs of f, $P_j f$, and $P_{j+1} f = P_j f + Q_j f$.

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, so that $Q_j f$ seems to be a linear combination of the Haar functions at scale j. This is indeed the case, and the next Lemma takes care of the details of this impressionistic comment.

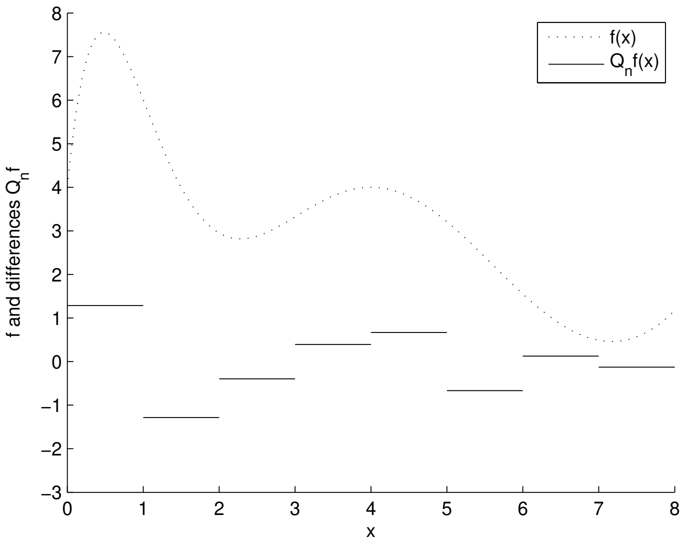


FIGURE 9.6. Graphs of f and $Q_n f$.

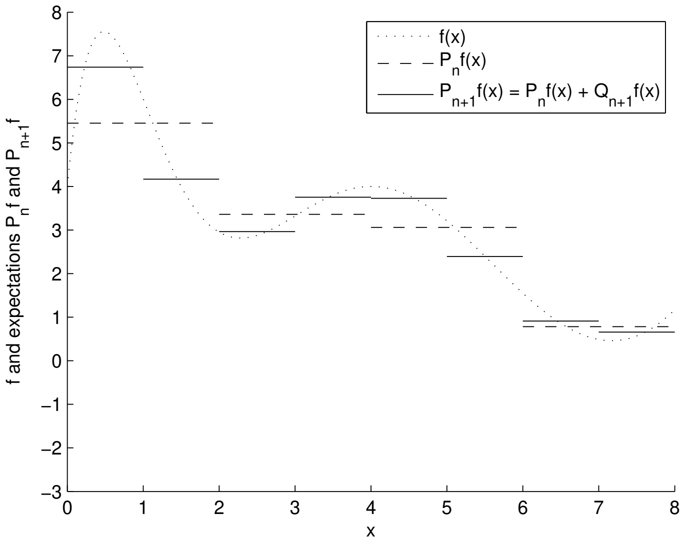


FIGURE 9.7. Graphs of f, $P_n f$, and $P_{n+1} f$. [$***n = j$]

LEMMA 9.25. *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. Recall that $m_I f$ denotes the average of the function f over the interval I:

$$m_I f = \frac{1}{|I|}\int_I f(x)\,dx.$$

By definition of the Haar functions, and noting that $|I| = 2|I_r| = 2|I_l|$, we see that

$$\langle f, h_I\rangle h_I(x) = \frac{|I|^{1/2}}{2}\left(\frac{1}{|I_r|}\int_{I_r} f - \frac{1}{|I_l|}\int_{I_l} f\right) h_I(x) = \frac{|I|^{1/2}}{2}(m_{I_r} f - m_{I_l} f) h_I(x).$$

Since $h_I(x) = |I|^{-1/2}$ if $x \in I_r$, and $h_I(x) = -|I|^{-1/2}$ if $x \in I_l$, we conclude that if $x \in I$, then

$$\langle f, h_I\rangle h_I(x) = \begin{cases} \frac{1}{2}(m_{I_r} f - m_{I_l} f), & \text{if } x \in I_r; \\ -\frac{1}{2}(m_{I_l} f - m_{I_r} f), & \text{if } x \in I_l. \end{cases} \tag{9.11}$$

On the other hand, if $x \in I \in \mathcal{D}_j$, then $P_j f(x) = m_I f$, and

$$P_{j+1} f(x) = \begin{cases} m_{I_r} f, & \text{if } x \in I_r; \\ m_{I_l} f, & \text{if } x \in I_l. \end{cases}$$

Hence

$$Q_j f(x) = \begin{cases} m_{I_r} - m_I f, & \text{if } x \in I_r; \\ m_{I_l} - m_I f, & \text{if } x \in I_l. \end{cases}$$

Here is a useful averaging property of integral averages on dyadic intervals:

$$m_I f = \frac{m_{I_l} f + m_{I_r} f}{2}. \tag{9.12}$$

In other words, the average $m_I f$ of f over I sits half-way between the averages $m_{I_l} f$ and $m_{I_r} f$ over the children. Informally, *the integral average over a parent interval is the average of the integral averages over its children.*

This averaging property implies that

$$m_{I_r} f - m_I f = \frac{m_{I_r} f - m_{I_l} f}{2} = m_I f - m_{I_l} f.$$

Hence

$$Q_j f(x) = \begin{cases} \frac{1}{2}(m_{I_r} f - m_{I_l} f), & \text{if } x \in I_r; \\ -\frac{1}{2}(m_{I_l} f - m_{I_r} f), & \text{if } x \in I_l. \end{cases}$$

Comparing to (9.11) we conclude that

$$Q_j f(x) = \langle f, h_I\rangle h_I(x) \qquad \text{for } x \in I \in \mathcal{D}_j.$$

This proves the lemma. $\square$

EXERCISE 9.26. Verify the averaging property in equation (9.12). $\Diamond$

9.3.4. Completeness of the Haar system. The Haar system of functions is complete if, 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 9.25, this condition is equivalent to the condition

$$f(x) = \lim_{M,N\to\infty} \sum_{-M \le j < N} Q_j f(x).$$

A telescoping series argument shows that

$$P_N f(x) - P_M f(x) = \sum_{M \le j < N} \big(P_{j+1} f(x) - P_j f(x)\big) = \sum_{M \le j < N} Q_j f(x). \tag{9.13}$$

Therefore, verifying completeness of the Haar system 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.

We need only prove the following theorem.

Theorem 9.27. *For $f \in L^2(\mathbb{R})$,*

$$\lim_{M\to-\infty} \|P_M f\|_2 = 0 \tag{9.14}$$

and

$$\lim_{N\to\infty} \|P_N f - f\|_2 = 0. \tag{9.15}$$

Exercise 9.28. Use Theorem 9.27 to show that, if $f \in L^2(\mathbb{R})$ is orthogonal to all Haar functions, then f must be zero in $L^2(\mathbb{R})$. (We already observed that $f(x) = 1$ is orthogonal to all Haar functions, however is not a square integrable function.) ◊

Recall that $I_j(x)$ is the unique dyadic interval in $\mathcal{D}_j$ that contains x. Figure 9.8 shows part of the "tower" of dyadic intervals $\{I_j(x)\}_{j\in\mathbb{Z}}$ shrinking to $\{x\}$ as j goes to infinity. Equation (9.15) says that given $x \in \mathbb{R}$, the averages of the function over the dyadic intervals $\{I_j(x)\}_{j\in\mathbb{Z}}$ converge to f in the L^2-sense as the intervals shrink:

$$\lim_{j\to\infty} \frac{1}{|I_j(x)|} \int_{I_j(x)} f(t)\,dt = f(x).$$

It turns out that the convergence also holds *almost everywhere*[6] (a.e.). This is the content of the celebrated *Lebesgue Differentiation Theorem* in $\mathbb{R}$.

Theorem 9.29 (Lebesgue Dominated Convergence Theorem on $\mathbb{R}$). *If $f \in L^1(\mathbb{R})$ (locally integrable suffices), then*

$$\lim_{x\in I, |I|\to 0} \frac{1}{|I|} \int_I f(y)\,dy = f(x) \quad a.e.\ x \in \mathbb{R}.$$

Here I denotes any interval, dyadic or not, that contains x.

In Chapter 4 we stated a version of this theorem for intervals $[x-h, x+h]$ centered at x, and $h \to 0$. For a proof see [**SS2**, Chapter 3].

[6]In other words, the convergence holds except on a set of measure zero.

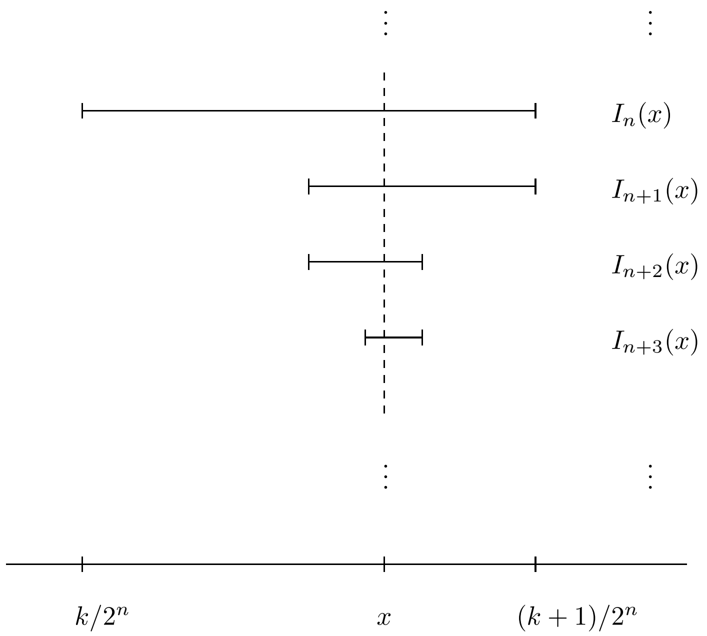


FIGURE 9.8. Part of the tower of nested dyadic intervals $\cdots \supset I_n(x) \supset I_{n+1}(x) \supset I_{n+2}(x) \supset I_{n+3}(x) \supset \cdots$ containing the point x, with $k/2^n \leq x \leq (k+1)/2^n$. Here n and k are integers.

EXERCISE 9.30. Verify the *Lebesgue Differentiation Theorem* for continuous functions, and show that for continuous functions the pointwise convergence holds everywhere. That is, for all $x \in \mathbb{R}$,

$$\lim_{[a,b]\to\{x\}} \frac{1}{b-a}\int_a^b f(t)\,dt = f(x). \tag{9.16}$$

Here the intervals $[a,b]$ need not be dyadic. $\Diamond$

Theorem 9.27 is a consequence of the following lemmata.

LEMMA 9.31. *For every function $f \in L^2(\mathbb{R})$ and every integer j, the operators P_j are* uniformly bounded *in $L^2(\mathbb{R})$. More precisely,*

$$\|P_j f\|_2 \leq \|f\|_2.$$

LEMMA 9.32. *If g is continuous and has compact support on the interval $[-K,K]$, then Theorem 9.27 holds.*

LEMMA 9.33. *The continuous functions with compact support are* dense *in $L^2(\mathbb{R})$. More precisely, 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 prove the theorem, and then the lemmata.

Proof of Theorem 9.27. By Lemma 9.33, given $\varepsilon > 0$ we can decompose $f = g + h$, where g is continuous with compact support on $[-K, K]$, and $h \in L^2(\mathbb{R})$ with $\|h\|_2 < \varepsilon/4$.

By Lemma 9.32, we can choose N large enough so that for all $j > N$,

$$\|P_{-j}g\|_2 \leq \varepsilon/2.$$

Now apply the triangle inequality, and use Lemma 9.31 to conclude that

$$\|P_{-j}f\|_2 \leq \|P_{-j}g\|_2 + \|P_{-j}h\|_2 \leq \frac{\varepsilon}{2} + \|h\|_2 \leq \varepsilon.$$

This proves equation (9.14).

Similarly, by Lemma 9.32, we can choose N large enough that for all $j > N$,

$$\|P_j g - g\|_2 \leq \varepsilon/2.$$

Now apply the triangle inequality, and use Lemma 9.31 to conclude that

$$\|P_j f - f\|_2 \leq \|P_j g - g\|_2 + \|P_j h - h\|_2 \leq \frac{\varepsilon}{2} + 2\|h\|_2 \leq \varepsilon.$$

This proves equation (9.15). □

We have twice used a very important principle from functional analysis: *If a sequence of linear operators is uniformly bounded in a Banach space, and the sequence converges to a bounded operator on a dense subset of the Banach space, then it converges in the whole space to the same operator.* This principle is called the *Uniform Boundedness Principle*, or the *Banach–Steinhaus Theorem*. In our case 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 9.31. In one case the operators converge to the zero operator, as $n \to -\infty$, and in the other case they converge to the identity operator, as $n \to \infty$.

Here is a precise statement of the Uniform Boundedness Principle, from [**Sch**, Chapter III].

Theorem 9.34 (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. That is, there exists a constant $C > 0$ such that for all $T \in W$, and all $x \in X$,

$$\|Tx\|_Y \leq C\|x\|_X.$$

In particular if the operators $\{T_n\}_{n \geq 1}$ are uniformly bounded, that is $\|T_n x\|_Y \leq C\|x\|_X$, and $\lim_{n\to\infty} \|T_n x\|_Y = 0$, for all x on a dense subset $A \subset X$, then $\lim_{n\to\infty} \|T_n x\|_Y = 0$ for all $x \in X$. Why? Consider a point $x \in X$. There are points $x_m \in A$ such that $\lim_{m\to\infty} \|x_m - x\|_X = 0$, and for those points we know that $\|T_n x_m\|_Y \to 0$ as $n \to \infty$ for all m. Now, $x = (x - x_m) + x_m$, the operators are linear, hence $T_n x = T_n(x - x_m) + T_n(x_m)$. We can estimate the norm of $T_n x$ using the triangle inequality and the uniform boundedness:

$$\|T_n x\|_Y \leq \|T_n(x - x_m)\|_Y + \|T_n x_m\|_Y \leq C\|x - x_m\|_Y + \|T_n x_m\|_Y.$$

Both terms on the right hand side can be made arbitrarily small. First, by choosing M large enough we can ensure that for a given ε,

$$\|x - x_m\|_Y < \epsilon/2C, \quad \text{for all } m > M.$$

Second, fix $m > M$ and choose N large enough that

$$\|T_n x_m\|_Y < \varepsilon/2.$$

We conclude that for all $\varepsilon > 0$, and for all $n > N$, $\|T_n x\|_Y < \varepsilon$, that is

$$\lim_{n\to\infty} \|T_n x\|_Y = 0.$$

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. We indicate in the next exercise how to deduce this.

EXERCISE 9.35. Let X be the Banach space of all real-valued continuous functions of period 2π with uniform norm, that is $X = C(\mathbb{T})$, and let $Y = \mathbb{C}$. For $f \in C(\mathbb{T})$, define $T_N(f) = S_N f(0) \in \mathbb{C}$, where $S_N f$ denotes the N^{th} partial Fourier sum of f. Recall that $S_N f = D_N * f$ where D_N denotes the periodic Dirichlet kernel (see Chapter 4). Verify that if $C_N > 0$ is such that

$$|T_N f| \leq C_N \|f\|_\infty,$$

then necessarily $C_N \geq c\|D_N\|_{L^1(\mathbb{T})}$. But we showed in Chapter 4 that $\|D_N\|_{L^1(\mathbb{T})} \sim \log N$. This implies that the operators T_N cannot be uniformly bounded, so there must exist $f \in C(\mathbb{T})$ such that

$$\sup_{N\geq 0} |S_N f(0)| = \infty.$$

Therefore for this function f the partial Fourier sums do not converge at $x = 0$. $\Diamond$

PROOF OF LEMMA 9.31. For $x \in I \in \mathcal{D}_j$,

$$|P_j f(x)|^2 = \left|\frac{1}{|I|}\int_I f(t)\,dt\right|^2 \leq \frac{1}{|I|^2}\left(\int_I 1^2\,dt\right)\left(\int_I |f(t)|^2\,dt\right) = \frac{1}{|I|}\int_I |f(t)|^2\,dt.$$

The inequality is a consequence of the Cauchy–Schwarz inequality.

Now integrate over the interval I to obtain

$$\int_I |P_j f(x)|^2 dx \leq \int_I |f(t)|^2 dt,$$

and add over all intervals in $\mathcal{D}_j$ (this is a disjoint family that covers the whole line!):

$$\int_{\mathbb{R}} |P_j f(x)|^2 dx = \sum_{I\in\mathcal{D}_n} \int_I |P_j f(x)|^2\,dx \leq \sum_{I\in\mathcal{D}_n} \int_I |f(t)|^2 dt = \int_{\mathbb{R}} |f(t)|^2\,dt.$$

The lemma is proved. $\square$

Proof of Lemma 9.32. The function g is continuous and has support on the interval $[-K,K]$. If j is large enough so that $K<2^j$ and $x\in[0,2^j)\subset\mathcal{D}_{-j}$, then

$$\begin{aligned}|P_{-j}g(x)| &= \frac{1}{2^j}\int_0^K |g(t)|dt\\ &\le \frac{1}{2^n}\left(\int_0^K 1^2\,dt\right)^{1/2}\left(\int_0^K |g(t)|^2\,dt\right)^{1/2}\\ &\le \frac{1}{2^j}\sqrt{K}\|g\|_2,\end{aligned}$$

where the last inequality is another application of the Cauchy–Schwarz inequality. The same inequality holds for $x<0$. If $|x|\ge 2^j$ then $P_{-j}g(x)=0$.

We can now estimate the L^2-norm of $P_{-j}g$:

$$\|P_{-j}g\|_2^2=\int_{-2^j}^{2^j}|P_{-j}g(x)|^2\,dx\le\frac{1}{2^{2j}}K\|g\|_2^2\int_{-2^j}^{2^j}1\,dx=2^{-j+1}K\|g\|_2^2.$$

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 $j>N$,

$$\|P_{-j}g\|_2\le\varepsilon.$$

This proves equation (9.14) for continuous functions with compact support.

We are assuming that g is continuous and 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 $|y-x|\le\delta$ for all $y\in I$, and

$$|P_jg(x)-g(x)|\le\frac{1}{|I|}\int_I|g(y)-g(x)|\,dy\le\frac{\varepsilon}{\sqrt{2^{M+1}}}.$$

Squaring and integrating over $\mathbb{R}$, we get

$$\int_{\mathbb{R}}|P_jg(x)-g(x)|^2dx=\int_{-2^M}^{2^M}|P_jg(x)-g(x)|^2dx<\frac{\varepsilon^2}{2^{M+1}}\int_{-2^M}^{2^M}1\,dx=\varepsilon^2.$$

Notice that if $|x|>2^M$, then for $n>N\ge M$, $P_jg(x)$ 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 $n>N$,

$$\|P_jg-g\|_2\le\varepsilon.$$

This proves equation (9.15) for continuous functions with compact support. □

Proof of Lemma 9.33. This lemma is an example of an approximation theorem in $L^2(\mathbb{R})$. How do we achieve it? We choose K large enough so that the tail of f has very small L^2-norm, in other words $\|f\chi_{\{x\in\mathbb{R}:|x|>K\}}\|_2\le\varepsilon/3$. Next we recall that on compact intervals, the continuous functions are dense in $L^2([-K,K])$; see Theorem 2.67. (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 $\|f\chi_{[-K,K]} - g_1\|_2 \le \varepsilon/3$. 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 $\|g_1 - g\|_2 \le \varepsilon/3$. Now let

$$h = f - g = f\chi_{\{x\in\mathbb{R}:|x|>K\}} + f\chi_{[-K,K]} - g_1 + g_1 - g.$$

By the triangle inequality,

$$\|h\|_2 \le \|f\chi_{\{x\in\mathbb{R}:|x|>K\}}\|_2 + \|f\chi_{[-K,K]} - g_1\|_2 + \|g_1 - g\|_2 \le \varepsilon. \qquad \square$$

We have shown (Lemma 9.32) that the step functions can approximate continuous functions with compact support in the L^2-norm. Lemma 9.33 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 9.36. (*Approximation by Step Functions*) Show that continuous functions with compact support can be approximated in the uniform norm by step functions. 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 on $\mathbb{R}$, and $\varepsilon > 0$, there exists $N > 0$ such that for all $j > N$, and for all $x \in \mathbb{R}$,

$$|P_j f(x) - f(x)| < \varepsilon.$$

$\Diamond$

Exercise 9.37. 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? $\Diamond$

9.4. Haar vs Fourier

We give two examples to illustrate how the Haar basis can outperform the Fourier basis when dealing with localized data. See [**Kra**, Section 7.4, p.285].

The first example is a caricature of the problem: What is the most localized "function" we could consider? The delta distribution. If we could find its Fourier series, 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.

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 9.30.

EXERCISE 9.38. Compute the Fourier transform of f_N. Check that if we view f_N as a periodic function on $[0,1)$, and if its M^{th} partial Fourier sum is

$$S_M(f_N)(x) = \sum_{|m|\leq M} \widehat{f}\,(m)e^{2\pi imx},$$

then

$$\|f_N - S_M(f_N)\|_{L^2([0,1))} \sim M^{-1/2}.$$

$\Diamond$

EXERCISE 9.39. Compute the Haar coefficients of f_N. Check that the partial Haar sum

$$P_M(f_N) = \sum_{j<M} Q_j(f_N) = \sum_{|j|<M} Q_j(f_N) + P_{-M+1}(f_N)$$

has an exponential decay rate:

$$\|f_N - P_M(f_N)\|_2 = 2^{-M/2}.$$

Hint: We have already seen this phenomenon in Exercise 9.22, for $N = 1$. $\Diamond$

The exponential decay rate exhibited in Exercise 9.39 is much better than the square root decay exhibited in Exercise 9.38. Suppose we wanted an approximation of 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 $M = 10\log_2 10 < 40$.

The localization properties of wavelets allow us to worry only about the wavelets 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 of those wavelets supported in the interval vanish, because wavelets share the zero integral property of Haar functions, namely $\int \psi = 0$. This is not the case with the trigonometric functions, whose support spreads over the whole real line.

EXERCISE 9.40. (*Truncated Cosine*) Consider the function f defined on the real line by

$$f(x) = \cos(\pi x)\,\chi_{[0,1]}(x).$$

Compute the Fourier transform of f and then reconstruct f, with the aid of a computer if necessary. With the aid of the computer, compute the Haar decomposition of f, and compare with the Fourier approximation. $\Diamond$

In Exercise 9.40, if we had considered the function $\cos(\pi x)\chi_{[-1,1]}(x)$ over a full period of the cosine function, then the Fourier series would have been the function itself, and no doubt that would have been the right way of doing the analysis. However, by chopping off the function we force other Fourier coefficients to come into play.

9.5. Unconditional bases and the martingale transform

The Haar basis happens to be a so-called *unconditional basis* for $L^p(\mathbb{R})$, $1 < p < \infty$. Roughly speaking, this means that we can approximate a function in the L^p-norm with finite linear combinations of Haar functions (basis). Furthermore, the coefficients corresponding to 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 those coefficients, that is, using some formula involving only $|\langle f, h_I\rangle|$. No information about the sign or argument of $\langle f, h_I\rangle$ is necessary. In particular, if $f \in L^p$ and

$$f = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I,$$

then the new functions defined by

$$T_\sigma f := \sum_{I\in\mathcal{D}} \sigma_I \langle f, h_I\rangle h_I, \tag{9.17}$$

where $\sigma_I = \pm 1$ (or more generally $|\sigma_I| = 1$), are also in L^p and their norms are comparable to that of f. There exist constants c, $C > 0$ such that for all choices σ of signs,

$$c\|f\|_p \leq \|T_\sigma f\|_p \leq C\|f\|_p. \tag{9.18}$$

For a given sequence $\sigma = \{\sigma_I\}$, the operator T_σ in equation (9.17) is called the *martingale transform*. It is an example of a *constant Haar multiplier*.

Let us illustrate the above paragraphs in the case $p = 2$. We already know that the Haar functions provide an orthonormal basis in $L^2(\mathbb{R})$. In particular, the Plancherel formula 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 must 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 (9.18) holds with $c = C = 1$. Therefore the martingale transform is an isometry in $L^2(\mathbb{R})$.

