Chapter 9

From Fourier to wavelets, emphasizing Haar

There are many things we would like to record, compress, transmit, or reconstruct: for instance, audio signals such as music or speech, video signals such as TV or still photos, and data in the form of words or numbers. Harmonic analysis provides the theoretical tools to do all these tasks effectively. Nowadays, in the toolbox of every engineer, statistician, applied mathematician, or other practitioner interested in signal or image processing, you will find not only classical Fourier analysis tools but also modern wavelet analysis tools.

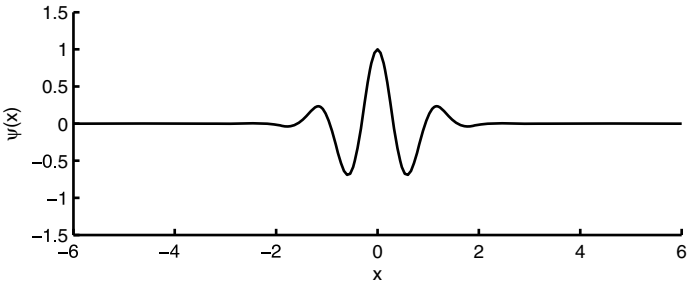

Figure 9.1. A Morlet wavelet given by $\psi(x) = e^{-t^2/2}\cos(5t)$.

Now, what is a wavelet? It is a *little wave that starts and stops*, such as the Morlet[1] wavelet shown in Figure 9.1. It differs from sines and cosines, which go on forever, and from truncated sines and cosines, which go on for the whole length of the truncation window.

The key idea of wavelets is to express functions or signals as sums of these little waves and of their translations and dilations. Wavelets play the rôle here that sines and cosines do in Fourier analysis. They can be more efficient, especially if the signal lasts only a finite time or behaves differently in different time periods. A third type of analysis, the windowed Fourier transform, unites some features of Fourier analysis with the localization properties of wavelets. Each of these three types of analysis has its pros and cons.

In this chapter we survey the windowed Fourier transform and its generalization known as the Gabor transform (Section 9.2) and introduce the newest member of the family, the wavelet transform (Section 9.3). We develop the first and most important example of a wavelet basis, the Haar basis (Section 9.4). We show that the Haar system is a complete orthonormal system in $L^2(\mathbb{R})$. We question the plausibility of this statement, and then we prove the statement. To do so, we introduce dyadic intervals, expectation (averaging) operators, and difference operators. We compare Fourier and Haar analysis and touch on some dyadic operators, whose boundedness properties hold the key to the unconditionality of the Haar basis in $L^p(\mathbb{R})$ (Section 9.5).

9.1. Strang's symphony analogy

Let us begin with the central idea of wavelet theory, through a musical analogy developed by Gilbert Strang; the quotations in this section are from his article [**Stra94**].

Imagine we are listening to an orchestra playing a symphony. It has a rich, complex, changing sound that involves dynamics, tempi, pitches, and so on. How do we write it down? How can we notate this sound experience? Musical notation is one way. Recording on

[1] Named after French geophysicist Jean Morlet (1931–2007), who coined the term *wavelet* to describe the functions he used. In 1981, Morlet and Croatian-French physicist Alex Grossman (born 1930) developed what we now call the wavelet transform.

CD is another. These two methods do not contain exactly the same information. We are expressing the symphony in different ways.

We consider three further ways of analyzing a symphony: in terms of (1) cosine waves (traditional Fourier series); (2) pieces of cosine waves (windowed Fourier series); and (3) wavelets.

(1) *Traditional Fourier analysis* "separate[s] the whole symphony into pure harmonics. Each instrument plays one steady note", with a specific loudness; for example the B flat above middle C at a given volume, or the F sharp below middle C at some other volume. The signal, meaning the whole symphony, becomes a sum $a_0 + a_1 \cos t + a_2 \cos 2t + \cdots$ of cosine waves. To store the signal, we need to record only the list of *amplitudes* $(a_0,\ a_1,\ a_2,\ \dots)$ and corresponding *frequencies* $(0,\ 1,\ 2,\ \dots)$.

In practice we cannot retain all these terms, but we may need to keep many of them to get a high-fidelity reproduction. Also, if the symphony has been playing forever and continues forever, then we need a continuum of frequencies (uncountably many musicians), and instead of a sum we get an integral, namely the Fourier transform.