We need to introduce a new operator, the *dyadic square function* S^d, defined by

$$S^d(f)(x) = \left(\sum_{I\in\mathcal{D}} \frac{|\langle f, h_I\rangle|^2}{|I|}\chi_I(x)\right)^{1/2}. \tag{9.19}$$

It turns out that in the case of the Haar functions, one has the following norm equivalence:

$$c_p\|f\|_p \leq \|S^d(f)\|_p \leq C_p\|f\|_p. \tag{9.20}$$

It can be considered as an L^p substitute for Plancherel for Haar functions.

EXERCISE 9.41. (*The Dyadic Square Function is an Isometry on* $L^2(\mathbb{R})$) Verify that $\|S^d(f)\|_2 = \|f\|_2$. ◊

EXERCISE 9.42. (*The Dyadic Square Function in Terms of Difference Operators*) Show that

$$S^d(f)(x) = \left(\sum_{j\in\mathbb{Z}} |Q_j f(x)|^2\right)^{1/2},$$

where Q_j is the difference operator defined in (9.10) (Lemma 9.25 will be useful). ◊

The definition of the dyadic square function (9.19) and inequality (9.20) imply inequality (9.18), because

$$S^d(f) = S^d(T_\sigma f).$$

We can estimate the L^p-norm of f from the knowledge of the absolute values of its Haar coefficients alone.

In the case of general wavelets we also have some *averaging* and *difference operators*, P_j and Q_j, and a corresponding *square function*. The same norm equivalence (9.20) holds in $L^p(\mathbb{R})$. As it turns out, wavelet bases will provide unconditional bases for a whole zoo of function spaces (Sobolev, Hölder, etc). There are many references, in particular [**Dau92**, Chapter 9].

The trigonometric system is an orthonormal basis in $L^2([0,1])$, however it does not provide an unconditional basis in $L^p([0,1])$ for $p \neq 2$. This reflects a lack of orthogonality in the trigonometric system in $L^p([0,1])$. There is a square function that plays the same role as in wavelets, but it involves more than just the absolute value of the Fourier coefficients:

$$Sf(x) = \left(\sum_j |\Delta_j f(x)|^2 \right)^{1/2},$$

where

$$\Delta_j f(x) = \sum_{2^j \leq |n| < 2^{j+1}} \widehat{f}(j) e^{2\pi i n x}.$$

It is true that $\|f\|_p$ is comparable to $\|S(f)\|_p$ in the sense of inequality (9.20). We are allowed to change the signs of the Fourier coefficients of f on *the dyadic blocks* of frequency. If we denote by $\mathcal{T}_\delta f$ the function reconstructed with the modified coefficients, that is,

$$\mathcal{T}_\delta f(x) := \sum_{j \geq 0} \delta_j \Delta_j f(x), \qquad \delta_j = \pm 1,$$

then $S(\mathcal{T}_\delta f) = S(f)$, 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 $S(\mathcal{T}_\delta f)$. In fact, the innocent-seeming operation of changing signs inside the dyadic blocks can cause us to "jump out" of $L^p([0,1])$.

The study of square functions is known as *Littlewood–Paley Theory*; it is a widely used tool in harmonic analysis. For an introduction to dyadic Harmonic Analysis, see the lecture notes by the first author [Per].

CHAPTER 10

Zooming properties of wavelets, and applications

We highlight the *zooming* properties of the Haar system, and how they can be mathematically encoded in the so-called *multiresolution analysis* (MRA). The MRA provides a framework for the construction of most wavelets. We discuss how to implement the wavelet transform via filter banks. We present, in a very informal manner, competing attributes we would like the wavelets to have, and a by no means exhaustive catalog of wavelets. We briefly discuss wavelet packets and two-dimensional wavelets used in image processing, as well as the use of wavelet decompositions in compression and denoising of images and signals.

10.1. Multiresolution analyses (MRAs)

An *orthogonal multiresolution analysis* is a collection of closed subspaces $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ such that

(1) $\cdots \subset \mathbf{V}_{-2} \subset \mathbf{V}_{-1} \subset \mathbf{V}_0 \subset \mathbf{V}_1 \subset \mathbf{V}_2 \subset \cdots \subset L^2(\mathbb{R})$ (nested)

(2) $\bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$ (trivial intersection),

(3) $\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j$ is dense in $L^2(\mathbb{R})$ (density in $L^2(\mathbb{R})$),

(4) $f(x) \in \mathbf{V}_j$ if and only if $f(2x) \in \mathbf{V}_{j+1}$ (scaling property),

(5) $f(x) \in \mathbf{V}_0$ if and only if $f(x-k) \in \mathbf{V}_0$ for any $k \in \mathbb{Z}$ (shift invariance),

(6) There exists a *scaling function* $\varphi \in \mathbf{V}_0$ such that its integer translates, $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$, form an orthonormal basis for $\mathbf{V}_0$.

Given the scaling function φ, we denote its integer translates and dyadic dilates with subscripts j, k, as we did for the wavelet ψ:

$$\varphi_{j,k} := 2^{j/2}\varphi(2^j x - k). \tag{10.1}$$

EXERCISE 10.1. Show that $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $\mathbf{V}_j$. ◊

Given an L^2-function f, let $P_j f$ be the *orthogonal projection of f onto* $\mathbf{V}_j$:

$$P_j f := \sum_{k\in\mathbb{Z}} \langle f, \varphi_{j,k}\rangle \varphi_{j,k}. \tag{10.2}$$

The function $P_j f$ is an approximation to the original function at scale 2^{-j}. More precisely, it is the *best approximation* in the subspace $\mathbf{V}_j$ to f. See Theorem 5.33, and Section 9.3.3.

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 an equally good approximation. How do we go from the approximation $P_j f$ to the better approximation $P_{j+1} f$? 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 Q_jf to P_jf. It is clear that $P_{j+1} = P_j + Q_j$. This defines Q_j as the orthogonal projection onto a closed subspace of $L^2(\mathbb{R})$, denoted $\mathbf{W}_j$, which we call the *detail subspace* at scale 2^{-j}. The space $\mathbf{W}_j$ is the *orthogonal complement* of $\mathbf{V}_j$ in $\mathbf{V}_{j+1}$. This means that $\mathbf{V}_j \perp \mathbf{W}_j$, and if $f \in \mathbf{V}_{j+1}$, there exist unique $g \in \mathbf{V}_j$ and $h \in \mathbf{W}_j$ such that $f = g+h$. In fact $g = P_jf$ and $h = Q_jf$. We use the notation already introduced at the end of Chapter 5 to denote the direct sum of two orthogonal subspaces:

$$\mathbf{V}_{j+1} = \mathbf{V}_j \oplus \mathbf{W}_j.$$

The nested subspaces $\{V_j\}$ define an MRA. Therefore the detail subspaces give an orthogonal decomposition of $L^2(\mathbb{R})$:

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} \mathbf{W}_j. \tag{10.3}$$

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 $\mathbf{W}_0$. Since $\mathbf{W}_j$ is a dilation of $\mathbf{W}_0$, we can define

$$\psi_{j,k} = 2^{j/2}\psi(2^jx - k),$$

and the family $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ forms an orthonormal basis for $\mathbf{W}_j$. Therefore the orthogonal projection Q_j onto $\mathbf{W}_j$ is given by

$$Q_jf = \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}.$$

The collection of functions $\psi_{j,k}$ forms a wavelet basis for $L^2(\mathbb{R})$. This result is Mallat's Theorem, which we state here and prove in Section 10.3.

Theorem 10.2 (Mallat). *Given an MRA with scaling function φ, there is a wavelet $\psi \in L^2(\mathbb{R})$ such that for each j, the family $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $\mathbf{W}_j$. Hence the family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$.*

Exercise 10.3. Given an orthogonal MRA with scaling function φ, show that if $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $\mathbf{W}_0$, then $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $\mathbf{W}_j$. Show that the two-parameter family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis for $L^2(\mathbb{R})$. ◊

Consider removing this paragraph and the wavelet phase plane fig

Notice that on the Fourier side

$$(\psi_{j,0})^\wedge(\xi) = 2^{-j/2}\widehat{\psi}\,(2^{-j}\xi).$$

The Heisenberg boxes change with the scale parameter j, and remain constant along translations k. For $j = 0$ they are essentially squares of area 1. For other j the dimensions are $2^j \times 2^{-j}$. The wavelet transform divides the phase plane differently from either the Fourier or Gabor bases. The wavelet phase plane is given in Figure 10.1.

Example 10.4. (*The Haar MRA*) The characteristic function $\varphi(t) = \chi_{[0,1]}(t)$ is the scaling function of an orthogonal MRA, namely the *Haar MRA*. The subspace $\mathbf{V}_j$ corresponds to step functions with steps on the intervals $[k2^{-j}, (k+1)2^{-j})$, and they have all the properties listed. The Haar function is the wavelet ψ in Mallat's theorem. We return to this example in Section 10.2.

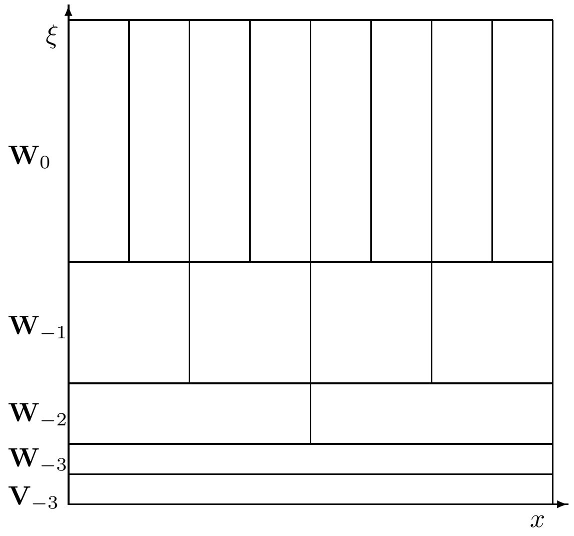


FIGURE 10.1. The wavelet phase plane.

In Section 9.3.3 we defined the expectation and difference operator for the Haar basis. These operators coincide with the ones associated with the *Haar MRA*. $\Diamond$

Are there other MRAs? Yes, there are.

EXAMPLE 10.5. (*The Shannon MRA*) We met the Shannon wavelet in Section 9.2. The scaling function is defined on the Fourier side by

$$\widehat{\varphi}(\xi) = \chi_{[-1/2,1/2)}(\xi),$$

The subspaces $\mathbf{V}_j$ consist of those functions that are band limited to the interval $[-2^{j-1}, 2^{j-1})$; that is, those functions f such that the support of their Fourier transform is contained on the interval $[-2^{j-1}, 2^{j-1})$. The subspaces $\mathbf{W}_j$ consist of functions that are band limited to the double-paned window $[-2^j, -2^{j-1}) \cup [2^{j-1}, 2^j)$. $\Diamond$

EXERCISE 10.6. Verify that the subspaces V_j defined in Example 10.5 do generate an MRA. Verify that the subspace $\mathbf{W}_j$ is the orthogonal complement of $\mathbf{V}_j$ in $\mathbf{V}_{j+1}$. $\Diamond$

While there are wavelets that do not come from an MRA, these are rare. If the wavelet has compact support then it does come from an MRA. 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 all these issues and more, consult the book by Hernández and Weiss [**HW**, Chapter 2] and the book by Wojtaszczyk [**Woj**, Chapter 2].

10.2. The Haar multiresolution analysis

Before considering how to construct wavelets from scaling functions associated to an MRA, we revisit the simple example that predates the main development of wavelets, namely the Haar multiresolution analysis.

The scaling function for the Haar MRA is the characteristic function of the unit interval,

$$\varphi(x) = \begin{cases} 1, & \text{for } 0 \leq x < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

The closure $\mathbf{V}_0$ of the linear span of the integer translates of the Haar scaling function φ,

$$\mathbf{V}_0 := \overline{\operatorname{span}(\{\varphi(x-k)\}_{k\in\mathbb{Z}})},$$

is the subspace consisting of piecewise constant functions with jumps only at the integers. The subspace

$$\mathbf{V}_j := \overline{\operatorname{span}(\{\varphi_{j,k}\}_{k\in\mathbb{Z}})}$$

is the subspace consisting of piecewise constant functions with jumps only at the integer multiples of 2^{-j}.

The orthogonal projection $P_j f$ onto $\mathbf{V}_j$ 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 9.3.3.) We showed in Lemma 9.25 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, defined by

$$\psi(x) = h(x) = \begin{cases} -1, & \text{for } 0 \leq x < 1/2; \\ 1, & \text{for } 1/2 \leq x < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Hence Q_j is the orthogonal projection onto $\mathbf{W}_j$, the closure of the linear span of $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$. The subspace

$$\mathbf{W}_j = \overline{\operatorname{span}(\{\psi_{j,k}\}_{k\in\mathbb{Z}})},$$

consists of piecewise constant functions with jumps only at integer multiples of $2^{-(j+1)}$, and average 0 between integers integer multiples of 2^{-j}.

We can view the averages $P_j f$ at resolution j as successive approximations to the original signal f. These approximations are the orthogonal projections $P_j f$ onto the approximation spaces $\mathbf{V}_j$. 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 $\mathbf{W}_j$. 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 $f \in L^2(\mathbb{R})$ in L^2-norm (this is the content of equation (9.15) in Theorem 9.27), and a.e. (Lebesgue Differentiation Theorem). Clearly the subspaces are nested, that is, $\mathbf{V}_j \subset \mathbf{V}_{j+1}$, and their intersection is the trivial subspace containing just the zero function (this is equation (9.14) in Theorem 9.27). Lo and behold, we have shown that the Haar scaling function generates an MRA.

We give an example of how to decompose a function into its projections onto the Haar subspaces. We have borrowed this example from [**MP**].

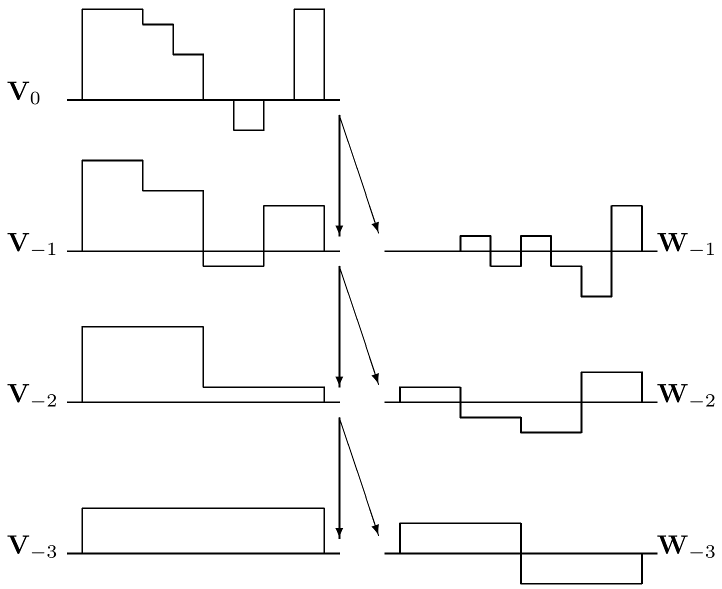


FIGURE 10.2. A wavelet decomposition: $\mathbf{V}_0 = \mathbf{V}_{-3} \oplus \mathbf{W}_{-3} \oplus \mathbf{W}_{-2} \oplus \mathbf{W}_{-1}$.

Consider changing the numbers in the vector and hence in the projections and pictures.

In practice, we select a coarsest scale $\mathbf{V}_{-n}$ and a finest scale $\mathbf{V}_0$, truncate the chain to

$$\mathbf{V}_{-n} \subset \cdots \subset \mathbf{V}_{-2} \subset \mathbf{V}_{-1} \subset \mathbf{V}_0,$$

and obtain

$$\mathbf{V}_0 = \mathbf{V}_{-n} \oplus \mathbf{W}_{-n} \oplus \mathbf{W}_{-n+1} \oplus \mathbf{W}_{-n+2} \oplus \cdots \oplus \mathbf{W}_{-1}. \tag{10.4}$$

We will go through the decomposition process in the text using vectors. It is also enlightening to look at the graphical version in Figure 10.2.

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 8 intervals of length 1, so that our function is supported on the interval $[0, 8]$. For our example, we choose the vector

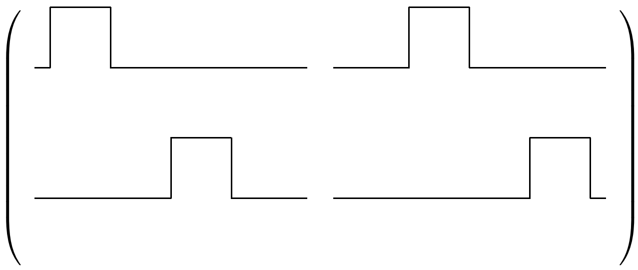
$$v_0 = [6, 6, 5, 3, 0, -2, 0, 6]$$

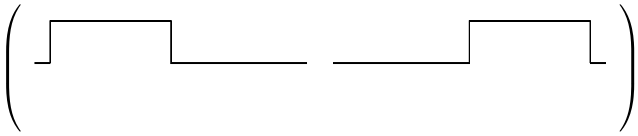
to represent our function in $\mathbf{V}_0$. To construct the projection onto $\mathbf{V}_{-1}$ we average pairs of values, obtaining

$$v_{-1} = [6, 6, 4, 4, -1, -1, 3, 3].$$

The difference is in $\mathbf{W}_{-1}$, so we have

$$w_{-1} = [0, 0, 1, -1, 1, -1, -3, 3].$$

$\mathbf{V}_{-1} = \text{span}$ 

$\mathbf{V}_{-2} = \text{span}$ 

$\mathbf{V}_{-3} = \text{span}$ 

FIGURE 10.3. The scaling function subspaces used in Figure 10.2.

By repeating this process, we obtain

$$\begin{aligned} v_{-2} &= [5, 5, 5, 5, 1, 1, 1, 1], \\ w_{-2} &= [1, 1, -1, -1, -2, -2, 2, 2], \\ v_{-3} &= [3, 3, 3, 3, 3, 3, 3, 3], \quad \text{and} \\ w_{-3} &= [2, 2, 2, 2, -2, -2, -2, -2]. \end{aligned}$$

For the example in Figure 10.2, the scaling function subspaces are shown in Figure 10.3 and the wavelet subspaces are shown in Figure 10.4. To compute the coefficients of the expansion (10.2), we must compute the inner product $\langle f, \varphi_{j,k} \rangle$ for the function (10.1). In terms of our vectors, we have for example

$$\langle f, \varphi_{0,3} \rangle = \langle [6, 6, 5, 3, 0, -2, 0, 6], [0, 0, 0, 1, 0, 0, 0, 0] \rangle = 3$$

and

$$\langle f, \varphi_{1,1} \rangle = \langle [6, 6, 5, 3, 0, -2, 0, 6], [0, 0, 1/\sqrt{2}, 1/\sqrt{2}, 0, 0, 0, 0] \rangle = 8/\sqrt{2}.$$

The scaling function φ satisfies the *two-scale recurrence equation*

$$\varphi(t) = \varphi(2t) + \varphi(2t - 1). \tag{10.5}$$

Therefore $\varphi_{j,k} = (\varphi_{j-1,2k} + \varphi_{j-1,2k+1})/\sqrt{2}$, and so

$$\langle f, \varphi_{j,k} \rangle = \frac{1}{\sqrt{2}}(\langle f, \varphi_{j-1,2k} \rangle + \langle f, \varphi_{j-1,2k+1} \rangle).$$

Thus we can also compute

$$\langle f, \varphi_{1,1} \rangle = \frac{1}{\sqrt{2}}(5 + 3).$$

The coefficients $\langle f, \varphi_{j,k} \rangle$ for fixed j are called the *averages* of f at scale j, and denoted $a_{j,k}$.

$\mathbf{W}_{-1} = \operatorname{span}$

$\mathbf{W}_{-2} = \operatorname{span}$

$\mathbf{W}_{-3} = \operatorname{span}$

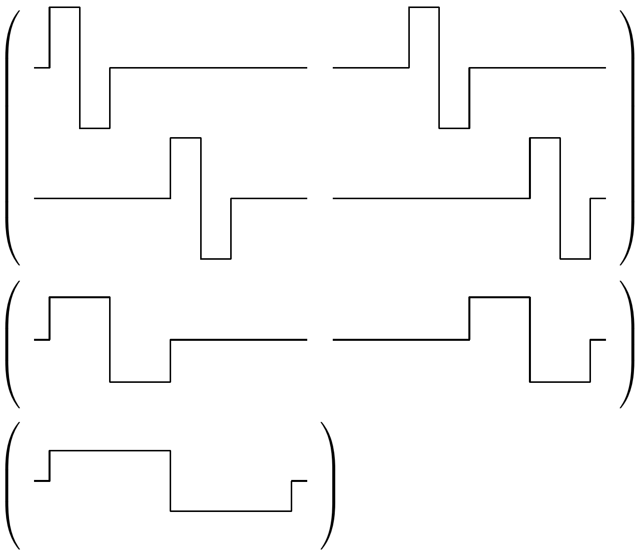


FIGURE 10.4. The wavelet subspaces used in Figure 10.2.

Similarly, the wavelet satisfies the *two-scale difference equation*,

$$\psi(t) = \varphi(2t) - \varphi(2t-1), \tag{10.6}$$

and thus we can recursively compute

$$\langle f, \psi_{j,k}\rangle = \frac{1}{\sqrt{2}}(\langle f, \varphi_{j-1,2k}\rangle - \langle f, \varphi_{j-1,2k+1}\rangle).$$

The coefficients $\langle f, \psi_{j,k}\rangle$ for fixed j are called the *differences* or *details* of f at scale j, and denoted $d_{j,k}$. Evaluating the whole set of Haar coefficients $d_{j,k}$ and averages $a_{j,k}$ requires $2(N-1)$ additions and $2N$ multiplications. The discrete wavelet transform can be performed using a similar *cascade algorithm* with complexity N, where N is the number of data points. 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. This is the same algorithm as the fast Haar transform we discussed in Chapter 6 in the language of matrices.

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.

First let us find necessary and sufficient conditions on the Fourier side that guarantee that the family of integer translates of an square-integrable function forms an orthonormal family.

LEMMA 10.7. *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 for almost every $\xi \in \mathbb{R}$*

$$\sum_{n\in\mathbb{Z}} |\widehat{f}(\xi+n)|^2 = 1.$$

PROOF. First note that by a change of variable,

$$\langle \tau_k f, \tau_m f\rangle = \langle \tau_{k-m} f, f\rangle.$$

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 $e^{-2\pi ik\xi}$ has period one. Therefore for all $k \in \mathbb{Z}$

$$\begin{aligned}
\delta_k &= \langle \widehat{\tau_k f}, \widehat{f}\rangle \\
&= \int_{\mathbb{R}} e^{-2\pi ik\xi} |\widehat{f}(\xi)|^2 \, d\xi \\
&= \sum_{n\in\mathbb{Z}} \int_n^{n+1} e^{-2\pi ik\xi} |\widehat{f}(\xi)|^2 \, d\xi \\
&= \int_0^1 e^{-2\pi ik\eta} \sum_{n\in\mathbb{Z}} |\widehat{f}(\eta+n)|^2 \, d\eta.
\end{aligned}$$

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. □

Let φ be the scaling function of an orthogonal MRA. Then its integer translates form an orthonormal family. Therefore, by Lemma 10.7, for almost every $\xi \in \mathbb{R}$

$$\sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi+n)|^2 = 1. \tag{10.7}$$

Also, $\varphi \in \mathbf{V}_0 \subset \mathbf{V}_1$, and the functions $\varphi_{1,k}(x) = \sqrt{2}\varphi(2x-k)$, for $k \in \mathbb{Z}$, form an orthonormal basis for $\mathbf{V}_1$. This means that the following *scaling equation* holds, for some coefficients $\{h_k\}$ such that $\sum_k |h_k|^2 < \infty$:

$$\varphi(x) = \sum_{k\in\mathbb{Z}} h_k \varphi_{1,k}(x) = \sqrt{2} \sum_{k\in\mathbb{Z}} h_k \varphi(2x-k). \tag{10.8}$$

On the Fourier side, the scaling equation reads

$$\widehat{\varphi}(\xi) = H(\xi/2)\widehat{\varphi}(\xi/2), \tag{10.9}$$

where $H(\xi) = (1/\sqrt{2}) \sum h_k e^{-2\pi ik\xi}$ is a function of period 1, called the *refinement mask*.

For convenience we assume that H is a trigonometric polynomial, so that all but finitely many of the coefficients $\{h_k\}$ vanish. There are multiresolution analyses whose refinement masks are trigonometric polynomials, for example the Haar MRA. In fact, for applications, these are the most useful, and they correspond to MRAs with compactly supported scaling functions $\boxed{\text{ref: Daubechies?}}$.

EXERCISE 10.8. Check that the scaling equation on the Fourier side is given by equation (10.9). ◊

We can now deduce a necessary property that the refinement mask H must satisfy, namely the so-called *quadrature mirror filter* (QMF) property.

LEMMA 10.9. *Given an orthogonal MRA with scaling function φ, and corresponding refinement mask H that is a trigonometric polynomial, then*

$$|H(\xi)|^2 + |H(\xi + 1/2)|^2 = 1.$$

PROOF. Insert equation (10.9) into equation (10.7), obtaining

$$\begin{aligned} 1 &= \sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi + n)|^2 \\ &= \sum_{n\in\mathbb{Z}} |H((\xi + n)/2)|^2 |\widehat{\varphi}((\xi + n)/2)|^2. \end{aligned}$$

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 (twice) equation (10.7), which holds for almost every point ξ:

$$\begin{aligned} 1 &= |H(\xi/2)|^2 \sum_{k\in\mathbb{Z}} |\widehat{\varphi}(\xi/2 + k)|^2 + |H(\xi/2 + 1/2)|^2 \sum_{k\in\mathbb{Z}} |\widehat{\varphi}((\xi + 1)/2 + k)|^2 \\ &= |H(\xi/2)|^2 + |H(\xi/2 + 1/2)|^2. \end{aligned}$$

Equality holds almost everywhere. Since H is a trigonometric polynomial, H is continuous and so equality must hold everywhere. □

Note that if $f \in \mathbf{V}_1$, 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

$$\widehat{f}(\xi) = m_f(\xi/2)\widehat{\varphi}(\xi/2). \tag{10.10}$$

In this notation, the refinement mask $H = m_\varphi$.

We now describe the functions in $\mathbf{W}_0$, the orthogonal complement of $\mathbf{V}_0$ in $\mathbf{V}_1$, where all the subspaces correspond to an orthogonal MRA with scaling function φ and refinement mask a trigonometric polynomial H.

LEMMA 10.10. *A function $f \in \mathbf{W}_0$ 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, $f \in \mathbf{W}_0$ if and only if $f \in \mathbf{V}_1$ and $f \perp \mathbf{V}_0$. The fact that $f \in \mathbf{V}_1$ allows us 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 10.7 shows that

$$\begin{aligned} 0 &= \langle \widehat{f}, \widehat{\varphi_{0,k}} \rangle \\ &= \int_{\mathbb{R}} \widehat{f}(\xi) e^{2\pi i k \xi} \overline{\widehat{\varphi}(\xi)}\, d\xi \\ &= \int_{\mathbb{R}} e^{2\pi i k \xi} m_f(\xi/2) \widehat{\varphi}(\xi/2) \overline{H(\xi/2)\widehat{\varphi}(\xi/2)}\, d\xi \\ &= \int_{\mathbb{R}} e^{2\pi i k \xi} m_f(\xi/2) \overline{H(\xi/2)} |\widehat{\varphi}(\xi/2)|^2\, d\xi. \end{aligned}$$

In the third line we have used equations (10.10) and (10.9).

At this point we want to use the same trick we used in Lemma 10.7: 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

$$\begin{aligned} 0 &= \sum_{n\in\mathbb{Z}} \int_n^{n+1} e^{2\pi i k \xi} m_f(\xi/2) \overline{H(\xi/2)} |\widehat{\varphi}(\xi/2)|^2\, d\xi \\ &= \int_0^1 e^{2\pi i k \xi} \sum_{n\in\mathbb{Z}} m_f((\xi+n)/2) \overline{H((\xi+n)/2)} |\widehat{\varphi}((\xi+n)/2)|^2\, d\xi. \end{aligned}$$

We would also like to take advantage of the periodicity of the functions m_f and H. These are function of period one, but we are adding half an integer. If we separate the last sum, as we did in Lemma 10.9, into the sums over the odd and even integers, we will win. In fact we obtain that for all $k \in \mathbb{Z}$

$$\begin{aligned} 0 &= \int_0^1 e^{2\pi i k \xi} [m_f(\xi/2) \overline{H(\xi/2)} \sum_{m\in\mathbb{Z}} |\widehat{\varphi}(\xi/2+m)|^2 \\ &\quad + m_f(\xi/2+1/2) \overline{H(\xi/2+1/2)} \sum_{m\in\mathbb{Z}} |\widehat{\varphi}((\xi+1)/2+m)|^2]\, d\xi \\ &= \int_0^1 e^{2\pi i k \xi} \left[m_f(\xi/2)\overline{H(\xi/2)} + m_f(\xi/2+1/2)\overline{H(\xi/2+1/2)} \right] d\xi, \end{aligned}$$

where the last equality is a consequence of applying equation (10.7) twice. This time the 1-periodic function (see Exercise 10.11)

$$F(\xi) = m_f(\xi/2)\overline{H(\xi/2)} + m_f(\xi/2+1/2)\overline{H(\xi/2+1/2)},$$

has been shown to have Fourier coefficients identically equal to zero. Hence it must be the zero function almost everywhere:

$$m_f(\xi)\overline{H(\xi)} + m_f(\xi+1/2)\overline{H(\xi+1/2)} = 0 \quad \text{a.e.} \tag{10.11}$$

This equation says that for a 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

$$\vec{w} = \big(H(\xi), H(\xi+1/2)\big).$$

Notice that the QMF property of H (see Lemma 10.9) 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 a two-dimensional space. Therefore the orthogonal complement of the one-dimensional subspace generated by the vector $\vec{w}$ is one-dimensional. It suffices to locate one non-zero vector $\vec{u} \in C^2$ that is orthogonal to $\vec{w}$ to completely characterize all the vectors $\mathbf{v}$ 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 $\mathbf{w}$. 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 10.12). The conclusion is that $f \in \mathbf{W}_0$ 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 10.11. 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 10.12. 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. ◊

We are now ready to present the wavelet ψ associated to the MRA with scaling function φ and refinement mask H, where H is a trigonometric polynomial. The function ψ we are looking for is in $\mathbf{W}_0$. Therefore on the Fourier side it must satisfy the equation

$$\widehat{\psi}(\xi) = m_\psi(\xi/2)\widehat{\varphi}(\xi/2),$$

where

$$m_\psi(\xi) = e^{2\pi i\xi}\sigma(\xi)\overline{H(\xi+1/2)}, \tag{10.12}$$

where $\sigma(\xi)$ 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 10.7 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 (10.12) 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.

Proof of Mallat's Theorem (Theorem 10.2). Choose the simplest possible function of period 1/2 that has absolute value 1 all the time, namely $\sigma(\xi) \equiv 1$. 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 10.10, $\psi \in \mathbf{W}_0$. Its integer translates are also in $\mathbf{W}_0$, because

$$\widehat{\psi_{0,k}}(\xi) = e^{-2\pi ik\xi} G(\xi/2)\widehat{\varphi}(\xi/2).$$

Furthermore G satisfies a QMF property, and so the family of integer translates of ψ is an orthonormal family in $\mathbf{W}_0$.

It remains to show that this family spans $\mathbf{W}_0$. By Lemma 10.10, 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 of 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

$$f = \sum_{k\in\mathbb{Z}} a_k \psi_{0,k}.$$

That is, f belongs to the span of the integer translates of ψ. The integer translates of ψ form an orthonormal basis of $\mathbf{W}_0$. By scale invariance, the functions $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis of $\mathbf{W}_j$. Thus the family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis of $L^2(\mathbb{R})$, as required. □

EXERCISE 10.13. Verify that if $\tau_k\psi$ is an orthonormal system, and $\psi \in \mathbf{W}_0$, then m_ψ satisfies the QMF property. ◊

10.4. Mallat's algorithm revisited

In this section we give a glimpse of how we would implement the wavelet transform once an MRA is at our disposal, in a similar way to the implementation for the Haar functions.

It turns out that all one needs for computations are the so-called *filter coefficients*, and not the scaling and wavelet functions. These filter coefficients are finite sequences of numbers if and only if the scaling function is *compactly supported*, as it is the case of the Haar scaling function.

As noted above, given an orthogonal MRA, the scaling function φ satisfies the following *scaling equation*, for some set of filter coefficients $h = \{h_k\}$ such that $\sum_k |h_k|^2 < \infty$:

$$\varphi(t) = \sum_{k\in\mathbb{Z}} h_k\varphi_{1,k}(t) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k\varphi(2t-k). \tag{10.13}$$

In general this sum is not finite, but whenever it is, both the scaling function φ and the wavelet ψ are *compactly supported.* In the case of the Haar MRA, we have $h_0 = h_1 = 1/\sqrt{2}$, and all other coefficients vanish. The sequence $h = \{h_k\}$ is the so-called *low-pass filter.* We assume that the low-pass filter has finite length L. (It turns out that such a filter always has even length, $L = 2M$ say). The *refinement mask* is given by $H(\xi) = (1/\sqrt{2})\sum h_k e^{-2\pi ik\xi}$, which can be viewed as a 1-periodic function, of the frequency variable ξ, whose Fourier coefficients are $\widehat{H}\,(n) = h_{-n}/\sqrt{2}$.

Not all wavelets have compact support. However, for applications it is a most desirable property, since compactly supported wavelets correspond to FIR filters.

The existence of a solution of the scaling equation can be expressed in the language of fixed-point theory. Given a low-pass filter H, define a transformation T by $T\varphi(t) := \sqrt{2}\sum_k h_k\varphi(2t-k)$. Does T have a fixed point? If yes, then the fixed point is a solution to the scaling equation $\boxed{\text{ref?}}$. However, we do not pursue this argument here; instead we use Fourier analysis.

On Fourier side the scaling equation becomes

$$\widehat{\varphi}(\xi) = H(\xi/2)\widehat{\varphi}(\xi/2). \tag{10.14}$$

We can iterate this formula to obtain

$$\widehat{\varphi}(\xi) = \left(\Pi_{j=0}^{N} H(\xi/2^j)\right)\widehat{\varphi}(\xi/2^N).$$

If $\widehat{\varphi}$ is continuous at $\xi = 0$, $\widehat{\varphi}(0) \neq 0$, and the infinite product converges, then there is a solution to the scaling equation. It turns out that to obtain orthonormality of the set $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ we must have $|\widehat{\varphi}(0)| = 1$, 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.

The conditions on the filter H that guarantee the existence of a solution φ to the scaling equation are now well understood (consult [**HW**] for more details). For example, the fact that the infinite product $\Pi_{j=0}^{\infty} H\left(\frac{\xi}{2^j}\right)$ must converge for each ξ, in particular for $\xi = 0$, forces $H(0) = 1$, or equivalently $\sum_{k=0}^{L-1} h_k = \sqrt{2}$. In Lemma 10.9 we showed that the orthonormality of the integer shifts of the scaling function implies that

$$|H(\xi)|^2 + |H(\xi + 1/2)|^2 = 1. \tag{10.15}$$

Equation (10.15) is known in the engineering community as a *quadrature mirror filter* (QMF) condition, necessary for exact reconstruction for a pair of filters. The QMF condition together with $H(0) = 1$ implies that $H(1/2) = 0$, which explains the name *low-pass filter*: low frequencies near $\xi = 0$ are kept, while high frequencies near $\xi = 1/2$ are removed (filtered out).

The wavelet ψ we are seeking is an element of $\mathbf{W}_0 \subset \mathbf{V}_1$. Therefore it is also a superposition of the basis elements $\{\varphi_{1,k}\}_{k\in\mathbb{Z}}$ of $\mathbf{V}_1$. So there are coefficients $\{g_k\}$ such that $\sum_{k\in\mathbb{Z}} |g_k|^2 < \infty$, and

$$\psi(t) = \sum_{k\in\mathbb{Z}} g_k\varphi_{1,k}(t). \tag{10.16}$$

Define the *high-pass filter* $g = \{g_k\}$ by

$$g_k = (-1)^{k-1}\overline{h_{1-k}}, \qquad G(\xi) = \frac{1}{\sqrt{2}}\sum_{k=0}^{L-1} g_k e^{2\pi i k\xi}. \tag{10.17}$$

With this choice of filter, the function ψ given by equation (10.16) is Mallat's wavelet, which we constructed in Section 10.3. 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 10.14.

If the high-pass filter G is itself a QMF, in other words if

$$|G(\xi)|^2 + |G(\xi + 1/2)|^2 = 1, \tag{10.18}$$

then it can be verified that for each scale j, the wavelets $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis for $\mathbf{W}_j$. Furthermore, the orthogonality between $\mathbf{W}_j$ and $\mathbf{V}_j$

implies that for all ξ,

$$H(\xi)\overline{G(\xi)} + H(\xi+1/2)\overline{G(\xi+1/2)} = 0. \tag{10.19}$$

This relation together with the knowledge that $H(0) = 1$, and $H(\pm 1/2) = 0$, implies that $G(0) = 0$ (equivalently $\sum g_k = 0$) and $G(\pm 1/2) = 1$, which 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 formula (10.17).

EXERCISE 10.14. Check that with this choice, $G(\xi) = e^{2\pi i\xi}\overline{H(\xi+1/2)}$, and that if H satisfies a QMF condition, so does G. Check that equation (10.19) 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 $V_j \perp W_j$. $\diamondsuit$

$\boxed{\text{Construct box diagram.}}$

Mallat's theorem provide a mathematical algorithm for constructing the wavelet from the MRA and the scaling function via the filter coefficients. After discussing some examples we will see in the next section that Mallat's theorem also provides a numerical algorithm that can be implemented very successfully, namely the *cascade algorithm* or *fast wavelet transform*. We represent the cascade algorithm schematically as follows.

$$\boxed{\begin{array}{c}\varphi\\ \text{scaling MRA}\end{array}} \to \boxed{\begin{array}{c}H\\ \text{low-pass filter}\end{array}} \to \boxed{\begin{array}{c}G\\ \text{high-pass filter}\end{array}} \to \boxed{\begin{array}{c}\psi\\ \text{Mallat's wavelet}\end{array}}$$

EXAMPLE 10.15. (*Haar Revisited*) The characteristic function of the unit interval $\chi_{[0,1]}$ generates an orthogonal MRA, namely the Haar MRA. The non-zero low-pass filter coefficients are $h_0 = h_1 = 1/\sqrt{2}$; hence the non-zero 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 refinement masks are

$$H(\xi) = \frac{1+e^{2\pi i\xi}}{2}, \qquad G(z\xi) = \frac{e^{-2\pi i\xi}-1}{2}.$$

Compute the infinite product $\Pi_{j=1}^{\infty} H(\xi/2^j)$ directly in this example, and compare it with $\widehat{\varphi}(\xi)$ (they should coincide). $\diamondsuit$

EXAMPLE 10.16. (*Shannon Revisited*) This time $\widehat{\varphi}(\xi) = \chi_{[-1/2,1/2)}(\xi)$ generates an orthogonal MRA. It follows from Exercise 10.8 that $H(\xi) = \chi_{[-1/2,1/4)\cup[1/4,1/2)}(\xi)$. Hence by Exercise 10.14,

$$G(\xi) = e^{2\pi i\xi}H(\xi+1/2) = e^{2\pi i\xi}\chi_{[-1/4,1/4)}(\xi)$$

(recall that we are viewing $H(\xi)$ and $G(\xi)$ as periodic functions on the unit interval), and

$$\widehat{\psi}(\xi) = e^{\pi i\xi}\chi_{\{1/2<|\xi|\leq 1\}}(\xi).$$

$\diamondsuit$

EXAMPLE 10.17. (*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. They are denoted in MATLAB by dbN. The wavelet $db1$ is the Haar wavelet. The coefficients corresponding to $db2$ are

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \qquad h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \qquad h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, \qquad h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}.$$

◊

EXERCISE 10.18. Check that the $db2$ filter is a QMF, and that $h_0+h_1+h_2+h_3 = \sqrt{2}$. ◊

It turns out that for finite filters H, the conditions (10.15) and $\sum_k h_k = \sqrt{2}$ are sufficient to guarantee the existence of a solution φ to the scaling equation. For infinite filters an extra decay assumption is necessary. However, it is not sufficient to guarantee the orthonormality $\boxed{\text{check!}}$ of the integer shifts of φ. But, if for example $\inf_{|\xi|\leq 1/4} |H(\xi)| > 0$ is also true, then $\{\varphi_{0,k}\}$ is an orthonormal set in $L^2(\mathbb{R})$. For more details, see for example [**Fra**, Chapter 5].

10.5. Cascade algorithm and filter banks

A *cascade algorithm*, similar to the one described for the Haar basis, can be implemented to provide a fast wavelet transform. Given the approximation coefficients $\{a_{J,k}\}_{k=0,1,\dots,N-1}$, $N = 2^J$ "scaled samples" of the function f defined on the interval $[0,1]$, namely

$$a_{J,k} := \langle f, \varphi_{J,k}\rangle.$$

Then the coarser approximation and detail coefficients $a_{j,k} := \langle f, \varphi_{j,k}\rangle$ and $d_{j,k} := \langle f, \psi_{j,k}\rangle$ for scales $j < J$ can be calculated in order LN operations, where L is the length of the filter from the coefficients in the finest approximation scale $a_{J,k} := \langle f, \varphi_{J,k}\rangle$. Let a_j denote the sequence $\{a_{j,k}\}$ and d_j the sequence $\{d_{j,k}\}$.

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 $\{\varphi_{1,m}\}$:

$$\begin{aligned}
\varphi_{0,\ell}(x) &= \varphi(x-\ell) = \sum_{k\in\mathbb{Z}} h_k \varphi_{1,k}(x-\ell) \\
&= \sum_{k\in\mathbb{Z}} h_k \sqrt{2}\varphi(2x-(2\ell+k)) \\
&= \sum_{m\in\mathbb{Z}} h_{m-2\ell}\varphi_{1,m} \\
&= \sum_{m\in\mathbb{Z}} \tilde{h}_{2\ell-m}\varphi_{1,m},
\end{aligned}$$

where $\tilde{h}_k := \overline{h_{-k}}$.

We can now compute $a_{0,\ell}$ in terms of a_1:

$$\begin{aligned} a_{0,\ell} &= \langle f, \varphi_{0,\ell}\rangle \\ &= \sum_{m\in\mathbb{Z}} \tilde{h}_{2\ell-m}\langle f, \varphi_{1,m}\rangle \\ &= \sum_{m\in\mathbb{Z}} \tilde{h}_{2\ell-m} a_{1,m} \\ &= \tilde{h} * a_1(2m), \end{aligned}$$

where $\widetilde{h} = \{\widetilde{h}_k\}$ is the conjugate flip of the low-pass filter h.

Similarly, for the detail coefficients $d_{0,\ell}$, the two-scale equation connects $\psi_{0,\ell}$ to $\{\varphi_{1,m}\}$. The only difference is that the filter coefficients are $\{g_k\}$ instead of $\{h_k\}$, so we obtain

$$\psi_{0,\ell}(x) = \sum_{m\in\mathbb{Z}} \tilde{g}_{2\ell-m}\varphi_{1,m},$$

and consequently,

$$d_{0,\ell} = \sum_{m\in\mathbb{Z}} \tilde{g}_{2\ell-m} a_{1,m} = \tilde{g} * a_1(2m),$$

where $\widetilde{g} = \{\widetilde{g}_k\}$ is the conjugate flip of the high-pass filter g.

Exercise 10.19. Verify that for all $j \in \mathbb{Z}$,

$$\begin{aligned} \varphi_{j,k} &= \sum_{n\in\mathbb{Z}} h_{m-2\ell}\varphi_{j+1,n} \\ \psi_{j,k} &= \sum_{n\in\mathbb{Z}} g_{m-2\ell}\varphi_{j+1,n} \end{aligned}$$

$\Diamond$

As a consequence of Exercise 10.19: we obtain for all $j \in \mathbb{Z}$

$$\begin{aligned} a_{j,k} &= \sum_{n\in\mathbb{Z}} \tilde{h}_{2k-n}\, a_{j+1,n} = \widetilde{h} * a_{j+1}(2k), \\ d_{j,k} &= \sum_{n\in\mathbb{Z}} \tilde{g}_{2k-n}\, a_{j+1,n} = \widetilde{g} * a_{j+1}(2k), \end{aligned}$$

We can compute the approximation and detail coefficients at a rougher scale (j) by convolving (circular convolution) the approximation coefficients at the finer scale ($j+1$) with the low and high-pass filters $\widetilde{h}$ and $\widetilde{g}$, and *down-sampling* by a factor of two. More precisely the *down-sampling operator* takes an N-vector and maps it into a vector half as long by discarding the odd entries,

$$Ds(n) = s(2n).$$

The down-sampling 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 schematically as follows.

$$\begin{array}{ccl} & & \boxed{*\widetilde{h}} \to \boxed{\downarrow 2} \to a_j \\ & \nearrow & \\ a_{j+1} \to & & \\ & \searrow & \\ & & \boxed{*\widetilde{g}} \to \boxed{\downarrow 2} \to d_j \end{array}$$

Another useful operation is *up-sampling*, which is the right inverse of down-sampling. The *up-sampling operator* takes an N-vector and maps it to a vector twice as long, 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 up-sampling operator is denoted by the symbol $\uparrow 2$.

EXERCISE 10.20. Compute the Fourier transform for the up-sampling and down-sampling operators in finite-dimensional space. $\Diamond$

EXERCISE 10.21. Verify that given a vector $s \in \mathbb{C}^N$ then $DUs = s$, but UDs is not always s. For which vectors s is $UDs = s$? $\Diamond$

The reconstruction of the "samples" at level $j+1$ from the samples and details at the previous level j is also an order N algorithm.

We will carefully analyze how to get from the coarse approximation and detail coefficients a_j and d_j to the approximation coefficients in 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

$$\sum_{m\in\mathbb{Z}} a_{j+1,m}\varphi_{j+1,m} = \sum_{\ell\in\mathbb{Z}} a_{j,\ell}\varphi_{j,\ell} + \sum_{\ell\in\mathbb{Z}} d_{j,\ell}\psi_{j,\ell}. \tag{10.20}$$

We have formulas that express $\varphi_{j,\ell}$ and $\psi_{j,\ell}$ in terms of $\{\varphi_{j+1,m}\}_{m\in\mathbb{Z}}$; see Exercise 10.19. Inserting those formulas in the right hand side (RHS) of the equality (10.20), and collecting all terms that are multiples of $\varphi_{j+1,m}$, we get

$$\begin{aligned} \text{RHS} &= \sum_{\ell\in\mathbb{Z}} a_{j,\ell} \sum_{m\in\mathbb{Z}} h_{m-2\ell}\varphi_{j+1,m} + \sum_{\ell\in\mathbb{Z}} d_{j,\ell} \sum_{m\in\mathbb{Z}} g_{m-2\ell}\varphi_{j+1,m} \\ &= \sum_{m\in\mathbb{Z}} \left[\sum_{\ell\in\mathbb{Z}} h_{m-2\ell}a_{j,\ell} + g_{m-2\ell}d_{j,\ell} \right] \varphi_{j+1,m}. \end{aligned}$$

This calculation, together with equation (10.20), 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 up-samplings, we obtain

$$a_{j+1,m} = h * Ua_j(m) + g * Ud_j(m), \tag{10.21}$$

where here we first up-sample the approximation and detail coefficients to restore the right dimensions, then convolve with the filters, and finally add the outcomes.