(2) *Windowed Fourier analysis*, also known as the short-time Fourier transform, looks at short segments of the symphony individually. "In each segment, the signal is separated into cosine waves as before. The musicians still play one note each, but they change amplitude in each segment. This is the way most long signals are carved up." Now we need to record lists of amplitudes and frequencies *for each segment*.

"One disadvantage: There are sudden breaks between segments." In an audio signal, we might be able to hear these breaks, while in a visual image, we might see an edge. This kind of artificial discontinuity is called a *blocking artifact*. An example is shown in the JPEG fingerprint image in Figure 10.3.

(3) In *wavelet analysis*, "[i]nstead of cosine waves that go on forever or get chopped off, the new building blocks are '*wavelets*'. These are little waves that start and stop", like the Morlet wavelet in Figure 9.1. In the symphony analogy "[t]hey all come from one basic wavelet $w(t)$,

which gives the sound level of the standard tune at time t." In practice, the Daubechies wavelet shown in Figure 9.4 is a typical standard tune.

We make up the symphony from different versions of this tune, played with specific amplitudes and starting times and at specific speeds. The *double basses* play the standard tune $w(t)$. "The *cellos* play the same tune but in half the time, at doubled frequency. Mathematically, this speed-up replaces the time variable t by $2t$. The first bass plays $b_1 w(t)$ and the first cello plays $c_1 w(2t)$, both starting at time $t = 0$. The next bass and cello play $b_2 w(t-1)$ and $c_2 w(2t-1)$, with amplitudes b_2 and c_2. The bass starts when $t - 1 = 0$, at time $t = 1$. The cello starts earlier, when $2t - 1 = 0$ and $t = 1/2$. There are twice as many cellos as basses, to fill the length of the symphony. Violas and violins play the same passage but faster and faster and all overlapping. At every level the frequency is doubled (up one octave) and there is new richness of detail. 'Hyperviolins' are playing at 16 and 32 and 64 times the bass frequency, probably with very small amplitudes. Somehow these wavelets add up to a complete symphony."

It is as if each possible signal (symphony) could be achieved by an orchestra where each musician plays only the song "Happy Birthday" but at his or her own speed and start time and with his or her own volume chosen according to the signal required. To store the signal, we need to record only the list of amplitudes, in order. At every level the frequency is doubled, which means that the pitch goes up by an octave. The rôle played by the cosines and sines in Fourier analysis is played here by *dilates and translates of a single function* $w(t)$. Here we see the fundamental idea of wavelets.

For a delightful introduction to these ideas for a broad audience, see Barbara Burke Hubbard's book [**Burke**].

9.2. The windowed Fourier and Gabor bases

We construct an orthonormal basis on the line by pasting together copies of the trigonometric basis on intervals (*windows*) that partition the line. Generalizing this idea, we obtain the Gabor bases.

9.2.1. The windowed Fourier transform. The continuous Fourier transform gives us a tool for analyzing functions that are defined on the whole real line $\mathbb{R}$, rather than on the circle $\mathbb{T}$. However, the trigonometric functions $\{e^{2\pi i\xi x}\}_{\xi\in\mathbb{R}}$ no longer form a countable basis, since there is one for each real number ξ. Also, for each fixed $\xi \in \mathbb{R}$, the function $e^{2\pi i\xi x}$ is not in $L^2(\mathbb{R})$. The windowed Fourier transform addresses the problem of finding a basis for $L^2(\mathbb{R})$ to fill the rôle played by the trigonometric basis for $L^2(\mathbb{T})$.

How can we obtain an *orthonormal basis* for $L^2(\mathbb{R})$? A simple solution is to split the line into unit segments $[k, k+1)$ indexed by $k \in \mathbb{Z}$ and on each segment use the periodic Fourier basis for that segment.

Theorem 9.1. *The functions*

$$g_{n,k}(x) = e^{2\pi inx}\chi_{[k,k+1)}(x) \qquad \text{for } n, k \in \mathbb{Z}$$

form an orthonormal basis for $L^2(\mathbb{R})$, where $\chi_{[k,k+1)}$ is the characteristic function of the interval $[k, k+1)$. The corresponding reconstruction formula *holds, with equality in the L^2 sense:*

$$f(x) = \sum_{n,k\in\mathbb{Z}} \langle f, g_{n,k}\rangle g_{n,k}(x).$$

Exercise 9.2. Prove Theorem 9.1. **Hint:** Use Lemma A.51. ◊

Definition 9.3. The *windowed Fourier transform* is the map G that assigns to each function in $L^2(\mathbb{R})$ the sequence of coefficients with respect to the windowed Fourier basis $\{g_{n,k}\}_{j,k\in\mathbb{Z}}$ defined in Theorem 9.1. More precisely, $G : L^2(\mathbb{R}) \to \ell^2(\mathbb{Z}^2)$ is defined by

$$Gf(n,k) := \langle f, g_{n,k}\rangle = \int_k^{k+1} f(x)e^{-2\pi inx}\,dx.$$

Here our signal f is a function of the continuous variable $x \in \mathbb{R}$, while its windowed Fourier transform Gf is a function of two discrete (integer) variables, n and k. ◊