The synthesis phase of the *subband filtering scheme* can be represented schematically as follows.

$$\begin{array}{ccccccc} a_j \rightarrow \boxed{\uparrow 2} \rightarrow \boxed{*h} & & & & \\ & \searrow & & & \\ & & \oplus \rightarrow a_{j+1} & & \\ & \nearrow & & & \\ d_j \rightarrow \boxed{\uparrow 2} \rightarrow \boxed{*g} & & & & \end{array}$$

EXERCISE 10.22. Verify (10.21). $\Diamond$

The process of computing the coefficients can be represented by the following tree or *pyramid scheme*.

$$\begin{array}{ccccccccc}
a_J & \to & a_{J-1} & \to & a_{J-2} & \to & a_{J-3} & \cdots \\
 & \searrow & & \searrow & & \searrow & & \\
 & & d_{J-1} & & d_{J-2} & & d_{J-3} & \cdots
\end{array}$$

The reconstruction can also be illustrated with a pyramid scheme, as follows:

$$\begin{array}{ccccccccc}
a_n & \to & a_{n+1} & \to & a_{n+2} & \to & a_{n+3} & \cdots \\
 & \nearrow & & \nearrow & & \nearrow & & \\
d_n & & d_{n+1} & & d_{n+2} & & \cdots &
\end{array}$$

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 10.23. The cascade algorithm can be used to produce very good approximations for both ψ and φ, and this is how pictures of the wavelets and the scaling functions are obtained. 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 $\mathbf{V}_0 \subset \mathbf{V}_j \perp \mathbf{W}_j$ 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,

$$H \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \cdots \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \{\langle \varphi, \varphi_{n+1,k}\rangle\}.$$

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).

This is done similarly for the wavelet ψ. Notice that 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:

$$G \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \cdots \to \boxed{\uparrow 2} \to \boxed{\;*H\;} \to \{\langle \psi, \varphi_{j,k}\rangle\}.$$

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. $\Diamond$

Exercise 10.24. Use the algorithm described in Example 10.23 to create a picture of the scaling and the wavelet functions corresponding to the Daubechies wavelets $db2$ after 4 iterations and after 10 iterations. The coefficients for $db2$ are listed in Example 10.17. $\Diamond$

Altogether, we have a *perfect reconstruction filter bank*:

$$
\begin{array}{ccccc}
 & & \boxed{*\widetilde{h}} \to \boxed{\downarrow 2} \to a_{j-1} \to \boxed{\uparrow 2} \to \boxed{*h} & & \\
a_j \to & \begin{array}{c}\nearrow \\ \searrow\end{array} & & \begin{array}{c}\searrow \\ \nearrow\end{array} & \oplus \to a_j \\
 & & \boxed{*\widetilde{g}} \to \boxed{\downarrow 2} \to d_{j-1} \to \boxed{\uparrow 2} \to \boxed{*g} & &
\end{array}
$$

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. If we replace $\widetilde{h}$ 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^*.

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 holds: (*Riesz basis property*) there exist A, $B > 0$ such that for all $f \in L^2(\mathbb{R})$,

$$\sum_{j,k} |\langle f, \psi^*_{j,k}\rangle|^2 \sim \|f\|_2^2 \sim \sum_{j,k} |\langle f, \psi_{j,k}\rangle|^2,$$

where $A \sim B$ 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.

EXERCISE 10.25. Consider two linearly independent vectors in $\mathbb{R}^2$, $\vec{v}_1 = (x_1, y_1)$, $\vec{v}_2 = (x_2, y_2)$. They form a basis of $\mathbb{R}^2$. Find a pair of so-called *dual vectors* $\vec{v}_1^*$, $\vec{v}_2^* \in \mathbb{R}^2$, such that given any vector $\vec{u} \in \mathbb{R}^2$,

$$\vec{u} = \langle \vec{u}, \vec{v}_1^*\rangle \vec{v}_1 + \langle \vec{u}, \vec{v}_2^*\rangle \vec{v}_2.$$

Show that $\vec{u} = \langle \vec{u}, \vec{v}_1\rangle \vec{v}_1^* + \langle \vec{u}, \vec{v}_2\rangle \vec{v}_2^*$. Find positive numbers c and C such that

$$c \sum_{j=1}^{2} |\langle \vec{u}, \vec{v}_j^*\rangle|^2 \leq |\vec{u}|^2 \leq C \sum_{j=1}^{2} |\langle \vec{u}, \vec{v}_j^*\rangle|^2.$$

Draw a picture explaining the orthogonality properties among these vectors. The numbers c, C should be related to the angle between the vectors $\vec{v}_1$ and $\vec{v}_2$. $\Diamond$

A general filter bank is a sequence of convolutions and other simple operations such as up- and down-sampling. 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 operator can cut out various frequencies if the corresponding Fourier multiplier vanishes (or is very small) at those frequencies. Consult Strang and Nguyen's book [**SN**] for more information.

Filter banks can be implemented quickly because of the Fast Fourier Transform. Remember that convolution becomes multiplication on the Fourier side, that is a diagonal matrix (need only N products, whereas ordinary matrix multiplication requires N^2 products), and to go back and forth from Fourier to time domain, we can use the Fast Fourier Transform, so the total number of operations is of the order $N \log N$.

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 is available online.

10.6. 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 properties. Some of them are competing against each other, so it is up to the user to decide which ones are most efficient for her problem. The most popular conditions are orthogonality, compact support, vanishing moments, symmetry and smoothness.

Orthogonality: Orthogonality allows for straightforward calculation of the coefficients (via inner products with the basis elements). It guarantees that energy is preserved. However sometimes orthogonality can be substituted by *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.

Compact support: We have already stressed that compact support is important for numerical purposes, such as implementation of the FIR. In terms of detecting point singularities, it is clear that if the signal f has a singularity at t_0 then if t_0 is inside the support of $\psi_{j,n}$, 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 their support contains t_0). The shorter the support the fewer wavelets interacting with the singularity.

We have already mentioned that compact support of the scaling function coincides with FIR. Moreover if the low-pass filter is supported on $[N_1, N_2]$, so is φ, and it is not hard to see that ψ has support of the same length $(N_2 - N_1)$ but centered at $1/2$.

Smoothness: The regularity of the wavelet has effect on the error introduced by thresholding or quantizing the wavelet coefficients. Suppose an error ε is added to the coefficient $\langle f, \psi_{j,k}\rangle$. Then we add an error of the form $\varepsilon\psi_{j,k}$ 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 has exponential decay. Therefore there is no hope of finding an orthogonal wavelet that is C^∞ and has compact support.

Vanishing moments: A function ψ *has p vanishing moments* if for all $k = 0, 1, \dots, p-1$,

$$\int_{\mathbb{R}} x^k \psi(x)\, dx = 0.$$

If ψ has p vanishing moments, then ψ is orthogonal to polynomials of degree $p-1$. Smooth functions f have small fine-scale wavelet coefficients. More precisely, if the function to be analyzed is k-regular, 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.

The constraints imposed on orthogonal wavelets imply that if ψ has p vanishing moments then its support is of length at least $2p-1$. 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. It 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. If we use *multiwavelets*[1], we can construct them to be orthogonal, smooth, compactly supported and symmetric. Some wavelets have been designed to be nearly symmetric (Daubechies symmlets, for example).

10.7. A catalog of wavelets

We list the main wavelets and indicate their properties.

Haar wavelet: Perfectly localized in time, less localized in frequency, discontinuous. Symmetric. Shortest possible support, only one vanishing moment, hence not well adapted to approximating smooth functions.

Shannon wavelet: This wavelet does not have compact support, however it is C^∞. It is band-limited, but its Fourier transform is discontinuous, hence $\psi(t)$ decays like $1/|t|$ at infinity. $\widehat{\psi}(\xi)$ is zero in a neighborhood of $\xi = 0$, hence all its derivatives are zero at $\xi = 0$, and so ψ has an infinite number of vanishing moments.

Meyer wavelet: This is a symmetric band-limited function whose Fourier transform is smooth, hence $\psi(t)$ has faster decay at infinity. The scaling function is also band-limited. Hence both ψ and φ are C^∞. ψ has an infinite number of vanishing moments. (This wavelet was found by Strömberg in 1983, and it went unnoticed for several years).

Battle-Lemarié spline wavelets: Polynomial splines of degree m. The wavelet ψ has $m+1$ vanishing moments. They don't have compact support, but they have exponential decay. Since they are polynomial splines of degree m, they are $m-1$ times continuously differentiable. For m odd, ψ is symmetric around $1/2$. For m even, it is antisymmetric around $1/2$. The linear spline wavelet is the Franklin wavelet. [*** Give a reference for Franklin? Ph. Franklin, *A set of continuous orthogonal functions*. Math. Annalen 100 (1928) pp.522–529.]

[1]In this case one has more than one scaling function and more than one wavelet function, whose dilates and translates provide bases in the MRA. [*** or 'a basis in the MRA'?]

Daubechies compactly supported wavelets: They have compact support of minimal length for any given number of vanishing moments. More precisely, if ψ has p vanishing moments, then the filters have length $2p$ (or $2p$ *taps*). For large p, φ and ψ are uniformly Lipschitz α of the order $\alpha \sim 0.2p$. They are asymmetric. When $p = 1$ we recover the Haar wavelet.

Daubechies symmlets: p vanishing moments, minimum support of length $2p$, as symmetric as possible.

Coiflets: ψ has p vanishing moments, minimum support, and φ has $p-1$ moments vanishing (from the second to the p^{th} moment, never the first since $\int \varphi = 1$. This extra property requires enlarging the support of ψ to length $(3p-1)$. This time if we approximate a regular function f by a Taylor polynomial, then the approximation coefficients satisfy

$$2^{J/2}\langle f, \varphi_{J,k}\rangle \sim f(2^J k) + O(2^{-(k+1)J}).$$

Hence at fine scale J, the approximation coefficients are close to the signal samples.

The coiflets were constructed by Daubechies after Coifman requested them for the purpose of applications to almost diagonalization of singular integral operators. We will come back to this point in Chapter 11.

Mexican hat: 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 to normalize it in L^2. It does not come from an MRA, and it is not orthogonal. It is appropriate for continuous wavelet transform. It has exponential decay but not compact support. According to Daubechies, this function is popular in vision analysis. [*** Reference?]

Morlet wavelet: 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 continuous wavelet transform. It has exponential decay but not compact support.

Spline biorthogonal wavelets: These are compactly supported. There are two positive integer parameters N, N^*. N^* determines the scaling function φ^*, it is a spline of order $[N^*/2]$. The other scaling function and both wavelets depend on both parameters. 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. ψ also has support increasing with N and vanishing moments as well. 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 performing image analysis for example.

All the previous wavelets are encoded in Matlab. One can review them and their properties online, and we encourage the reader to do so.

10.8. Wavelet packets

To perform the wavelet transform we iterate 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, and 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*. Wickerhauser's book is a good source of information on this topic [**Wic**].

Denote the spaces by $\mathbf{W}_{j,n}$, where j is the scale as before, and n determines the "frequency".

$$
\begin{tikzcd}
 & & \mathbf{W}_{-2,3} \\
 & \mathbf{W}_{-1,1} \arrow[ur] \arrow[dr] & \\
\mathbf{W}_{0,0} \arrow[ur] \arrow[ddr] & & \mathbf{W}_{-2,2} \\
 & & \mathbf{W}_{-2,1} \\
 & \mathbf{W}_{-1,0} \arrow[ur] \arrow[dr] & \\
 & & \mathbf{W}_{-2,0}
\end{tikzcd}
$$

1. Haar packet binary tree: two levels.
Leaves determine Walsh basis.

$$
\begin{tikzcd}
 & \mathbf{W}_{-1,1} & \\
\mathbf{W}_{0,0} \arrow[ur] \arrow[dr] & & \mathbf{W}_{-2,1} \\
 & \mathbf{W}_{-1,0} \arrow[ur] \arrow[dr] & \\
 & & \mathbf{W}_{-2,0}
\end{tikzcd}
$$

2. Connected binary subtree.
Leaves determine Haar basis.

Notice that $\mathbf{W}_{0,0} = \mathbf{V}_0$, more generally, $\mathbf{W}_{j,0} = \mathbf{V}_j$, and $\mathbf{W}_{j,1} = \mathbf{W}_j$. We also know that the spaces $\mathbf{W}_{j-1,2n}$ and $\mathbf{W}_{j-1,2n+1}$ are orthogonal and their direct sum is $\mathbf{W}_{j,n}$. Therefore the leaves of every connected binary subtree[2] of the wavelet packet tree correspond to an orthogonal basis of the initial space.

Each of the bases encoded in the wavelet packet corresponds to a dyadic tiling of the phase plane in Heisenberg boxes of area one. They provide a much richer time–frequency analysis.

Each of the spaces $\mathbf{W}_{j,n}$ is generated by the integer shifts of a wavelet function at scale j and frequency n. More precisely, let $\omega_{j,k,n}(t) = 2^{-j/2}\omega_n(2^{-j}t - k)$, where $n \in \mathbb{N}$, j, $k \in \mathbb{Z}$, and

$$
\omega_{2n}(t) = \sqrt{2}\sum h_k \omega_n(2t-k), \quad \omega_0 = \varphi,
$$
$$
\omega_{2n+1}(t) = \sqrt{2}\sum g_k \omega_n(2t-k), \quad \omega_1 = \psi,
$$

where $\{h_k\}$ and $\{g_k\}$ are the low and high-pass filters of the MRA. Then

$$
\mathbf{W}_{j,n} = \operatorname{span}\{\omega_{j,k,n} : k \in \mathbb{Z}\} = \Big\{ f = \sum_k a_{j,k,n} w_{j,k,n} : \sum_k |a_{j,k,n}|^2 < \infty \Big\}.
$$

Example 10.26. (*The Walsh Functions*) For the Haar function, the corresponding wavelet packet is described by the equations

$$
\begin{aligned}
\omega_{2n}(t) &= \omega_n(2t) + \omega_n(2t-1) \\
\omega_{2n+1}(t) &= \omega_n(2t) - \omega_n(2t-1).
\end{aligned}
$$

The functions so obtained are the *Walsh functions*. They are step functions that play the role of the sines and cosines. ◊

[2]Starting from the "top" node ($\mathbf{W}_{0,0}$), we allow the node to have either two offspring or none. The same rule applies to the offspring, if there are any. The nodes that have no offspring are the *leaves*.

EXERCISE 10.27. Identify all possible bases for the Haar packet binary tree with three levels. How many bases can you find? Draw the corresponding phase plane diagrams. In particular, draw the Walsh functions. ◊

EXERCISE 10.28. Use MATLAB to plot the n^{th} Walsh function $\omega_n(t)$ and the functions $\sin nt$ and $\cos nt$ for $-2\pi \leq t \leq 2\pi$ on the same axes, for $n = 1, 2, \ldots, 5$. Notice that the Walsh functions resemble piecewise-constant versions of the trigonometric functions, with approximately the right frequencies. ◊

EXERCISE 10.29. What are the Walsh functions in the finite-dimensional case? ◊

We have seen that a signal of length $N = 2^J$ can be decomposed in 2^N different ways, the number of binary subtrees of a complete binary tree of depth J. This is a large number, and one would like to search efficiently in the tree to obtain the best basis with respect to some criteria.

Functionals verifying an additive-type property are well suited for searches of this type. Coifman and Wickerhauser introduced a number of such functionals [**CW**] [*** Make sure this is the correct paper], among them some *entropy criteria*. Given a signal s and $(s_i)_i$ its coefficients in an orthonormal basis. The entropy E must be an additive cost function such that $E(0) = 0$ and $E(s) = \sum_i E(s_i)$. There are fast algorithms that allow one to search in the wavelet packet tree for the orthonormal basis that minimizes some given entropies. (Four such algorithms are encoded in MATLAB.) Furthermore, the search can be performed in $O(N \log(N))$ operations. This search procedure is implemented in the Wavelet Toolbox.

The wavelet packet and *cosine packet* libraries create large libraries of orthogonal bases, all of which have fast algorithms. The Fourier and wavelet bases are particular examples in this time–frequency library; so are Gabor-like bases.

10.9. Two-dimensional wavelets

There is a standard procedure to construct bases in 2-D space from given bases in 1-D, the *tensor product*. In particular, given a wavelet basis $\{\psi_{j,k}\}$ 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, that is the scaling parameters i, j can be anything.

EXERCISE 10.30. Show that if $\{\psi_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for a closed subspace $\mathbf{V} \subset L^2(\mathbb{R})$, then the family of functions defined on $\mathbb{R}^2$ by

$$\psi_{m,n}(x,y) = \psi_m(x)\psi_n(y),$$

is an orthonormal basis for $\mathbf{V} \otimes V$. Here $\mathbf{V} \otimes V$ is the closure of the linear span of the functions $\{\psi_{m,n}\}_{m,n\in\mathbb{N}}$. ◊

EXERCISE 10.31. What would the trigonometric basis in $L^2([0,1]^2)$ be? How about the finite-dimensional trigonometric basis, say in two dimensions? ◊

We would like to use this idea but at the level of the approximation spaces $\mathbf{V}_j$ in the MRA. For each scale j, the family $\{\varphi_{j,k}\}_k$ is an orthonormal basis of $\mathbf{V}_j$.

Consider the tensor products $\varphi_{j,k,n}(x,y) = \varphi_{j,k}(x)\varphi_{j,n}(y)$ of these functions. Let $\mathcal{V}_j$ be the closure in $L^2(\mathbb{R}^2)$ of the linear span of those functions, in other words

$$\mathcal{V}_j = \mathbf{V}_j \otimes \mathbf{V}_j := \Big\{ f(x,y) = \sum_{n,k} a_{j,n,k}\varphi_{j,k,n}(x,y) : \sum_{n,k} |a_{j,n,k}|^2 < \infty \Big\}.$$

Notice that we are not mixing scales at the level of the MRA. It is not hard to see that the spaces $\mathcal{V}_j$ form an MRA in $L^2(\mathbb{R}^2)$, with scaling function

$$\varphi(x,y) = \varphi(x)\varphi(y).$$

Therefore 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 2 on both variables, and the other conditions are clear.

The orthogonal complement of $\mathcal{V}_j$ in $\mathcal{V}_{j+1}$ is denoted by $\mathcal{W}_j$. The space $\mathcal{W}_j$ can be viewed as the direct sum of three tensor products, namely

$$\mathcal{W}_j = (\mathbf{W}_j \otimes \mathbf{W}_j) \oplus (\mathbf{W}_j \otimes \mathbf{V}_j) \oplus (\mathbf{V}_j \otimes \mathbf{W}_j).$$

Therefore *three wavelets* are necessary to span the detail spaces:

$$\psi^d(x,y) = \psi(x)\psi(y), \qquad \psi^v(x,y) = \psi(x)\varphi(y), \qquad \psi^h(x,y) = \varphi(x)\psi(y),$$

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 in those directions.

The same approach works in higher dimensions. There are $2^n - 1$ wavelet functions and one scaling function, where n is the dimension.

EXERCISE 10.32. Describe a three-dimensional MRA. (These are useful for video compression.) $\Diamond$

This construction has the advantage that the bases are separable, implementing the fast two dimensional wavelet transform is not difficult. In fact it can be done by successively applying the one-dimensional FWT. The disadvantage is that the analysis is very axis-dependent, which might not be desirable for certain applications.

EXAMPLE 10.33. (*The Two-Dimensional Haar Basis*) The scaling function is the characteristic function

$$\varphi(x,y) = \chi_{[0,1]^2}(x,y)$$

of the unit cube. The following pictures help us to understand the nature of the two-dimensional Haar wavelets and scaling function.

$$\underset{\varphi(x,y)}{\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 1 & 1 \\ \hline\end{array}} \qquad \underset{\psi^d(x,y)}{\begin{array}{|c|c|}\hline -1 & 1 \\ \hline 1 & -1 \\ \hline\end{array}} \qquad \underset{\psi^h(x,y)}{\begin{array}{|c|c|}\hline 1 & -1 \\ \hline 1 & -1 \\ \hline\end{array}} \qquad \underset{\psi^v(x,y)}{\begin{array}{|c|c|}\hline -1 & -1 \\ \hline 1 & 1 \\ \hline\end{array}}$$

$\Diamond$

EXERCISE 10.34. Describe the two-dimensional discrete Haar MRA. $\Diamond$

There are *non-separable* two-dimensional MRA's. The most famous one corresponds to an analogue of the Haar basis. The scaling function is the characteristic function of a two-dimensional set. It turns out that the set has to be rather complicated. In fact it is a self-similar set with fractal boundary, the so-called *twin dragon*. [*** Cool. Include a reference. Point to the project too.]

10.10. 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 case of Amanda and her mother in Chapter 1. Having fewer data points allows for faster transmission and easier storage. In the presence of noise, one also wants to separate the noise from the signal, and one would like to have a basis that concentrates the signal in a few large coefficients and delegates the noise to very small coefficients.

In both the deterministic and the noisy case, the steps to follow are:

- Transform the data, find coefficients with respect to a given basis.
- Threshold the coefficients. Essentially one keeps the large ones and discards the small ones. Information is lost in this step, so perfect reconstruction is no longer possible.
- Reconstruct with the thresholded coefficients, and hope that the resulting compressed signal is a good approximation to your original signal, in other words that you have successfully denoised the original.

Wavelet bases are good for decorrelating coefficients. They are also good for denoising in the presence of *white noise*.

The crudest approach would be to use the projection into an approximation space as your compressed signal, discarding all the details after 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 technique is required, called *thresholding*. There are different thresholding techniques. The most popular are *hard thresholding* (it is a keep-or-toss thresholding), and *soft thresholding* (the coefficients are attenuated following a linear scheme). There is also the issue about thresholding individual coefficients or *block-thresholding*. 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. Both thresholding and threshold selection principles are encoded in MATLAB.

In traditional approximation theory there are two possible methods, *linear* approximation and *non-linear* approximation.

- *Linear approximation:* In linear approximation, one selects *a priori* N elements in the basis and projects onto the subspace generated by those elements, regardless of the function that is being approximated. It is a linear scheme:

$$P_N^l f = \sum_{n=1}^{N} \langle f, \psi_n \rangle \psi_n.$$

- *Non-linear approximation:* In non-linear approximation, one chooses 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:

$$P_N^{nl} f = \sum_{n=1}^{N} \langle f, \psi_{n,f} \rangle \psi_{n,f}.$$

The non-linear approach has proven quite successful. There is a lot more information about these issues in [**Mall98**, Chapters 9 and 10]. See also Project 12.5.

CHAPTER 11

The Hilbert transform

In this chapter we introduce a very important operator in harmonic analysis, the *Hilbert transform*. It is a prototypical example for two general classes of operators: *Fourier multipliers* and *singular integral operators*. These operators appear naturally in many areas of mathematics, including complex analysis, Fourier analysis, partial differential equations, and stochastic processes. We will emphasize here the connections of the Hilbert transform with Fourier theory.

We are also interested in the mapping properties of the Hilbert transform, and in particular what it does to L^p-functions. We had a similar (abbreviated) discussion for the Fourier transform in Section 8.6. Here we use this discussion to introduce interpolation techniques, as well as the concepts of strong and weak boundedness of operators. The interpolation results provide quick proofs of some of the most beautiful and useful inequalities in analysis. We have used these inequalities throughout the book: the Hausdorff–Young inequality (for the Fourier transform), the general Hölder inequality, and Young's inequality (for convolutions). We take the opportunity to discuss other inequalities mentioned earlier: Hölder's inequality, Minkowski's inequality (the triangle inequality for L^p), and Minkowski's integral inequality. To close, we use these inequalities to prove an approximation of the identity result in $L^p(\mathbb{R})$.

11.1. Introduction

The *Hilbert transform* can be defined on the Fourier side by the formula

$$(Hf)^\wedge(\xi) := -i \operatorname{sgn}(\xi)\widehat{f}(\xi). \tag{11.1}$$

Here the signum function sgn is defined by

$$\operatorname{sgn}(\xi) = \begin{cases} -1, & \text{if } \xi < 0; \\ 0, & \text{if } \xi = 0; \\ 1, & \text{if } \xi > 0. \end{cases}$$

Since the Fourier transform is a bijection on $L^2(\mathbb{R})$, formula (11.1) does define an operator on $L^2(\mathbb{R})$: the Hilbert transform of f is the function whose Fourier transform is $-i\operatorname{sgn}(\xi)$ times the Fourier transform of f. Moreover, the Hilbert transform is an *isometry* on $L^2(\mathbb{R})$, since

$$\|Hf\|_2 = \|\widehat{Hf}\|_2 = \|\widehat{f}\|_2 = \|f\|_2,$$

using Plancherel's Identity twice. The definition (11.1) also implies that $H^2 = -I$, where $-I$ denotes the negative of the identity operator, since

$$(H^2 f)^\wedge(\xi) = -i \operatorname{sgn}(\xi)(Hf)^\wedge(\xi) = (-i \operatorname{sgn}(\xi))^2 \widehat{f}(\xi) = -\widehat{f}(\xi).$$

EXERCISE 11.1. Let F denote the Fourier transform: $Ff = \widehat{f}$. Show that F^4 is the identity operator on those spaces where the Fourier transform is a bijection: $\mathcal{S}(\mathbb{R})$, $L^2(\mathbb{R})$, and $\mathcal{S}'(\mathbb{R})$. $\Diamond$

It is desirable to obtain alternative proofs of the fact that the Hilbert transform is an isometry, independent of the Fourier transform. There are several alternative proofs of a seemingly less powerful statement. Namely, there exists a positive constant C such that for all $f \in L^2(\mathbb{R})$,

$$\|Hf\|_2 \leq C\|f\|_2.$$

This inequality means that the operator H is *bounded* in L^2.

DEFINITION 11.2. An operator T is *bounded on* a space X if $T : X \to X$ and the X-norm of each Tf, $f \in X$, is no more than a fixed constant multiple of the X-norm of f:

$$\|Tf\|_X \leq C\|f\|_X,$$

where C is a constant independent of f.

The Hilbert transform is an example of a *Fourier multiplier*. A *multiplier* is an operator that is defined on the Fourier side, at least for nice, smooth functions, by multiplication of the Fourier transform of f by a function $m(\xi)$:

$$(Tf)^\wedge(\xi) := m(\xi)\widehat{f}(\xi). \tag{11.2}$$

The function $m(\xi)$ is called the *symbol* of the operator. In the case of the Hilbert transform the symbol is $m_H(\xi) = -i\,\mathrm{sgn}(\xi)$, a bounded function. If $m \in L^\infty(\mathbb{R})$, then the Fourier multiplier T_m with symbol m is a bounded operator in $L^2(\mathbb{R})$. Indeed, using Plancherel's Identity twice,

$$\|T_m f\|_2 = \|\widehat{T_m f}\|_2 = \left(\int_{\mathbb{R}} |m(\xi)\widehat{f}(\xi)|^2\,d\xi\right)^{1/2} \leq \|m\|_\infty \|\widehat{f}\|_2 = \|m\|_\infty \|f\|_2.$$

The converse is also true; it depends on the properties of essentially bounded functions.

Similarly, we can define boundedness in other L^p-spaces. Given p and q with $1 \leq p \leq \infty$ and $1 \leq q \leq \infty$, we say that an *operator T is bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$*[1] if there exists a positive constant C such that

$$\|Tf\|_q \leq C\|f\|_p \qquad \text{for all } f \in L^p(\mathbb{R}).$$

We discussed boundedness properties in $L^p(\mathbb{R})$ for the Fourier transform in Section 8.6. In particular we mentioned that the Fourier transform is a bounded mapping from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$, provided that $1 \leq p \leq 2$ and p, q are dual exponents, that is, $\frac{1}{p} + \frac{1}{q} = 1$. Furthermore, the *Hausdorff–Young Inequality* holds for all such p, q, and for all $f \in L^p(\mathbb{R})$:

$$\|\widehat{f}\|_q \leq \|f\|_p.$$

We prove it in Section 11.5 as Corollary 11.27.

[1] For linear operators, boundedness is equivalent to continuity at zero. Thus a linear operator T is bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$ if and only if

$$\|Tf_n\|_q \to 0 \qquad \text{whenever } \|f_n\|_p \to 0 \text{ and } f_n \in L^p(\mathbb{R}).$$

If we try to use this inequality as we used Plancherel's Identity to obtain an L^p-boundedness result for the Hilbert transform for $p \neq 2$, we face some problems. Let us try. Start with $f \in L^p(\mathbb{R})$, $1 \leq p < 2$. We know that $\widehat{f} \in L^q(\mathbb{R})$. Furthermore, if we use the Fourier multiplier definition (the only definition we have so far), at least for Schwartz functions ϕ with zero mean (i.e. $\widehat{\phi}(0) = 0$) (which are dense in $L^p(\mathbb{R})$ for all p such that $1 \leq p < \infty$), and for which we can use the inversion formula to obtain a function in $\mathcal{S}(\mathbb{R})$ that will be its Hilbert transform, namely

$$H\phi = (m_H \widehat{\phi})^{\vee} \in \mathcal{S}(\mathbb{R}).$$

By definition, $\|(H\phi)^{\wedge}\|_q = \|m_H\widehat{\phi}\|_q = \|\widehat{\phi}\|_q$, but to obtain meaningful estimates not on the Fourier side the only tool we have is the Hausdorff–Young inequality, which tells us that $\|\widehat{\phi}\|_q \leq \|\phi\|_p$ for appropriate p, q, and $\|(H\phi)^{\wedge}\|_q \leq \|H\phi\|_p$. When we try to put together all this information, we realize that some inequalities are going in the wrong direction (the beauty of equalities, such as Plancherel's Identity, is that they go in both directions!). New ideas are needed.

Our experience with the time–frequency dictionary suggests that at least formally, on the time domain, the Hilbert transform should be given by convolution with a kernel (remember that convolution is transformed into multiplication by the Fourier transform):

$$Hf = (m_H \widehat{f}\,)^{\vee} = k_H * f,$$

where $k_H(x) = (m_H)^{\vee}(x)$. The problem is, what is the inverse Fourier transform of the symbol? For bounded symbols, like m_H, the inverse can certainly 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 for all bounded Fourier multipliers: they correspond to convolution operators with distributional kernels.

One can compute the inverse Fourier transform of $\widehat{Hf}$, at least when applied to Schwartz functions f. It is given formally by the following principal value integral:

$$(11.3) \qquad Hf(x) = \text{p.v.}\frac{1}{\pi}\int \frac{f(y)}{x-y}dy := \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|x-y|>\varepsilon}\frac{f(y)}{x-y}\,dy = k_H * f(x),$$

where $k_H(x) = \text{p.v.}\dfrac{1}{\pi x}$ is the principal value distribution introduced in Example Such-and-such. ***

[***Add an example: the distribution given by p.v. integration from -1 to 1 of $1/x$.]

The Hilbert transform is given by convolution with the distributional kernel $k_H(x) = \text{p.v.}\frac{1}{x}$, which is not an integrable function. Had the kernel been an integrable function $k \in L^1(\mathbb{R})$, then the *integral operator* T given by convolution against the kernel,

$$Tf(x) = k * f(x) = \int_{\mathbb{R}} k(x-y)f(y)\,dy$$

would have been automatically bounded in $L^p(\mathbb{R})$ for all $1 \leq p \leq \infty$, by *Young's inequality*,

$$\|f * k\|_p \leq \|k\|_1\|f\|_p,$$

which we prove in Section 11.5 as Corollary 11.34. Unfortunately the Hilbert transform kernel is not integrable, but still the Hilbert transform is bounded in $L^p(\mathbb{R})$ for all $1 < p < \infty$; and although it is not bounded at the endpoints $p = 1$ and $p = \infty$, there are appropriate substitutes.

THEOREM 11.3. *The Hilbert transform is a bounded operator in $L^p(\mathbb{R})$ for all $1 < p < \infty$: there exists a constant $C_p > 0$ such that for all $f \in L^p(\mathbb{R})$,*

$$\|Hf\|_p \leq C_p \|f\|_p.$$

The Hilbert transform is the prototypical example of a generalized Calderón–Zygmund singular integral operator. These are operators given by integration against a kernel $K(x, y)$,

$$Tf(x) = \text{p.v.} \int_{\mathbb{R}} K(x, y) f(y)\, dy.$$

The kernel grows like the Hilbert kernel near the diagonal $y = x$, and therefore the integral must be understood in some principal value sense. The cancellation encoded in the symmetry of p.v.$\frac{1}{x}$ is replaced by an appropriate estimate on the gradient or on differences of the kernel[2].

The informal justification for formula (11.3) is the calculation of the Fourier transform of the perfectly well defined kernel $k_{\varepsilon,R}(y) = \frac{1}{\pi y}\chi_{\{y\in\mathbb{R}:\varepsilon<|y|<R\}}(y) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$,

$$\begin{aligned}
\widehat{k_{\varepsilon,R}}(\xi) &= \int_{\varepsilon<|y|<R} \frac{1}{\pi y} e^{-2\pi i y\xi}\, dy \\
&= \int_{\varepsilon}^{R} \frac{1}{\pi y} e^{-2\pi i y\xi}\, dy + \int_{-R}^{-\varepsilon} \frac{1}{\pi y} e^{-2\pi i y\xi}\, dy \\
&= \int_{\varepsilon}^{R} \frac{1}{\pi y} e^{-2\pi i y\xi}\, dy - \int_{\varepsilon}^{R} \frac{1}{\pi y} e^{2\pi i y\xi}\, dy \\
&= -2\int_{\varepsilon}^{R} \frac{\sin 2\pi y\xi}{\pi y}\, dy \\
&= -\frac{2}{\pi}\int_{2\pi\varepsilon\xi}^{2\pi R\xi} \frac{\sin t}{t}\, dt \\
&= -\operatorname{sgn}(\xi)\left[\frac{2}{\pi}\int_{2\pi\varepsilon|\xi|}^{2\pi R|\xi|} \frac{\sin t}{t}\, dt\right].
\end{aligned}$$

Notice that the quantity in brackets tends to $\frac{2}{\pi}\int_0^\infty \frac{\sin t}{t}\, dt = 1$ as $\varepsilon \to 0$ and $R \to \infty$. Therefore, the well-defined truncated Hilbert transforms $H_{\varepsilon,R}$ in $L^2(\mathbb{R})$, given by the equivalent formulas

$$H_{\varepsilon,R}f = k_{\varepsilon,R} * f, \qquad \widehat{H_{\varepsilon,R}f}(\xi) = \widehat{k_{\varepsilon,R}}(\xi)\widehat{f}(\xi),$$

can be seen to be uniformly bounded (in R and ε) in $L^2(\mathbb{R})$, by Plancherel. Furthermore, in the limit as $R \to \infty$ and $\varepsilon \to 0$, the symbols $\widehat{k_{\varepsilon,R}}(\xi)$ converge to the symbol $m_H(\xi) = -\operatorname{sgn}(\xi)$ of the Hilbert transform. Also, in $L^2(\mathbb{R})$, $H_{\varepsilon,R}f \to Hf$, where H is given by formula (11.3). By the continuity of the Fourier transform, it follows that

$$\widehat{H_{\varepsilon,R}f} \to m_H\widehat{f} \quad \Rightarrow \quad \widehat{Hf} = m_H\widehat{f}.$$

[2]A one-dimensional *standard kernel* K is a function on $\mathbb{R}\backslash\{0\}$ 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 $|x-y| > 2|x-w|$.

Recall that by definition $\int_{\mathbb{R}} = \lim_{R\to\infty} \int_R^{-R}$, so the symmetry at infinity is built in to the definition of the improper integral. The symmetry around zero in the truncation provides the right cancellation necessary to carry on the calculation.

See Exercise 11.10 for another idea on how to prove formula (11.3).

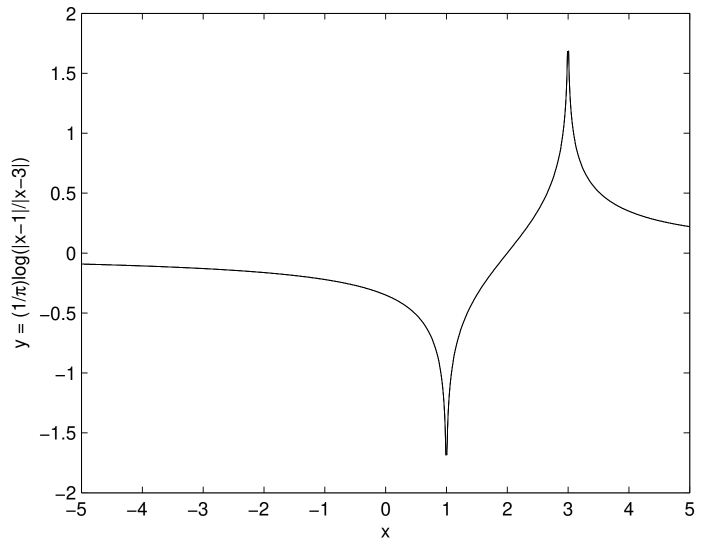


FIGURE 11.1. Graph of the Hilbert transform $H\chi_{[1,3]}$ of the characteristic function of the interval $[1,3]$.

EXERCISE 11.4. Show that the Hilbert transform of the characteristic function of the interval $I = [a, b]$ is given by

$$H\chi_I(x) = \frac{1}{\pi} \log \left| \frac{x-a}{x-b} \right|.$$

Deduce that the Hilbert transform does not map L^∞ to itself. $\Diamond$

The graph of $H\chi_{[1,3]}$ is shown in Figure 11.1.

11.2. Some history

Why did mathematicians get interested in the Hilbert transform? Here are a couple of classical problems where the Hilbert transform appeared naturally. The first one is in connection to Fourier series, the second one to complex analysis.

Natural extensions of the Hilbert transform to higher dimensions, the Riesz transforms, are connected to partial differential equations, although we will not explore these connections here. [*** References?]

11.2.1. Connection to Fourier series.

[*** Square definition of $\mathbb{T}$ with earlier $[-\pi, \pi]$.]

For functions integrable on $\mathbb{T} = [0, 1]$, the n^{th} Fourier coefficient is defined by the formula

$$\widehat{f}(n) := \int_0^1 f(\theta)e^{-2\pi in\theta}\,d\theta.$$

Since $L^2(\mathbb{T}) \subset L^1(\mathbb{T})$, the above integral is also well defined for square integrable functions f. We already know that the trigonometric system $\{e^{2\pi in\theta}\}_{n\in\mathbb{Z}}$ is an orthonormal complete system in $L^2(\mathbb{T})$. Therefore the following reconstruction and isometry formulae hold in $L^2(\mathbb{T})$:

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

The N^{th} partial sum is given by

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

Mathematicians asked, in the nineteenth century, for which 2π-periodic functions f is it true that

$$\lim_{N\to\infty} S_N f(\theta) = f(\theta)$$

at a given point $\theta \in \mathbb{T}$? Some partial answers were given, and something more than continuity at the point θ was always required, for example Dini's condition. In 1889, Du Bois-Reymond found a continuous function whose partial Fourier sum diverges at a point. Summation methods were introduced to counter these so-called pathologies. We discussed some of these issues in Section 1.4. With the advent of measure theory and L^p-spaces, new questions were formulated.

- **Is there convergence a.e.?** That is, does there exist an exceptional set E of measure zero, such that for all $\theta \in \mathbb{T}$ not included in E

$$\lim_{N\to\infty} S_N f(\theta) = f(\theta)\ ?$$

- **Is there convergence in the L^p-sense?** That is, does

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

The answer to the second question is positive for $1 < p < \infty$. For $p = 2$ it is Plancherel's Identity, and for $1 < p < \infty$ it is a consequence of the boundedness of the periodic analogue of the Hilbert transform in such L^p spaces, as we show below. The first question is much more difficult. For $p = 2$ the positive answer was given by L. Carleson in his celebrated paper [**Car**] published in 1965, settling the question for periodic continuous functions, which had remained open until then. Two years later, R. Hunt extended the result for the remaining values of p with $1 < p < \infty$; see [**Hun**]. The case $p = 1$ had been ruled out by A. Kolmogorov's famous example of an integrable function whose Fourier series diverges everywhere; see [**Kol**].

Define the *periodic Hilbert transform* on the Fourier side by

$$(\widetilde{H}f)^\wedge(n) = -i\ \mathrm{sgn}(n)\widehat{f}(n).$$

[*** Careful! Do we really want to use $\widetilde{H}$ here? So far we've tried to use $\widetilde{H}$ only for the reflection of a function.]

Exercise 11.5. Given $f(\theta) = \sin(2\pi\theta)$ on $[0, 1]$, find $\widetilde{H}f$. $\Diamond$

This is an example of a *Fourier multiplier* on the circle. These operators are linear transformations on $L^2(\mathbb{T})$ given by the formula

$$(Tf)^\wedge(n) = m(n)\widehat{f}(n),$$

where the sequence $\{m(n)\}_{n\in\mathbb{Z}}$ is the *symbol*.

If the sequence $\{m(n)\}_{n\in\mathbb{Z}}$ is a bounded sequence ($\ell^\infty(\mathbb{Z})$), more precisely, if there is $M > 0$ such that $|m(n)| \leq M$ for all $n \in \mathbb{Z}$, then

$$\|Tf\|^2_{L^2(\mathbb{T})} = \sum_{n\in\mathbb{Z}} |(Tf)^\wedge(n)|^2 = \sum_{n\in\mathbb{Z}} |m(n)\widehat{f}(n)|^2 \leq M^2 \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2 = M^2\|f\|^2_{L^2(\mathbb{T})},$$

where the first and last equalities are applications of Parseval's Identity. We conclude that if the sequence $\{m(n)\}_{n\in\mathbb{Z}}$ is in $\ell^\infty(\mathbb{Z})$ then the multiplier with symbol $\{m(n)\}_{n\in\mathbb{Z}}$ is bounded in $L^2(\mathbb{T})$. The converse is also true, and easy to prove in the periodic case.

Lemma 11.6. *A Fourier multiplier T_m on the circle $\mathbb{T}$ with symbol $\{m(n)\}_{n\in\mathbb{Z}}$ is bounded in $L^2(\mathbb{T})$ if and only if the sequence $\{m(n)\}_{n\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z})$.*

Proof. We have just proved the 'if' direction. For the 'only if' direction, suppose that T_m is bounded in $L^2(\mathbb{T})$. Then there exists a constant $C > 0$ such that for all $f \in L^2(\mathbb{T})$,

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

In particular this is true for the exponential functions $e_n(\theta) = e^{2\pi in\theta}$, $n \in \mathbb{Z}$. The exponential functions form an orthonormal basis in $L^2(\mathbb{T})$, so in particular they have norm 1. Their Fourier coefficients have a simple form:

$$\widehat{e_n}(k) = \begin{cases} 1, & \text{if } k = n; \\ 0, & \text{if } k \neq n. \end{cases}$$

Therefore, using Parseval 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 $n \in \mathbb{Z}$, and the sequence $\{m(n)\}_{n\in\mathbb{Z}}$ is bounded, as required. □

The operators H and $\widetilde{H}$ share similar boundedness properties. In particular, $\widetilde{H}$ is bounded in $L^p(\mathbb{T})$ for $1 < p < \infty$, that is, there exists a constant $C_p > 0$ such that for all $f \in L^p(\mathbb{T})$,

$$\|\widetilde{H}f\|_{L^p(\mathbb{T})} \leq C_p\|f\|_{L^p(\mathbb{T})}.$$

Note that the N^{th} partial Fourier sum of a nice function is also a Fourier multiplier, namely,

$$(S_N f)^\wedge(n) = m_N(n)\widehat{f}(n),$$

whose symbol is the sequence

$$m_N(n) = \begin{cases} 1, & \text{if } |n| \leq N; \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 11.7. Check that

$$m_N(n) = \frac{1}{2}\big(\operatorname{sgn}(n-N) - \operatorname{sgn}(n+N)\big).$$

Let M_N denote the modulation operator, defined by $M_N f(\theta) := f(\theta)e^{2\pi i\theta N}$. Verify that

$$\begin{aligned} i(M_N \widetilde{H} M_{-N})^\wedge(n) &= \operatorname{sgn}(n-N)\widehat{f}\,(n), \\ i(M_{-N} \widetilde{H} M_N)^\wedge(n) &= \operatorname{sgn}(n+N)\widehat{f}\,(n). \end{aligned}$$

(Remember that the Fourier transform converts modulations into translations.) $\Diamond$

The exercise implies, after applying the inverse Fourier transform, that

$$S_N = \frac{i}{2}(M_N \widetilde{H} M_{-N} - M_{-N}\widetilde{H} M_N). \tag{11.4}$$

The modulation operators preserve L^p-norms. In other words,

$$\|M_N f\|_{L^p(\mathbb{T})} = \|f\|_{L^p(\mathbb{T})} \qquad \text{for all } N \in \mathbb{Z}.$$

Exercise 11.8. Verify the preceding statement: modulations preserve L^p-norms. $\Diamond$

Theorem 11.9. *The N^{th} partial Fourier sums of a function $f \in L^p(\mathbb{T})$ converge to f in L^p-norm as $N \to \infty$.*

Proof. Formula (11.4) shows that the partial sums S_N are uniformly (in N) bounded in $L^p(\mathbb{T})$, for each $1 < p < \infty$. More precisely, take L^p-norms on both sides of (11.4), apply the triangle inequality in $L^p(\mathbb{T})$, then use the boundedness properties of M_N and $\widetilde{H}$, as follows:

$$\begin{aligned} \|S_N f\|_{L^p(\mathbb{T})} &\leq \frac{1}{2}\left[\|M_N \widetilde{H} M_{-N} f\|_{L^p(\mathbb{T})} + \|M_{-N}\widetilde{H} M_N f\|_{L^p(\mathbb{T})}\right] \\ &= \frac{1}{2}\left[\|\widetilde{H}(M_{-N} f)\|_{L^p(\mathbb{T})} + \|\widetilde{H}(M_N f)\|_{L^p(\mathbb{T})}\right] \\ &\leq \frac{C_p}{2}\left[\|M_{-N} f\|_{L^p(\mathbb{T})} + \|M_N f\|_{L^p(\mathbb{T})}\right] \\ &= C_p \|f\|_{L^p(\mathbb{T})}. \end{aligned}$$

We also know that the trigonometric polynomials are dense in $L^p(\mathbb{T})$. Given a trigonometric polynomial $f(\theta) = \sum_{|m|\leq M} a_m e^{2\pi i m\theta}$, 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 $L^p(\mathbb{T})$. We can now deduce from the Uniform Boundedness Principle (Theorem 9.34) that

$$\lim_{N\to\infty} \|S_N f - f\|_p = 0. \qquad \square$$

Therefore, the convergence in $L^p(\mathbb{T})$ of the partial Fourier sums is a consequence of the boundedness of the periodic Hilbert transform in those spaces.

11.2.2. Connection to complex analysis. Consider a real valued function $f \in L^2(\mathbb{R})$ and let $F(z)$ be twice its analytic extension to the upper half-plane $\mathbb{R}^2_+ = \{z = x + iy : y > 0\}$, suitably normalized. $F(z)$ can be explicitly computed by means of the well known *Cauchy integral formula*:

$$F(z) = \frac{1}{\pi i} \int_{\mathbb{R}} \frac{f(t)}{z - t}\, dt, \qquad z \in \mathbb{R}^2_+.$$

Notice the resemblance with the Hilbert transform. No principal value is needed here since the singularity is never achieved: $z \in \mathbb{R}^2_+$ cannot equal $t \in \mathbb{R}$. By separating the real and imaginary parts of the kernel, one can obtain explicit formulae for the real and imaginary parts of $F(z) = u(z) + iv(z)$ in terms of convolutions with the *Poisson* and *conjugate Poisson kernels* that we encountered in Section 7.6. Here are the formulae:

$$u(x + iy) = f * P_y(x), \qquad v(x + iy) = f * Q_y(x).$$

The function u is the *harmonic extension* of f to the upper half-plane, and the function v is the *harmonic conjugate* of u.

Exercise 11.10. Show that the Poisson kernel is given by $P_y(x) = \frac{1}{\pi}\frac{y}{x^2+y^2}$, and the conjugate Poisson kernel by $Q_y(x) = \frac{1}{\pi}\frac{x}{x^2+y^2}$. You have already shown in Exercise 8.43 that for each $y > 0$, $\widehat{Q_y}(\xi) = -i\,\mathrm{sgn}(\xi) e^{-2\pi|y\xi|}$. Therefore, as $y \to 0$, $\widehat{Q_y}(\xi)$ approaches $-i\,\mathrm{sgn}(\xi)$. On the other hand, show that as $y \to 0$, Q_y approaches the principal value distribution $\frac{1}{\pi x}$, which means in this case that for all test functions $\phi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{y\to 0} \int_{\mathbb{R}} Q_y(x)\phi(x)\, dx = \lim_{\varepsilon \to 0} \int_{|x|>\varepsilon} \frac{\phi(x)}{x}\, dx.$$

$\Diamond$

The Poisson kernel is an example of an approximation of the identity (Section 7.5). As such, the limit as $y \to 0$ of $u = P_y * f$ is f in the L^2-sense and almost everywhere. 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 $(Hf)^\wedge(\xi) = -i\,\mathrm{sgn}(\xi)\widehat{f}\,(\xi)$.

Figure 11.2. Picture illustrating.

By a limiting procedure on the unit disk, similar to the one described in the upper half-plane, one can conclude that the boundary value of the harmonic conjugate of the harmonic extension of a periodic, real-valued, continuously differentiable function f on $\mathbb{T}$ is given by

$$\widetilde{H} f(\theta) = \text{p.v.} \frac{1}{\pi} \int_0^1 f(t) \cot\big(\pi(\theta - t)\big)\, dt. \tag{11.5}$$

Here we are identifying the function $f(\theta)$ with $f(z)$, for $z = e^{2\pi i\theta}$. The singularity at the diagonal is comparable to that of the Hilbert transform. This is the appropriate analogue of the Hilbert transform on the unit circle. The analogy gets reinforced when we notice that on the Fourier side,

$$(\widetilde{H}f)^\wedge(n) = -i \,\mathrm{sgn}(n)\widehat{f}(n).$$

That is how we defined the periodic Hilbert transform in the previous section. [***Clarify which section.]

11.3. The function space Weak $(1,1)$

When $p = 1$, one can get away with a weaker notion of boundedness, which requires some of the ideas of measure theory. Suppose $g \in L^1(\mathbb{R})$, and let

$$E_\lambda(g) := \{x \in \mathbb{R} : |g(x)| > \lambda\} \subset \mathbb{R}.$$

First of all, functions in $L^1(\mathbb{R})$ are *measurable*. This means that the sets $E_\lambda(g)$ are themselves *measurable*; in other words we can assign a "length" or "measure" to each such set, which coincides with ordinary length when the set is an interval. This "measure" is additive for finite collections of disjoint sets, and has other convenient properties.

Assume T is a linear bounded operator in $L^1(\mathbb{R})$, namely $T : L^1(\mathbb{R}) \to L^1(\mathbb{R})$, and

$$\|Tf\|_1 = \int_{\mathbb{R}} |Tf(x)|\,dx \le C\|f\|_1.$$

Even if we are not fully acquainted with the Lebesgue integral, our experience with Riemann integration tells us that the following inequality is plausible,

$$\|Tf\|_1 = \int_{\mathbb{R}} |Tf(x)|\,dx \ge \int_{E_\lambda(Tf)} |Tf(x)|\,dx.$$

Now observe that by definition, if $x \in E_\lambda(Tf)$, then $|Tf(x)| > \lambda$, hence we can also believe[3] that

$$\int_{E_\lambda(Tf)} |Tf(x)|\,dx \ge \int_{E_\lambda(Tf)} \lambda\,dx = \lambda \int_{E_\lambda(Tf)} 1\,dx = \lambda \int_{\mathbb{R}} \chi_{E_\lambda(Tf)}(x)\,dx.$$

Finally if E_λ is an interval, then this last integral is equal to its length. Since we are assuming that $Tf \in L^1(\mathbb{R})$ then $E_\lambda(Tf)$ is a *measurable set*, and we can extend the notion of length to such sets, the so-called *Lebesgue measure*. The Lebesgue measure of a measurable set A is denoted by $m(A)$, and it coincides with the Lebesgue integral[4] of χ_A on $\mathbb{R}$, in other words

$$m(A) = \int_A 1\,dx = \int_{\mathbb{R}} \chi_A(x)\,dx.$$

(It reads "the measure of the set A".)

---

[3]The properties of the integral that we are using are: (i) for $f \ge 0$, and $A \subset B$ then $\int_A f \le \int_B f$; (ii) if $f \le g$ on A, then $\int_A f \le \int_B g$; (iii) if $\lambda \in \mathbb{R}$, then $\int_A \lambda f = \lambda \int_A f$; (iv) $\int_A 1dx = \int \chi_A(x)dx$.

[4]We do not define the Lebesgue integral in this book. To give a general idea, though, the difference between the Riemann and Lebesgue integrals is like the difference between (Riemann) adding loose change one coin at a time, and (Lebesgue) collecting them by denomination (pennies, dimes, quarters, etc), multiplying each denomination by the number of coins of that denomination, and then adding up.

What we have shown is that if T is bounded $L^1(\mathbb{R})$, then there is a constant $C \geq 1$ such that

$$m(\{x \in \mathbb{R} : |Tf(x)| > \lambda\}) \leq \frac{C}{\lambda}\|f\|_1. \tag{11.6}$$

An operator that satisfies inequality (11.6) is said to be of *weak type* $(1,1)$. We have just shown that bounded in $L^1(\mathbb{R})$ implies weak $(1,1)$; but the converse is, in general, false.

Exercise 11.11. Show that if $f \in L^p(\mathbb{R})$, then

$$m(\{x \in \mathbb{R} : |f(x)| > \lambda\}) \leq \left(\frac{\|f\|_p}{\lambda}\right)^p.$$

[*** For which p? Also spell out Chebyshev's inequality.] The special case when $p = 1$ is the celebrated *Chebyshev's inequality.* ♢

A measurable function f is said to belong to *weak*-$L^p(\mathbb{R})$ if there exists a constant $C > 0$ such that for all $\lambda > 0$

$$m(\{x \in \mathbb{R} : |f(x)| > \lambda\}) \leq \left(\frac{C}{\lambda}\right)^p.$$

Denote the smallest constant C that can be used in the above inequality by $[f]_p$, the *weak L^p-norm* of f. The previous exercise shows that $L^p(\mathbb{R})$ is a subset of weak $L^p(\mathbb{R})$, the next example shows that it is a proper subset.

Example 11.12. The function $f(x) = |x|^{-1/p}$ is not in $L^p(\mathbb{R})$, however it is in weak-$L^p(\mathbb{R})$.

Figure 11.3. Picture of f.

We can compute explicitly that $E_\lambda(f) = [-\lambda^{-p}, \lambda^{-p}]$. Therefore

$$m(E_\lambda(g)) = 2\lambda^{-p} = \left(\frac{2}{\lambda}\right)^p.$$

Hence f is in weak-$L^p(\mathbb{R})$. Furthermore, $[f]_p = 2$. ♢

An operator is said to be of *weak-type* (p,q) if there is a $C > 0$ such that for all $\lambda > 0$,

$$m(\{x \in \mathbb{R} : |Tf(x)| > \lambda\}) \leq \left(\frac{C\|f\|_p}{\lambda}\right)^q. \tag{11.7}$$

Exercise 11.13. Show that if T is bounded in $L^p(\mathbb{R})$, then T is of weak-type (p,p). ♢

Theorem 11.14. *The Hilbert transform is of weak-type* $(1,1)$*, but is not bounded in* $L^1(\mathbb{R})$*.*

Exercise 11.15 below shows that H cannot be bounded in $L^1(\mathbb{R})$, nor in $L^\infty(\mathbb{R})$. We will not show here that H is of weak-type (1,1). This is a consequence of the celebrated *Calderón–Zygmund decomposition* which, if we had space, could be discussed in a book at the level of this one. It is an example of another technique widely used in modern harmonic analysis, *stopping time techniques*. Bounded functions are mapped into the larger space BMO of functions of *Bounded Mean Oscillation*, which contains unbounded functions such as $\log|x|$. See [**Gar**, Chapter VI].

This behavior is shared by a large class of very important operators, known as *Calderón–Zygmund Singular Integral Operators*. The departure point of the Calderón–Zygmund theory is an *a priori* L^2-estimate; everything else unfolds from there. Having means other than Fourier analysis to obtain such an L^2-estimate becomes important in other contexts.

Exercise 11.15. Verify that the Hilbert transform of $\chi_{[0,1]}$ is given by

$$H\chi_{[0,1]}(x) = \log\left|\frac{x}{x-1}\right|.$$

Show that $H\chi_{[0,1]}$ is neither an integrable function nor a bounded function, despite the fact that the characteristic function $\chi_{[0,1]}$ is both. However, $H\chi_{[0,1]}$ must be in $L^2(\mathbb{R})$, and furthermore

$$\int_{\mathbb{R}} \left|\log\left|\frac{x}{x-1}\right|\right|^2 dx = \int_{\mathbb{R}} \left|\chi_{[0,1]}(x)\right|^2 dx = 1,$$

because H is an isometry.

Find $C > 0$ such that for all $\lambda > 0$,

$$m\{x \in \mathbb{R} : |H\chi_{[0,1]}(x)| \geq \lambda\} \leq \frac{C}{\lambda}.$$

This does not prove that H is of weak-type $(1,1)$, but it shows that the function $H\chi_{[0,1]}$ is in weak-$L^1(\mathbb{R})$. One could attempt to prove that H is of weak-type $(1,1)$ by observing that if f is a step function with L^1-norm one, $\|f\|_1 = 1$, then Hf is uniformly in weak $L^1(\mathbb{R})$, and then attempt to generalize to all L^1-functions by an approximation argument. ◊

[*** Clarify: would that approach work?]

We mentioned the Lebesgue Differentiation Theorem (equation (9.16)) in Section 9.3.4. We know the theorem holds for continuous functions as a consequence of the Fundamental Theorem of Calculus. A related operator, defined for all $x \in \mathbb{R}$ (possibly to be ∞), is the *Hardy–Littlewood maximal operator*

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

The maximal function is also of weak-type $(1,1)$; this is also a consequence of the Calderón–Zygmund Decomposition Lemma. This fact, together with the knowledge of the existence of the limit on a dense subset, allows us to prove the Lebesgue Differentiation Theorem for integrable functions. Maximal functions appear all over in harmonic analysis as controllers for other operators, and understanding their boundedness properties is very important.

11.3.1. The distribution function and L^p. Suppose f is a function such that $E_\lambda(f)$ is measurable for all $\lambda \geq 0$ (these are exactly the *measurable functions*. Step functions are always measurable). Then we can define the *distribution function* d_f of f on $[0, \infty)$ by

$$d_f(\lambda) = m\big(E_\lambda(f)\big) = m\{x \in \mathbb{R} : |f(x)| > \lambda\}.$$

Notice that d_f is always a non-negative decreasing function.

EXAMPLE 11.16. Let $f = 2\chi_{[-1,2]} - \chi_{[3,5]} + 7\chi_{[10,11]}$. Then the distribution function of f is

$$d_f = 6\chi_{[0,1)} + 4\chi_{[1,2)} + \chi_{[2,7)}.$$

The graphs of f and d_f are shown in Figure 11.3.1. $\Diamond$

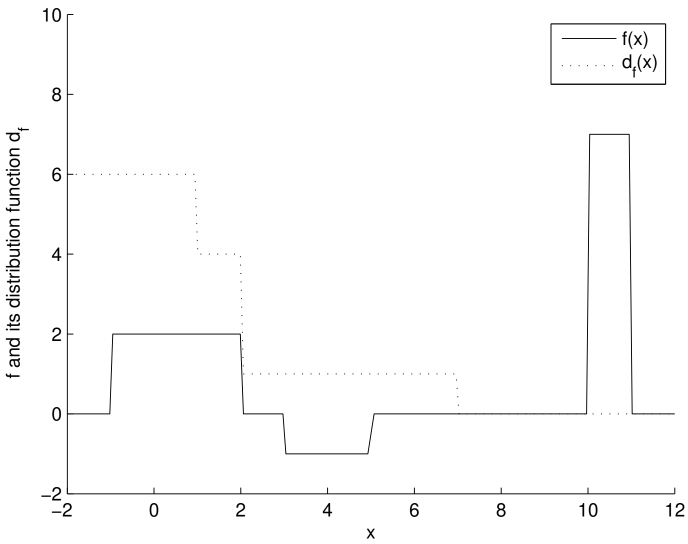


FIGURE 11.4. Graphs of the function $f(x) = 2\chi_{[-1,2)} - \chi_{[3,5)} + 7\chi_{[10,11)}$ and its distribution function $d_f = 6\chi_{[0,1)} + 4\chi_{[1,2)} + \chi_{[2,7)}$. [*** Shouldn't the formula for d_f account for d_f being nonzero for $x < 0$, since d_f is nondecreasing? Or is it only nondecreasing for nonnegative x? Fix. In fact, we have two variables x and λ on horizontal axis. And we've defined d_f only for nonnegative λ. Resolve.]

EXERCISE 11.17. Let $\{I_k : k = 1, \ldots, n\}$ be disjoint bounded intervals. Let f be the step function given by $f = \sum_{k=1}^n a_k \chi_{I_k}$. Find its distribution function d_f. $\Diamond$

EXERCISE 11.18. Show that $d_f(\lambda) = d_{\tau_h f}(\lambda)$ for all $h \in \mathbb{R}$, and for f a step function like the ones in Exercise 11.17. In other words, the distribution function of f coincides with the distribution function of any of its translates. (In fact this translation invariance holds for all measurable functions, not just for step functions.) $\Diamond$

The previous exercise shows that the distribution function provides information about the size of f, but not about its behavior near any given point. However, it does provide sufficient information to compute L^p-norms. The L^p-norm is invariant under translation:

$$\|f\|_p = \|\tau_h f\|_p \qquad \text{for all } h \in \mathbb{R},\ f \in L^p(\mathbb{R}).$$

THEOREM 11.19. *For $f \in L^p(\mathbb{R})$, $1 \le p < \infty$ we have that*

$$\|f\|_p^p = p\int_0^\infty \lambda^{p-1} d_f(\lambda)\, d\lambda.$$

Before we present a proof, let us calculate the L^p-norm and the λ integral for the step function in Example 11.16:

$$\|f\|_p^p = 2^p \int_{-1}^{2} dx + 1\int_3^5 dx + 7^p \int_{10}^{11} dx = 2^p \times 3 + 1 \times 2 + 7^p \times 1.$$

On the other hand,

$$\begin{aligned} p\int_0^\infty \lambda^{p-1} d_f(\lambda)\, d\lambda &= 6\int_0^1 p\lambda^{p-1}\, d\lambda + 4\int_1^2 p\lambda^{p-1}\, d\lambda + \int_2^7 p\lambda^{p-1}\, d\lambda \\ &= 6\ \lambda^p\big|_0^1 + 4\ \lambda^p\big|_1^2 + \lambda^p\big|_2^7 \\ &= 6(1^p - 0^p) + 4(2^p - 1^p) + (7^p - 2^p) \ =\ \|f\|_p^p. \end{aligned}$$

EXERCISE 11.20. Verify Theorem 11.19 for step functions like the ones in Exercise 11.17, as we just did for the very specific step function in Example 11.16. $\Diamond$

PROOF OF THEOREM 11.19. By definition of d_f, we have

$$p\int_0^\infty \lambda^{p-1} d_f(\lambda)\, d\lambda = p\int_0^\infty \int_{\{x:|f(x)|>\lambda\}} \lambda^{p-1}\, dx\, d\lambda.$$

FIGURE 11.5. Picture illustrating Fubini.

By Fubini's Theorem we may interchange the order of integration, obtaining

$$\int_{\mathbb{R}} \int_0^{|f(x)|} p\lambda^{p-1}\, d\lambda\, dx = \int_{\mathbb{R}} |f(x)|^p\, dx = \|f\|_p^p. \qquad \square$$

EXERCISE 11.21. Show that for any continuously differentiable increasing function ϕ on $[0, \infty)$, we have

$$\int_{\mathbb{R}} \phi(|f|)\, dx = \int_0^\infty \phi'(\lambda) d_f(\lambda)\, d\lambda.$$

$\Diamond$

11.4. Interpolation

In this section we state some very useful theorems that go under the generic name of *interpolation theorems*. There are whole books devoted to the subject. We will present a proof in a very simplified scenario. Essentially, these theorems tell us that if we know the boundedness properties of a linear operator at two end-points, then we automatically obtain boundedness at many other points.

11.4.1. The Marcinkiewicz Interpolation Theorem.

THEOREM 11.22. *Let T be a linear operator defined on $L^p(\mathbb{R}) + L^q(\mathbb{R})$, where $1 \leq p < q \leq \infty$. Assume T is of weak-type (p,p) and (q,q): there are positive constants A_p, A_q such that for all $\lambda > 0$,*

$$d_{Tf}(\lambda) = m(\{x \in \mathbb{R} : |Tf(x)| > \lambda\}) \leq \left(\frac{A_p\|f\|_p}{\lambda}\right)^p \qquad \text{for all } f \in L^p(\mathbb{R}),$$

and

$$d_{Tf}(\lambda) = m(\{x \in \mathbb{R} : |Tf(x)| > \lambda\}) \leq \left(\frac{A_q\|f\|_q}{\lambda}\right)^q \qquad \text{for all } f \in L^q(\mathbb{R}).$$

Then T is bounded in $L^r(\mathbb{R})$ for all r such that $p < r < q$. Moreover,

$$\|Tf\|_r \leq 2\left(\frac{r}{r-p} + \frac{r}{q-r}\right)^{1/r} A_p^t A_q^{1-t}\|f\|_r,$$

where $t \in (0,1)$ satisfies $\frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}$.

We will prove a result similar in spirit to the Marcinkiewicz Interpolation Theorem in the case when T is the identity operator. The result is a statement on the nature of $L^p(\mathbb{R})$ spaces. The proof of the Marcinkiewicz interpolation theorem can be found in most books in harmonic analysis or real analysis.

Replacing the weak-type hypothesis by boundedness on $L^p(\mathbb{R})$ and $L^q(\mathbb{R})$, one can get the same result with constant one: T is bounded in $L^r(\mathbb{R})$, and

$$\|Tf\|_r \leq A_p^t A_q^{1-t}\|f\|_r.$$

This result can be obtained as a corollary of the Riesz–Thorin interpolation theorem (see Section 11.4.3), which we will state but not prove either.

PROPOSITION 11.23. *If $f \in L^p(\mathbb{R}) \cap L^q(\mathbb{R})$, $1 \leq p < q \leq \infty$, then $f \in L^r(\mathbb{R})$ for all $p < r < q$. Moreover,*

$$\|f\|_r \leq \left(\frac{r}{r-p} + \frac{r}{q-r}\right)^{1/r} \|f\|_p^t\|f\|_q^{1-t},$$

where

$$\frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}.$$

PROOF. Assume $q < \infty$. By Chebyshev's inequality (see Exercise 11.11), if $f \in L^p(\mathbb{R}) \bigcap L^q(\mathbb{R})$, then,

$$d_f(\lambda) = m(\{x \in \mathbb{R} : |f(x)| > \lambda\}) \leq \min\left\{\frac{\|f\|_p^p}{\lambda^p}, \frac{\|f\|_q^q}{\lambda^q}\right\}.$$

Let $B = \left(\frac{\|f\|_q^q}{\|f\|_p^p}\right)^{1/(q-p)}$, and note that if $0 \le \lambda \le B$, then the above minimum is $\|f\|_p^p \lambda^{-p}$, and if $B \le \lambda \le \infty$, then the minimum is $\|f\|_q^q \lambda^{-q}$. We can now estimate the L^r-norm of f. By Theorem 11.19, we have that

$$\begin{aligned}
\|f\|_r^r &= r\int_0^\infty \lambda^{r-1} d_f(\lambda)\,d\lambda \\
&\le r\int_0^\infty \lambda^{r-1} \min\left\{\frac{\|f\|_p^p}{\lambda^p}, \frac{\|f\|_q^q}{\lambda^q}\right\} d\lambda \\
&= r\int_0^B \lambda^{r-1-p}\|f\|_p^p\,d\lambda + r\int_B^\infty \lambda^{r-1-q}\|f\|_q^q\,d\lambda \\
&= \frac{r}{r-p}\|f\|_p^p B^{r-p} + \frac{r}{q-r}\|f\|_q^q B^{r-q} \\
&= \left(\frac{r}{r-p} + \frac{r}{q-r}\right)\|f\|_p^{p\frac{(q-r)}{(q-p)}}\|f\|_q^{q\frac{(r-p)}{(q-p)}}.
\end{aligned}$$

Finally, note that if $t = \frac{p}{r}\frac{(q-r)}{(q-p)}$ then $\frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}$, and $1-t = \frac{q}{r}\frac{(r-p)}{(q-p)}$.

The case when $q = \infty$ is left as an exercise for the reader. $\square$

The inequality in the proposition actually holds with constant one instead of $\left(\frac{r}{r-p} + \frac{r}{q-r}\right)$. See Exercise 11.29. We have presented this proof because the same proof goes through if we replace L^p by weak-L^p, and then the resemblance to the Marcinkiewicz Interpolation Theorem is more noticeable. All we needed in this proof was Chebyshev's inequality, which always holds for functions in $L^p(\mathbb{R})$. But we could have asked the function to satisfy weak L^p- and L^q-estimates; in other words, replace $\|f\|_p$ by the quantity

$$[f]_p := \sup_{\lambda>0} \lambda(d_f(\lambda))^{1/p},$$

and $\|f\|_q$ by $[f]_q$. We would still have concluded that if both $[f]_p$ and $[f]_q$ are finite, then $f \in L^r(\mathbb{R})$ and

$$\|f\|_r \le \left(\frac{r}{r-p} + \frac{r}{q-r}\right)^{1/r} [f]_p^t [f]_q^{1-t}, \qquad \text{where } \frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}.$$

Exercise 11.24. Verify the following two statements.

(i) If $f \in L^p(\mathbb{R}) \cap L^\infty(\mathbb{R})$ then $f \in L^r(\mathbb{R})$ for all $r > p$. Moreover,

$$\|f\|_r \le \|f\|_p^t \|f\|_\infty^{1-t},$$

where $1/r = t/p$.

(ii) If $f \in L^\infty(\mathbb{R})$ and f is in weak-$L^p(\mathbb{R})$, then $f \in L^r(\mathbb{R})$ for all $r > p$. Moreover,

$$\|f\|_r \le \left(\frac{r}{r-p}\right)^{1/r} [f]_p^t \|f\|_\infty^{1-t},$$

where $1/r = t/p$.

$\lozenge$

11.4.2. L^p-boundedness of the Hilbert transform.

COROLLARY 11.25. *H is bounded in $L^p(\mathbb{R})$ for $1 < p < \infty$.*

PROOF. We mentioned (without proof) that H is of weak-type (1,1), and H is also bounded in $L^2(\mathbb{R})$. Then the Marcinkiewicz interpolation theorem implies that H is bounded in $L^p(\mathbb{R})$ for all p with $1 < p < 2$. The *adjoint* of H is $-H$, as the following formal calculation shows:

$$\begin{aligned}
\int Hf(x)g(x)\,dx &= \int (k_H * f)(x)g(x)\,dx \\
&= \int \left(\int k_H(x-y)f(y)\,dy \right) g(x)\,dx \\
&= \iint k_H(x-y)f(y)g(x)\,dx\,dy \qquad \text{(by Fubini)} \\
&= \int f(y) \left(\int k_H(x-y)g(x)\,dx \right) dy \\
&= \int f(y)\,(-Hg(y))\,dy.
\end{aligned}$$

It turns out that if H is bounded in $L^p(\mathbb{R})$, then its adjoint $-H$ is bounded in $L^{p'}(\mathbb{R})$, where p' is the dual exponent of p, that is, $\frac{1}{p} + \frac{1}{p'} = 1$. This duality argument now takes care of the range $2 < p < \infty$. $\square$

11.4.3. The Riesz–Thorin Interpolation Theorem. Here is a variant of the interpolation theorem above, where the operator does not necessarily map a given space into itself. It is not exactly a generalization of the Marcinkiewicz interpolation theorem because it does not deal with operators of weak-type, but it allows for different L^p-spaces as inputs and outputs of the operator. The constant that appears is simpler than the one in Marcinkiewicz Interpolation Theorem; it is just one. This theorem has as almost immediate corollaries some of the most famous inequalities in analysis.

THEOREM 11.26. *Let $1 \le p_1, p_2, q_1, q_2 \le \infty$, and for $0 \le t \le 1$ define p and q by*

$$\frac{1}{p} = \frac{t}{p_1} + \frac{1-t}{p_2}, \qquad \frac{1}{q} = \frac{t}{q_1} + \frac{1-t}{q_2}.$$

Suppose T is 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

$$\|Tf\|_{q_1} \le A_1 \|f\|_{p_1} \qquad \textit{for all } f \in L^{p_1}(\mathbb{R}),$$

and

$$\|Tf\|_{q_2} \le A_2 \|f\|_{p_2} \qquad \textit{for all } f \in L^{p_2}(\mathbb{R}).$$

Then T is bounded from $L^p(\mathbb{R})$ into $L^q(\mathbb{R})$. Moreover,

$$\|Tf\|_q \le A_1^t A_2^{1-t} \|f\|_p \qquad \textit{for all } f \in L^p(\mathbb{R}).$$

The proof of Theorem 11.26 uses the so-called *three-lines theorem* for analytic functions, which can be found in many harmonic analysis textbooks. See for example [**SW**, Chapter 5], [**Graf**, Chapter 1].

FIGURE 11.6. Picture in coordinates $1/p$ and $1/q$ showing the known data and the interpolation segment.

11.5. A festival of inequalities

COROLLARY 11.27 (Hausdorff–Young Inequality). *If $f \in L^p(\mathbb{R})$, $1 \leq p \leq 2$, then $\widehat{f} \in L^q(\mathbb{R})$, for $\frac{1}{p} + \frac{1}{q} = 1$, and*

$$\|\widehat{f}\|_q \leq \|f\|_p.$$

PROOF. We will prove the Hausdorff–Young inequality by interpolating between Plancherel,

$$\|\widehat{f}\|_2 = \|f\|_2,$$

and the L^1–L^∞ boundedness discussed in Chapter 8,

$$\|\widehat{f}\|_\infty \leq \|f\|_1.$$

These are both strong estimates, and one of them involves different domain and range, so we should use the Riesz–Thorin Interpolation Theorem with exponents

$$p_1 = q_1 = 2, \quad p_2 = 1, \quad q_2 = \infty$$

and constants

$$A_1 = A_2 = 1.$$

FIGURE 11.7. Picture in coordinates $1/p$ and $1/q$ showing the known data and the segment obtained by interpolation: a piece of the secondary diagonal of the unit square.

We can see from the figure that the pair $x = 1/p$, $y = 1/q$ reached by the interpolation theorem must lie on the line $x + y = 1$, for $x \in [1/2, 1]$. Therefore p and q are indeed dual exponents:

$$\frac{1}{p} + \frac{1}{q} = 1, \qquad \text{and } 1 \leq p \leq 2.$$

Finally, the interpolation theorem gives us the desired inequality with constant $A_1^t A_2^{1-t} = 1$ for all $t \in [0, 1]$:

$$\|\widehat{f}\|_q \leq \|f\|_p,$$

for the above range of p, q, as claimed. $\square$

Corollary 11.28 (Generalized Hölder's inequality). *Given $f \in L^p(\mathbb{R})$, and $g \in L^q(\mathbb{R})$, then $fg \in L^r(\mathbb{R})$, where $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, and*

$$\|fg\|_r \leq \|f\|_p \|g\|_q.$$

Proof. Fix $1 \leq q \leq \infty$, and $g \in L^q$. First notice that if $p = \infty$, then $r = q$ and the following inequality holds:

$$\|fg\|_q \leq \|f\|_\infty \|g\|_q.$$

In fact,

$$\|fg\|_q = \left(\int_{\mathbb{R}} |f(x)|^q |g(x)|^q \, dx \right)^{1/q} \leq \left(\|f\|_\infty^q \int_{\mathbb{R}} |g(x)|^q \, dx \right)^{1/q} \leq \|f\|_\infty \|g\|_q.$$

Next remember *Hölder's inequality*, which corresponds to the case $r = 1$, $p = q'$, where $\frac{1}{q} + \frac{1}{q'} = 1$:

$$\|fg\|_1 = \int_{\mathbb{R}} |f(x)g(x)| \, dx \leq \|f\|_{q'} \|g\|_q. \tag{11.8}$$

Now apply the Riesz–Thorin interpolation theorem with $p_1 = \infty$, $q_1 = q$, $p_2 = q'$, $q_2 = 1$, and the linear operator given by multiplication by g, $Tf = fg$. Notice that this time the constants are $A_1 = A_2 = \|g\|_q$.

Figure 11.8. Picture in coordinates $1/p$ and $1/q$ showing the known data and the segment obtained by interpolation.

It can be seen from the figure that the reciprocals of the targeted exponents (x, y) solve the equation of the line of slope 1 that passes through the point $(0, 1/q)$, that is the segment $y = x + \frac{1}{q}$, $0 \leq x \leq 1/q'$. Therefore, denoting the exponents by $p = 1/x$ and $r = 1/y$, we conclude that they must satisfy

$$\frac{1}{r} = \frac{1}{p} + \frac{1}{q},$$

and the interpolation theorem tells us that

$$\|fg\|_r \leq A_1^t A_2^{1-t} \|f\|_p = \|g\|_p^t \|g\|_p^{1-t} \|f\|_p = \|f\|_p \|g\|_q.$$

The inequality has been verified. $\square$

Exercise 11.29. (*Interpolation*) Show that Proposition 11.23 holds with constant one. In other words, show that if $1 \leq p < r < q \leq \infty$, and $f \in L^p(\mathbb{R}) \cap L^q(\mathbb{R})$, then $f \in L^r(\mathbb{R})$ and

$$\|f\|_r \leq \|f\|_p^t \|f\|_q^{1-t}, \qquad \text{where } \frac{1}{r} = \frac{t}{p} + \frac{1-t}{q}.$$

Hint: If $q < \infty$, use Hölder's inequality (11.8) with dual exponents $p/(tr)$ and $q/((1-t)r)$, and $f \to |f(x)|^{tr}$, $g \to |f(x)|^{(1-t)r}$. $\Diamond$

We have already encountered Hölder's inequality (11.8). In the case $q = q' = 2$, it is the Cauchy–Schwarz inequality, and we used it in several arguments to do with the Haar basis. The general case appeared when we discussed L^p-functions as distributions, in Section 8.6.

EXERCISE 11.30. (*Finite-dimensional analogue of Hölder's Inequality*) Let a_i, $b_i \in \mathbb{R}$ for $i = 1, \ldots, n$. Prove that

$$\sum_{i=1}^{n} |a_i b_i| \leq \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |b_i|^{p'}\right)^{1/p'}, \qquad \text{where } \frac{1}{p} + \frac{1}{p'} = 1.$$

(The case $n = 2$ and $p = 2 = p'$ is a restatement of the geometric fact that $|\cos\theta| \leq 1$.) $\Diamond$

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, from Hölder's inequality. In this case the inequality carries the name of Minkowski.

LEMMA 11.31 (Minkowski's Inequality). *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 can estimate pointwise by means of the triangle inequality for complex numbers,

$$|f(x) + g(x)|^p \leq (|f(x)| + |g(x)|)|f(x) + g(x)|^{p-1},$$

and apply Hölder's inequality, observing that if $\frac{1}{p} + \frac{1}{p'} = 1$, then $(p-1)p' = p$, therefore,

$$\begin{aligned}
\int_{\mathbb{R}} |f(x) + g(x)|^p \, dx &\leq \int_{\mathbb{R}} |f(x)||f(x) + g(x)|^{p-1} \, dx + \int_{\mathbb{R}} |g(x)||f(x) + g(x)|^{p-1} \, dx \\
&\leq \left(\int_{\mathbb{R}} |f(x)|^p \, dx\right)^{1/p} \left(\int_{\mathbb{R}} |f(x) + g(x)|^p \, dx\right)^{1/p'} \\
&\qquad + \left(\int_{\mathbb{R}} |g(x)|^p \, dx\right)^{1/p} \left(\int_{\mathbb{R}} |f(x) + g(x)|^p \, dx\right)^{1/p'} \\
&= (\|f\|_p + \|g\|_p) \left(\int_{\mathbb{R}} |f(x) + g(x)|^p \, dx\right)^{1/p'}.
\end{aligned}$$

The left-hand side is equal to the right-hand side raised to the $1/p' < 1$ power. We know it is not zero, so we can divide both sides by this quantity, and get exactly what we want on the left-hand side:

$$\|f + g\|_p = \Big(\int |f(x) + g(x)|^p \, dx\Big)^{1-1/p'} \leq \|f\|_p + \|g\|_p,$$

as required. $\square$

EXERCISE 11.32. (*Finite-dimensional analogue of Minkowski's Inequality*) Let a_i, $b_i \in \mathbb{R}$, for $i = 1, \ldots, n$. Prove that

$$\left(\sum_{i=1}^{n} |a_i + b_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n} |b_i|^p\right)^{1/p}.$$

$\Diamond$

One can have n summands, or even a continuum of summands (an integral!). That is the content of the following inequality, which we mentioned in the proof of Theorem 7.38.

LEMMA 11.33 (Minkowski's Integral Inequality). *Let $F(x,y)$ be a function of two variables such that for a.e. $y \in \mathbb{R}$, the function $F_y(x) := F(x,y)$ belongs to $L^p(\mathbb{R}, dx)$, and the function $G(y) = \|F_y\|_p$ 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,*

$$\left\| \int_{\mathbb{R}} F(x,y)\, dy \right\|_{L^p(dx)} \leq \int_{\mathbb{R}} \|F(x,y)\|_{L^p(dx)}\, dy.$$

This inequality is a staple of a real analysis course. It is helpful to keep in mind that it is the triangle inequality for integrals instead of sums.

COROLLARY 11.34 (Young's Inequality). *If $f \in L^p(\mathbb{R})$, and $g \in L^q(\mathbb{R})$, then $f * g \in L^r(\mathbb{R})$, where $\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}$, and*

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

PROOF. We will prove Young's Inequality by interpolation. We can check directly that for fixed $g \in L^p(\mathbb{R})$, if $f \in L^1(\mathbb{R})$, then

$$\|f * g\|_p \leq \|g\|_p \|f\|_1,$$

and if $f \in L^{p'}(\mathbb{R})$ then

$$\|f * g\|_\infty \leq \|g\|_p \|f\|_{p'},$$

where $\frac{1}{p} + \frac{1}{p'} = 1$. Then use the Riesz–Thorin Interpolation Theorem. (The second inequality is a direct application of Hölder's inequality. The first one is an easy consequence of Minkowski's Integral Inequality. [***Work out details.]) □

We end this chapter by proving an approximation-of-the-identity theorem for L^p functions.

THEOREM 11.35. *Assume K is a positive function on $\mathbb{R}$ such that $\|K\|_1 = \int_{\mathbb{R}} K(x)\, dx = 1$. The functions $K_t(x) = t^{-1}K(t^{-1}x)$, for $t > 0$, define an approximation of the identity in $\mathbb{R}$. Then if $f \in L^p(\mathbb{R})$,*

$$\lim_{t \to 0} \|K_t * f - f\|_p = 0, \qquad \text{for } 1 \leq p < \infty.$$

PROOF. We have already proved this theorem for Schwartz functions (Theorem 7.38). We also know that $\mathcal{S}(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ (see Appendix 2.4) for each $1 \leq p < \infty$. Furthermore the operators T_t given by convolution,

$$T_t(f) = K_t * f,$$

are uniformly bounded in $L^p(\mathbb{R})$ by Young's Inequality:

$$\|T_t f\|_p = \|K_t * f\|_p \leq \|K_t\|_1 \|f\|_p = \|K\|_1 \|f\|_p = \|f\|_p.$$

Theorem 7.38 says that on $\mathcal{S}(\mathbb{R})$ the operators T_t converge as $t \to 0$ to the identity operator, which is certainly bounded on every $L^p(\mathbb{R})$. We can now apply the Uniform Boundedness Principle to conclude that for all $f \in L^p(\mathbb{R})$,

$$\|T_t f - f\|_p \to 0 \qquad \text{as } t \to 0.$$ □

Since the conclusion of Theorem 11.35 holds for functions in $L^p(\mathbb{R})$, it also holds for functions $f \in L^p(\mathbb{T})$. Here is an application. Since the Fejér and Poisson kernels on $\mathbb{T}$ are approximations of the identity, then convergence in $L^p(\mathbb{T})$ of the Cesàro and Abel means follows:

$$\|\sigma_N f - f\|_p \to 0, \qquad \text{as } N \to \infty; \tag{11.9}$$

$$\|A_r f - f\|_p \to 0, \qquad \text{as } r \to 1, \tag{11.10}$$

without any reference to the Hilbert transform. We cannot use this argument for the partial Fourier sums, because the Dirichlet kernels do not define an approximation of the identity.

Of course, we could have deduced this for continuous functions on $\mathbb{T}$ from the uniform convergence of the Cesàro and Abel means, and then used the fact that continuous functions on $\mathbb{T}$ are dense in $L^p(\mathbb{T})$ for all $1 \leq p < \infty$.

CHAPTER 12

Projects for students

In this final chapter, we present ideas for projects in harmonic analysis that students can work on rather independently. The projects explore several areas of modern harmonic analysis, using the material in our book as a springboard. We have found that such projects help to get students 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 projects vary in their level of difficulty and sophistication, though they are all more challenging than the exercises in the book. We have written them to be flexible and open-ended, and we encourage instructors to adapt them and to create their own projects. Our projects are suitable for individuals and for teams of students; in fact, more ambitious projects such as Project 12.4 on the Gibbs phenomenon lend themselves to being divided up among several teams of students. It is not necessary for students to have covered most of the book before starting on the projects. Indeed, we recommend that students choose (or are assigned to) projects early in the course. The interplay between their ongoing work on the project and the material presented in lectures helps to motivate students to follow lectures, and leads them to make constructive contributions that stem from their projects.

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. In this way, skills are developed in various types of mathematical communication, and students with differing strengths get a chance to shine. Instructors might consider allocating the last two or three weeks of lectures to student presentations, if there are few enough students to make this practicable. It may also be fruitful to set up some project milestones throughout the course.

For instance, we like to have an initial ten-minute conversation with each student (or team) about the project they plan to do, during which we discuss a few preliminary sentences they have written with some early ideas about what the project will consist of. Such a meeting enables timely intervention if the proposed scope of the project is not feasible, for instance. One can also set a series of target dates for such things 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. Finally, if time permits, students may benefit from a technical discussion or two with the instructor on the content of their project, once they have got their teeth into it.

Students may find it helpful to use a mathematical typesetting package such as 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. It may be appropriate and natural to give some formal or informal training in the use of such tools and in mathematical writing and speaking; we recommend Higham's book [**Hig**] as an excellent place to start. Instructors contemplating the task of designing and planning for semester-long student projects will find much food for thought in Bean's book [**Bea**].

We thank the following students who wrote reports and gave presentations on some of the projects: Mike Klopfer for Project 12.5, ...

12.1. Fourier analysis on finite groups

See [**Ter**].

12.2. Two extremal problems on curves

See [**Kor**, Chapters 35 and 36].

See also [**KL**, p.151].

Hurwitz 1901: Isoperimetric Problem.

Pfluger 1980: Curves of constant width.

Is there some relevant result using wavelets? How about Peter Jones' beta numbers and the Travelling Salesman Problem?

12.3. Distribution of prime numbers

See [**Kor**, Chapter 3] on Weyl's equidistribution theorem.

Is there some relevant result using wavelets?

12.4. The Gibbs phenomenon for Fourier series and for wavelets

We have seen that if a function f is just a little better than continuous, for instance (Theorem 3.12) if f is in the space $C^1(\mathbb{T})$, then the Fourier series of f converges uniformly to f on $\mathbb{T}$. What happens 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 and its analogues for certain wavelet 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 *sawtooth* (or *ramp*) function

$$f(x) = \frac{\pi}{2}\,\mathrm{sgn}(x) - \frac{x}{2}, \qquad -\pi < x < \pi,$$

and a step function $g(x)$.

Animate your plots to create a movie showing what happens as N increases.

From your results, 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 sawtooth function in part (a) does not tend to zero as $N \to \infty$. One approach is as follows. Calculate the Fourier coefficients $\widehat{f}(n)$. Show that the Fourier series for f can be rewritten in the form

$$\sum_{n=-\infty}^{\infty} \widehat{f}(n)e^{inx} = \lim_{n\to\infty} \left(\frac{1}{2} \int_0^{\infty} D_n(t)\, dt \right) - \frac{x}{2},$$

where D_n is the Dirichlet kernel. Show that the N^{th} Fourier partial sum evaluated at the point π/N has the interesting limit

$$\lim_{N\to\infty} S_N\left(\frac{\pi}{N}\right) = \int_0^{\pi} \frac{\sin u}{u}\, du =: \mathrm{Si}(\pi).$$

Look up, 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 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 [**Kor**, Chapters 8 and 17], Prestini [**Pre**], and Gottlieb and Shu [**GS**] give overviews of the history, while more detail is in Gibbs' two letters to *Nature* and Michelson's letter to the same journal, and in the papers by Wilbraham, Bôcher, and Michelson and Stratton. Search the Internet for more information. Try to find a photo of the *harmonic analyzer* machine built by Michelson, and find out what role it played in the study of the Gibbs phenomenon. Kelvin also built a harmonic analyzer. The website `http://www.sosmath.com/fourier/fourier3/gibbs.html` includes animations of partial sums.

(e) Does the Gibbs phenomenon occur for Cesàro sums, that is if we replace the Dirichlet kernel by the (good) Fejér kernel? Investigate using MATLAB. Form a conjecture, and try to prove it analytically.

What happens for other good kernels?

We saw in Chapter 4 that replacing the Dirichlet kernel by the Fejér kernel led to an improvement, in the sense of better convergence of the resulting partial sums. How does this idea relate to your results?

If any of the kernels you have considered do eliminate the Gibbs phenomenon, why don't we always use such a kernel instead of the Dirichlet kernel? Explore the literature to clarify this point.

(f) Does the Gibbs phenomenon occur for wavelet bases? For the sawtooth function f, consider the Haar projections $P_j(f)$ as $j \to \infty$. Use MATLAB to investigate.

(g) How about for the various Daubechies wavelets? Use MATLAB to investigate what happens for these and for other wavelet bases. Is there any relation between the occurrence or non-occurrence of the Gibbs phenomenon and the smoothness of the scaling function?

(h) Look at the literature to build your understanding of which wavelets have the Gibbs phenomenon. You could start with the work of Susan Kelly [**Kel**], of Hong-Tae Shim and Hans Volkmer [**SV**], and of Abdul J. Jerri.

(i) The title of Gottlieb and Shu's paper [**GS**] is *On the Gibbs phenomenon and its resolution*. In what sense do they resolve the phenomenon? Write a report explaining the content of their paper.

[*** Add refs to bibliography; add cite commands to parts (d) and (h); list in index.]

See Jorge's notes in the Park City Lecture, Summer 2003.

*** Add to bibliography: Here are two books (that I have not seen but that might have the elementary examples we are looking for):

The Gibbs Phenomenon in Fourier Analysis, Splines and Wavelet (Mathematics and Its Applications), by A. J. Jerri. Springer, 1998.

The Gibbs Phenomenon in Fourier Analysis, Splines and Wavelet Approximations by Abdul J. Jerri, Kluwer Academic Publishers, Dordrecht/Boston/London, 1998.

12.5. Linear and nonlinear approximation in Fourier and Haar bases

In Section 10.10 we introduced the ideas of linear approximation and nonlinear approximation. The *linear approximation error* $\varepsilon_l[N]$ for the N^{th} partial sum of the Fourier series of a function f is defined by

$$\varepsilon_l[N] := ||f - S_N f||_2^2 = \sum_{|n|>N} |\widehat{f}(n)|^2.$$

Now consider the same coefficients, arranged in decreasing order. We call the resulting sum the *nonlinear approximation error* $\varepsilon_n[N]$ for the N^{th} partial sum of the Fourier series of f:

$$\varepsilon_n[N] := \sum_{k\in\mathbb{N}, |n_k|>N, |\widehat{f}(n_k)|\geq|\widehat{f}(n_{k+1})|} |\widehat{f}(n_k)|^2.$$

(a) We have seen that the smoother a function is, the faster its Fourier coefficients shrink as $n \to \infty$. For example, Theorem 3.13 says that if $f \in C^k(\mathbb{T})$, then its Fourier coefficients decay at least as fast as n^{-k}: there is a constant C such that

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

Therefore there is a similar rate of decay for the linear approximation error in the Fourier basis. Specifically, show that for $f \in C^k(\mathbb{T})$,

$$\varepsilon_l[N] \leq \frac{2}{(2k-1)N^{2K-1}}.$$

(b) Consider a function that has a discontinuity, namely $\chi_{[0,2/3)}(x)$ on $(0,1)$. Show that the linear approximation error and the nonlinear approximation error in the Fourier basis are both of order N^{-1}:

$$\varepsilon_l[N] \sim \frac{1}{N} \quad \text{and} \quad \varepsilon_n[N] \sim \frac{1}{N}.$$

(c) Replacing the Fourier basis by the Haar basis, write definitions for $\varepsilon_l(N)$, the *linear approximation error in the Haar basis* for the N^{th} Haar partial sum of a function f, and for $\varepsilon_n(N)$, the *nonlinear approximation error in the Haar basis* for the N^{th} Haar partial sum of a function f. Pay attention to the fact that there are n Haar functions corresponding to intervals of length 2^{-n}. For example, the first 15 Fourier basis functions correspond to 15 different frequencies, while the first 15 Haar basis functions correspond to only 4 frequencies.

(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 constant C such that

$$\varepsilon_l(N) \geq \frac{C}{N} \qquad \text{for all } N \in \mathbb{N}.$$

This estimate is not to be confused with estimates for

$$||f - P_j f||_2^2 > \frac{C}{2^j};$$

the preceding expression is not $\varepsilon_l(j)$ but $\varepsilon_l(2^j)$.

(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) [**Mall98**, p.384].

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 $(0,1)$,

$$\varepsilon_n(N) = \frac{C}{2^N}.$$

(f) For a certain class of discontinuous functions, wavelet bases can 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. You might start with [**Mall98**, Prop 9.4].

12.6. Monsters

In this project we explore some unusual functions. Could cite Hermite and Poincaré's comments, or are they too negative? See Körner page 42.

(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 3.26.) Write an explanation of the construction. Your intended audience is another student at the same stage of their studies as you are now. Some possible sources: the original construction by du Bois-Reymond is explained in [**Kor**, Chapter 18]. Another construction, relying on the Uniform Boundedness Principle, is given in [**Duo**, Chapter 1].

(b) 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 [**Bre**, Chapter 6] works through the proof for Weierstrass's example. Körner [**Kor**, Chapter 11] shows that the same properties hold for another version of Weierstrass's example, namely

$$\sum_{n=0}^{\infty} \frac{1}{n!} \sin\big((n!)^2 x\big).$$

Körner mentions that Cellérier independently discovered the same function, but did not publish it, and that 'many years earlier Bolzano seems to have been close to a continuous nowhere differentiable function constructed along the same lines, but he too did not publish'.

Construction of an example à la Du Bois-Reymond.
Weierstrass and Riemann nowhere differentiable functions.
See [**Kor**, Chapters 11 and 18].
Quote from [**JMR**, p.3]:

> "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."

Perform comparative smoothness analysis: wavelets are more sensitive than Fourier! Design experiments to see this (see Torresani).

- Holshneider and P. Tchamitchian, *Pointwise analysis of Riemann's nowhere differentiable function.* Invent. Math.,105 (1991) pp.157-176
- S. Jaffard, *Local behavior of Riemann's function.* Contemp. Mat. 189 (1995) pp.278-307.
- S. Jaffard, *Old friends revisited: The multifractal nature of some classical functions.* J. Fourier Anal. Appl. 3 (1997) pp. 1-22.
- See Chapter 10 in the book by S. Jaffard, Y. Meyer, R. D. Ryan, *Wavelets: Tools for Science and technology.* SIAM, 2001 [JMR]
- See also [**KL**, Section 4, p. 382]

12.7. Local cosine and sine bases

Borrow from description in Notes by Cris and Martin (2004).

Malvar-Wilson wavelets (Kenneth Wilson (Nobel Prize in Physics in 1982). See [**JMR**, Chapter 6].

12.8. The twin dragon and fractals in wavelets

Could be paired with a general Cantor sets project, or made into one project (I like this idea, it justifies the appearance of the Cantor 2/3 set in the text, but here we present it in the context of wavelets which is one of the topics in the book!).

Just building the twin-dragon set from mathematical analogy to Haar basis in 2-D to writing the software is a very nice project. If we understand how this is done, then variations leading to other fractal sect can be part of the project!

Images and information can be found at:

- `http://www.math.umd.edu/$\sim$jjb/2d-wavelet.jpg`
- Jelena Kovacevic has an applet that I cannot view, but that is another reference

  `http://lcavwww.epfl.ch/$\sim$minhdo/wavelet$_-$course/applets/TwinDragon/index.html`

- She also has a link to the math background:

  `http://lcavwww.epfl.ch/$\sim$minhdo/wavelet$_-$course/applets/TwinDragon/math.html`

- I'll bet in her book we can find the details.
  Jelena Kovacevic and Martin Veterli, *Wavelets and subband coding*, Prentice Hall, Signal Processing Series, Englewoods cliffs, NJ, 1995.
- The above I found linked from:

  `http://log24.com/log03/0827.htm`

- Another source is an article that appeared in the Notices:
  Matrix factorizations, algorithms, wavelets, by Palle E. T. Jorgenson. Notices of the AMS, Volume 50, Number 8, p. 880-894.
- Benedetto and his students Manuel León and Songkid Sumetkijatka have results:

  `http://www.math.umd.edu/$\sim$jjb/fractal.jpg`

- Also see the work of Ursula Molter, Carlos Cabrelli, Christopher Heil, Lars Villemois, etc. (these are probably too advanced, but worth keeping in mind when pointing to more advanced literature.)

12.9. Fourier analysis and wavelets on graphs

Fourier analysis on graphs. See [**Chu**]. [*** Cris to find details for this reference.]

Here present a toy model graph and ask the students to find the Fourier basis on the graph.

Diffusion wavelets
Using the same toy model find a diffusion wavelet basis.

- R. Coifman, M. Maggioni, *Diffusion wavelets.* Submitted to Elsevier Science 2005. Preprint found online in Maggioni's webpage.
- See Hamilton Link's Ph D Thesis 2007 or 2008 (he is the one who introduced me (MCP) to this!

12.10. The discrete Hilbert transform and Cotlar's Lemma

As in Bernadette Mendoza-Spencer's MS Thesis, Spring 2006.

- Define DHT. Find a periodic function whose Fourier coefficients are $1/n$ for $n \neq 0$. Deduce boundedness in ℓ^2.
- Introduce Cotlar's lemma, and sketch a proof in the ℓ^2 case. Almost-orthogonalization technique. As an application, use it to obtain a complicated proof of boundedness in ℓ^2 of the discrete Hilbert transform. Explain matrix setting and meaning at that linear algebra level, on how to find the right candidates for this decomposition.
- Perhaps come up with other infinite matrices and strategies for effective use of Cotlar's Lemma?

12.11. The Hilbert transform as an average of Haar multipliers

Read Stefanie's Comptes Rendus paper.

S. Petermichl, *Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol.* C. R. Acad. Sci. Paris Sr. I Math. 330 (2000) #6, 455-460.

12.12. The Calderón-Zygmund decomposition

An example of an stopping-time argument, and how to use it to show that the Hilbert transform is of weak-type (1,1).

12.13. Non-standard representation of SIO

See Beylkin, Coifman,and Rokhlin's original paper.

- G. Beylkin, *On the representation of operators in bases of compactly supported wavelets.* SIAM J. Numer. Anal. Vol 6, No. 6, pp. 1716–1740 (1992).
- G. Beylkin, R. Coifman, V. Rokhlin, *Fast wavelet transform and numerical algorithms.* Comm. Pure Appl. Math., **44**, pp. 141–183 (1991).
- See [**KL**, Section 5, p.360], and also I (MCP) have lecture notes by I. Daubechies where this is explained.
- Dig out Paul Hubbard's MS Thesis 1999.

Include MATLAB experiments to see standard and non-standard matrix representations.

12.14. The $T1$ Theorem proved via Haar

- See paper by Coifman, Jones, and Semmmes, *Two elementary proofs of the L^2 boundedness of Cauchy integrals on Lipschitz curves.* J. of the AMS **2** (1989), 553-564.
- See Chapter 4 of the review paper by the first author. Maria Cristina Pereyra, *Lecture notes on dyadic harmonic analysis.* Contemp. Math. **289** (2001) p. 1-60.
- See [**KL**, Section 3, p.378].

Include how these estimates connect with the fast algorithm in the NS decomposition of BCR.

APPENDIX A

Some useful facts from analysis

An appendix with a list of useful facts from analysis goes here. State them without proof. Include references with index entry, e.g. [**Rud**]. Mostly results about uniform continuity and uniform convergence. Include a discussion, à la Bressoud, of why we care: these are the results that say when we can freely exchange the sum and the integral, or the limit and the integral, etc. (*** Much of this is now in the density section in the Interlude: Chapter 2.)

Results we need: (1) In proof of convolution properties, need to exchange dx and dy integrals over $\mathbb{T}$: ok for cts fns. (2) In same proof, need $\lim \int = \int \lim$ if integrands $f_k \to f$ uniformly. (3) Fubini's theorem; we use it in the proof of the Fourier Inversion Formula for $\mathcal{S}$ in Section 7.2.]

In a couple of places Chapter 8 mentions the fact that $\mathcal{S}(\mathbb{R})$ is dense in L^p, and maybe in other spaces as well, referring to this appendix. If we keep that structure, we should state/prove those density results here.

APPENDIX B

Vector spaces, norms, inner products

Here we record the definitions and the main properties of vector spaces, inner-product vector spaces, normed spaces, Banach and Hilbert spaces.

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. We also discuss orthonormal bases and orthogonal projections onto closed subspaces.

B.1. Vector spaces

We begin with the definition of a vector space.

Definition B.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) $x + y \in V$, $\lambda x \in V$. (*V is closed under addition and scalar multiplication.*)
(ii) $x + y = y + x$, $x + (y + z) = (x + y) + z$. (*Addition is commutative and associative.*)
(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) $\lambda(x + y) = \lambda x + \lambda y$. (*Addition is distributive w.r.t. scalar multiplication.*)
(vi) $\lambda_1(\lambda_2 x) = (\lambda_1 \lambda_2)x$. (*Scalar multiplication is associative.*)

$\Diamond$

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

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

Example B.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 2.4. $\Diamond$

Example B.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 Lebesgue spaces $L^p(\mathbb{R})$;

the collection $\mathcal{S}(\mathbb{R})$ of Schwartz functions; infinitely often differentiable and rapidly decaying functions all of whose derivatives also are rapidly decaying; see Section 7.2 for precise definitions. $\Diamond$

Example B.5. The space $\ell^2(\mathbb{Z})$ of sequences defined by equation (5.1) in Section 5.2, with componentwise addition and scalar multiplication. See also Example B.12. $\Diamond$

B.2. Normed spaces

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

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

(i) $\|x\| \geq 0$,
(ii) $\|x\| = 0$ if and only if $x = 0$,
(iii) $\|\alpha x\| = |\alpha| \, \|x\|$,
(iv) It satisfies the *triangle inequality*

$$\|x + y\| \leq \|x\| + \|y\|. \tag{B.1}$$

$\Diamond$

Here are some examples of normed spaces.

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

Example B.8. The set $\mathbb{R}^n$ with norm the ℓ^p-norm, defined for each $x = (x_1, x_2, \ldots, x_n)$, and for each $1 \leq p < \infty$, by

$$\|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p}.$$

$\Diamond$

In finite-dimensional spaces *all norms are equivalent.* In particular, for $\mathbb{R}^n$ and the ℓ^p-norms there are positive constants A, $B > 0$ depending only on p and q such that

$$A\|x\|_p \leq \|x\|_q \leq B\|x\|_p.$$

The situation is different for infinite-dimensional spaces.

Example B.9. The continuous functions over an interval I, or over $\mathbb{R}$, with the uniform norm. $\Diamond$

Example B.10. The continuous functions over a closed interval I, with the L^p-norm. $\Diamond$

B.3. Inner-product vector spaces

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

Definition B.11. 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) $\langle x, y\rangle = \overline{\langle y, x\rangle}$. (*The inner product is* 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) $\langle x, x\rangle = 0 \Longleftrightarrow x = 0$. ((iii) *&* (iv) *together: strictly positive-definite.*)

A vector space with such an inner product is called an *inner-product vector space.* ◊

Notice that (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 B.20. This norm is known as the *induced norm* on V.

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

Example B.12. $V = \mathbb{C}^d$. The elements of $\mathbb{C}^d$ look like $z = (z_1, z_2, \ldots, z_n)$, with $z_j \in \mathbb{C}$. The inner product and its induced norm are defined by

$$\begin{aligned} \langle z, w\rangle &:= z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n}, \\ \|z\|_{\mathbb{C}^n} &:= \langle z, z\rangle^{1/2} = (|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2)^{1/2}. \end{aligned}$$

◊

Example B.13. (*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 $\{a_n\}_{n\in\mathbb{Z}} = (\ldots, a_{-n}, \ldots, a_{-1}, a_0, a_1, \ldots, a_n, \ldots)$, 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 ℓ^2 by

$$\begin{aligned} \langle x, y\rangle &:= \sum_{n=-\infty}^{\infty} a_n\overline{b_n}, \\ \|x\|_{\ell^2(\mathbb{Z})} &:= \langle x, x\rangle^{1/2} = \left(\sum_{n=-\infty}^{\infty} |a_n|^2\right)^{1/2}. \end{aligned}$$

◊

Example B.14. $V = L^2(\mathbb{T})$. The inner product and its induced norm are defined by

$$\begin{aligned} \langle f, g\rangle &:= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}\, dx, \\ \|f\|_{L^2(\mathbb{T})} &:= \langle f, f\rangle^{1/2} = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2\, dx\right)^{1/2}. \end{aligned}$$

◊

The fact that the infinite sums in Example B.13 and the integrals in Example B.14 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 B.3.1.

Had we considered $\mathcal{R}(\mathbb{T})$, the set of Riemann-integrable functions on $\mathbb{T}$, with the L^2-inner product and norm, we would not have encountered the difficulty with the integrals, since functions in $\mathcal{R}(\mathbb{T})$ are bounded, and so the integrals are automatically finite. However we would still have had the problem with the infinite sums, since boundedness, or even decay to zero, do not guarantee the convergence of the series.

B.3.1. Three key results involving 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 very 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 B.15. Two 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 $y \in Y$. ◊

In Chapter 5 we show that the trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are *orthonormal* in $L^2(\mathbb{T})$. In other words, not only are they an orthogonal set, but each element in the set has norm one (the vectors have been *normalized*). Checking orthonormality is equivalent to verifying that

$$\langle e_n, e_m\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{in\theta}e^{-im\theta}\,d\theta = \begin{cases} 1, & \text{if } k = m;\\ 0, & \text{if } k \neq m.\end{cases}$$

EXERCISE B.16. Define vectors $e_k \in \mathbb{C}^N$ by $e_k(n) = (1/\sqrt{N})\,e^{2\pi ikn/N}$, for k, $n \in \{0, 1, 2, \ldots, N-1\}$. Show that the vectors $\{e_k\}_{k=0}^{N-1}$ form an orthonormal family in $\mathbb{C}^N$. ◊

The Pythagorean Theorem, the Cauchy–Schwarz Inequality, and the triangle inequality are well known in the two-dimensional case described above. They also hold in every inner-product vector space V over $\mathbb{C}$.

THEOREM B.17 (Pythagorean Theorem). *If x, $y \in V$ are orthogonal, then*

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

THEOREM B.18 (Cauchy–Schwarz Inequality). *For all x, y in V,*

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

THEOREM B.19 (Triangle Inequality). *For all x, $y \in V$,*

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

EXERCISE B.20. Prove Theorems B.17, B.18, and B.19, first for the model case $V = \mathbb{R}^2$, and then for a general inner-product vector space V over $\mathbb{C}$. In particular, show that $\|x\| := \sqrt{\langle x, x\rangle}$ is a norm in V. ◊

Exercise B.21. 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 B.22. 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),

$$a_1 x_{\lambda_1} + \cdots + a_n x_{\lambda_n} = 0 \quad \text{if and only if} \quad a_1 = \cdots = a_n = 0.$$

♢

B.3.2. Orthonormal bases. Orthogonality implies *linear independence* (see Exercise B.22), and if our geometric intuition is right, orthogonality is in some sense the most linearly independent a set could be. In Exercise B.16 we found 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 will call it the *N-dimensional Fourier basis*. In Chapter 6 we point out the analogies between this finite-dimensional theory, which only requires linear algebra facts, and the discrete Fourier theory.

The trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ are also an orthonormal basis of $L^2(\mathbb{T})$, but it takes some time to verify this fact. See Chapter 5. Having an infinite orthonormal family (hence a linearly independent family) tells us that the space is infinite-dimensional. So far, 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 B.23 (Complete system/orthonormal basis). Let $\{x_n\}_{n\in\mathbb{N}}$ be an orthonormal family 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 $\{a_n\}_{n\in\mathbb{N}}$ such that

$$\lim_{N\to\infty} \Big\| x - \sum_{n=1}^{N} a_n\, x_n \Big\| = 0.$$

Equivalently, $x = \sum_{n=1}^{\infty} a_n x_n$ where equality holds in the norm of V.

The coefficients a_n of x are uniquely determined. They can be calculated by pairing the vector with the basis elements:

$$a_n = \langle x, x_n\rangle.$$

♢

How can we tell whether a given orthonormal family X in a vector space V is actually an orthonormal basis for V? One criterion for completeness is that the only vector orthogonal to all the vectors in X is the zero vector. Another is that Plancherel's Identity must hold. These ideas are summarized in Theorem B.24, which holds for all inner-product vector spaces V over $\mathbb{C}$. In Chapter 5 we prove Theorem B.24 for the particular case of $V = L^2(\mathbb{T})$ with the trigonometric basis. Showing that the same proofs extend to the general case is left as an exercise.

THEOREM B.24. *Let V be an inner-product vector space over $\mathbb{C}$ and $X = \{x_n\}_{n\in\mathbb{N}}$ an orthonormal family in V. Then the following are equivalent:*

(i) *X is a complete system, hence an orthonormal basis of V.*
(ii) *If $x \perp X$ then $x = 0$.*
(iii) (Plancherel's Identity, or Infinite-Dimensional Pythagorean Theorem) *For all $x \in V$,*

$$\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, x_n\rangle|^2.$$

B.3.3. Banach and Hilbert spaces. Some, but not all, inner-product vector spaces satisfy another property, to do with convergence of Cauchy sequences of vectors. Namely, they are *complete inner-product vector spaces*, also known as *Hilbert spaces*. For example, $L^2(\mathbb{T})$ is a Hilbert space, but $\mathcal{R}(\mathbb{T})$ with the L^2-inner product is not a Hilbert space.

First note that since we have the norm induced by the inner product, we can certainly talk about sequences of vectors converging to another vector. Specifically if V is a vector space with a norm (a *normed space*), then we say a sequence $\{x_n\} \in V$ *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 B.25. The sequence $\{x_n\} \in V$ is *Cauchy* if for every $\varepsilon > 0$ there exists $N > 0$ such that $\|x_n - x_m\| < \varepsilon$ for all n, $m > N$. $\lozenge$

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 $\mathbb{Q}$; 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.

DEFINITION B.26. A *Banach space* is a complete normed space $\mathcal{B}$. In other words, every sequence in $\mathcal{B}$ that is Cauchy with respect to the norm of $\mathcal{B}$ converges to a limit in $\mathcal{B}$. $\lozenge$

EXAMPLE B.27. 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 [**Tao2**, Theorem 14.4.5]. $\lozenge$

EXAMPLE B.28. The spaces $L^p(I)$ defined in Chapter 2 are Banach spaces. The proof of this fact belongs to a more advanced course. However it is crucial for us, and we assume this result without proof. $\lozenge$

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

DEFINITION B.29. 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 $\mathcal{H}$. $\lozenge$

Notice that here we are using the word "complete" with a different meaning than that in Definition B.23, where we consider "complete systems of orthonormal functions". The context indicates whether we are talking about a "complete space" or a "complete system $\{f_n\}$ of functions".

Here are some canonical examples of Hilbert spaces.

Example B.30. C^n, a finite-dimensional Hilbert space. $\Diamond$

Example B.31. $\ell^2(\mathbb{Z})$, an infinite-dimensional Hilbert space. $\Diamond$

Example B.32. $L^2(\mathbb{T})$, the Lebesgue square-integrable functions on $\mathbb{T}$, consisting of the collection $\mathcal{R}(\mathbb{T})$ of all Riemann square-integrable functions on $\mathbb{T}$, together with all the functions that arise as limits of Cauchy sequences in $\mathcal{R}(\mathbb{T})$ (similarly to the construction of the real numbers from the rational numbers). In other words, $L^2(\mathbb{T})$ is the *completion* of $\mathcal{R}(\mathbb{T})$ with respect to the L^2-metric. $\Diamond$

Example B.33. $L^2(\mathbb{R})$, the Lebesgue square-integrable functions on $\mathbb{R}$. $\Diamond$

Bibliography

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

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

[Bart1] R. G. Bartle, *The elements of integration and Lebesgue measure*, John Wiley & Sons, Inc., 1966.

[Bart2] R. G. Bartle, *A modern theory of integration*, Graduate Studies in Mathematics, vol. 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.

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

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

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

[Chu] F. Chung, **** Fill in details: for project on Fourier analysis on graphs*

[CM] 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.

[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, vol. 61, SIAM, Philadelphia, PA, 1992.

[Duo] J. Duoandikoetxea, *Fourier analysis*, English transl. by D. Cruz-Uribe, SFO, Graduate Studies in Mathematics, vol. 29, Amer. Math. Soc., Providence, RI, 2001.

[DM] H. Dym and H. P. McKean, *Fourier series and integrals*, Probability and Mathematical Statistics, vol. 14, Academic Press, New York–London, 1972.

[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.

[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.

[Graf] L. Grafakos, *Classical and modern Fourier analysis*, Prentice Hall, 2004.

[Grat] I. Grattan-Guinness, *The development of the foundations of mathematical analysis from Euler to Riemann*, MIT Press, Cambridge, MA, 1970.

[Haa] A. Haar, *Zur Theorie der orthogonalen Funktionensysteme*, Math. Annal. **69** (1910), 331–371.

[HW] E. Hernandez and G. Weiss, *A first course on wavelets*, CRC Press, Boca Raton, FL, 1996.

[Hig] N. J. Higham, *Handbook of Writing for the Mathematical Sciences*, SIAM, Philadelphia, PA, 1998.

[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, pp. 235–255.

[JMR] S. Jaffard, Y. Meyer, and R. D. Ryan, *Wavelets: Tools for science and technology*, revised edition, SIAM, Philadelphia, PA, 2001.

[KK] J.-P. Kahane and Y. Katznelson, *Sur les ensembles de divergence des séries trigonométriques*, Studia Math. **26** (1966), 305–306.

[KL] J.-P. Kahane and P.-G. Lemarié-Rieusset, *Fourier series and wavelets*, Gordon and Breach Publishers, 1995.

[Kat] Y. Katznelson, *Introduction to harmonic analysis*, Dover Publications Inc., 1976.

[Kel] S. E. Kelly, *Gibbs phenomenon for wavelets*, Applied and Computational Harmonic Analysis **3** (1996), 72–81.

[Kol] A. Kolmogorov, *Une série de Fourier–Lebesgue divergente partout*, C. R. Acad. Sci. Paris **183** (1926), 1327–1328.

[Kor] T. W. Körner, *Fourier analysis*, Cambridge Univ. Press, Cambridge, 1988.

[Kra] S. G. Krantz, *A panorama of harmonic analysis*, The Carus Mathematical Monographs, vol. 27, Amer. Math. Soc., Providence, RI, 1999.

[Mall89] S. Mallat, *Multiresolution approximations and wavelet orthonormal bases for* $L^2(\mathbb{R})$, Trans. Amer. Math. Soc. **315** (1989), 69–87.

[Mall98] S. Mallat, *A wavelet tour of signal processing*, Academic Press, San Diego, CA, 1998.

[Malv] H. Malvar, *Lapped transforms for efficient transform/subband coding*, IEEE Trans. Acoust. Speech Signal Process **38** (1990), 969–978.

[MP] M. Mohlenkamp and M. C. Pereyra, *Wavelets, Their Friends, and What They Can Do for You*, EMS Series of Lectures in Mathematics, European Mathematical Society Publishing House, Zurich, to appear.

[Mun] J. R. Munkres, *Topology: A first course*, Prentice Hall, Englewood Cliffs, NJ, 1975.

[Pin] M. A. Pinsky, *Introduction to Fourier analysis and wavelets*, The Brooks/Cole Series in Advanced Mathematics (Paul J. Sally, Jr, ed.), Brooks/Cole, 2001.

[Pre] E. Prestini, *The evolution of applied harmonic analysis: Models of the real world*, Birkhäuser, Boston, MA, 2004.

[RS] M. Reed and B. Simon, *Methods of modern mathematical physics, Vol. II: Fourier analysis, self-adjointness*, Academic Press, 1975.

[Roy] H. L. Royden, *Real analysis*, 2nd ed., MacMillan, New York, NY, 1968.

[Rud] W. Rudin, *Principles of mathematical analysis*, McGraw–Hill, New York, NY, 1976.

[Sch] M. Schechter, *Principles of functional analysis*, Graduate Studies in Mathematics, vol. 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.

[SV] H.-T. Shim and H. Volkmer, *On the Gibbs phenomenon for wavelet expansions*, Journal of Approximation Theory **84** (1996), 74–95.

[Spi] M. Spiegel, *Schaum's outline of Fourier analysis with applications to boundary value problems*, McGraw–Hill, New York, NY, 1974.

[SS1] E. M. Stein and R. Shakarchi, *Fourier analysis: an introduction*, Princeton Lectures in Analysis I, Princeton Univ. Press, Princeton, NJ, 2003.

[SS2] 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.

[SW] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton Univ. Press, Princeton, NJ, 1971.

[SN] G. Strang and T. Nguyen, *Wavelets and filter banks*, Wellesley–Cambridge Press, Wellesley, MA, 1996.

[Str] R. S. Strichartz, *The way of analysis*, Jones & Bartlett Publishers; revised edition, 2000.

[Str2] R. S. Strichartz, *A guide to distribution theory and Fourier transforms*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1994.

[Tao1] T. Tao, *Analysis I*, Text and Readings in Mathematics, vol. 37, Hindustan Book Agency, India, 2006.

[Tao2] T. Tao, *Analysis II*, Text and Readings in Mathematics, vol. 37, Hindustan Book Agency, India, 2006.

[Ter] A. Terras, *Fourier analysis on finite groups and applications*, London Math. Soc. Student Texts, vol. 43, Cambridge Univ. Press, Cambridge, 1999.

[VK] M. Vetterli and J. Kovačević, *Wavelets and subband coding*, Prentice Hall, Upper Saddle River, NJ, 1995.

[Wic] V. Wickerhauser, *Adapted wavelet analysis from theory to software*, A. K. Peters, Wellesley, 1994.

[Woj] P. Wojtaszczyk, *A mathematical introduction to wavelets*, London Math. Soc. Student Texts, vol. 37, Cambridge Univ. Press, Cambridge, 2003.