We can think of each characteristic function $\chi_{[k,k+1)}(x)$ as giving us a window through which to view the behavior of f on the interval $[k, k+1)$. Both the function $\chi_{[k,k+1)}(x)$ and the associated interval $[k, k+1)$ are commonly called *windows*. These windows all have unit

length and they are integer translates of the fixed window $[0,1)$, so that $\chi_{[k,k+1)}(x) = \chi_{[0,1)}(x-k)$.

Instead, we could use windows of varying sizes. Given an arbitrary partition $\{a_k\}_{k\in\mathbb{Z}}$ of $\mathbb{R}$ into bounded intervals $[a_k, a_{k+1})$, $k \in \mathbb{Z}$, let $L_k = a_{k+1} - a_k$, and on each window $[a_k, a_{k+1})$ use the corresponding L_k Fourier basis. Then the functions

$$(1/\sqrt{L_k})e^{-2\pi inx/L_k}\chi_{[a_k,a_{k+1})}(x) \qquad \text{for } n,k \in \mathbb{Z}$$

form an orthonormal basis of $L^2(\mathbb{R})$. This generalization lets us adapt the basis to the function to be analyzed. For instance, if the behavior of f changes a lot in some region of $\mathbb{R}$, we may want to use many small windows there, while wherever the function does not fluctuate much, we may use wider windows.

We hope to get a fairly accurate reconstruction of f while retaining only a few coefficients. On the wider windows, a few low frequencies should contain most of the information. On the smaller windows, retaining a few big coefficients may not be very accurate, but that may not matter much if the windows are small and few in number. Alternatively one may have many small windows, on each of which few coefficients are retained.

One seeks a balance between the chosen partition of $\mathbb{R}$ and the number of significant coefficients to keep per window. This preprocessing may require extra information about the function and/or extra computations which may or may not be affordable for a specific application. On the other hand, by adapting the windows to the function, we lose the translation structure that arises from having all the windows the same size.

Exercise 9.4 (*The Gibbs Phenomenon for the Windowed Fourier Transform*)**.** Another problem arises from the discontinuity of the windows at the endpoints: the Gibbs phenomenon (see also the project in Section 3.4). The *hat function* is defined to be $f(x) := 1 - |x|$ if $-1 \le x < 1$ and $f(x) := 0$ otherwise. Compute the windowed Fourier transform of the hat function using windows on the intervals $[k, k+1)$, for $k \in \mathbb{Z}$, and plot using MATLAB. You should see corners at $x = -1$, 0, 1. Do the same with windows $[2k-1, 2k+1)$. Do you see any corners? How about with windows $[k/2, (k+1)/2)$? ◊

9.2.2. Gabor bases. In numerical calculations, the *sharp windows* $\chi_{[k,k+1)}(x)$ used in the windowed Fourier transform produce at the edges the same *artifacts* that are seen when analyzing periodic functions at discontinuity points (the Gibbs phenomenon, or *corners* at the divisions between windows). So smoother windows are desirable. The sharp windows given by $\chi_{[0,1)}(x)$ and its modulated integer translates $e^{2\pi inx}\chi_{[0,1)}(x-k)$ can be replaced by a *smooth window* g and its modulated integer translates.

Definition 9.5. A function $g \in L^2(\mathbb{R})$ is a *Gabor*[2] *function* if the family of its modulated integer translates

$$g_{n,k}(x) = g(x-k)e^{2\pi inx} \qquad \text{for } n,k \in \mathbb{Z} \tag{9.1}$$

is an orthonormal basis for $L^2(\mathbb{R})$. Such a basis is a *Gabor basis.* ♢

Example 9.6. The sharp window $g(x) = \chi_{[0,1)}(x)$ is a Gabor function. Figure 9.2 shows the imaginary part of $g_{n,k}(x)$ with the sharp window $g(x)$ and the parameters $n = 6$, $k = 0$. ♢

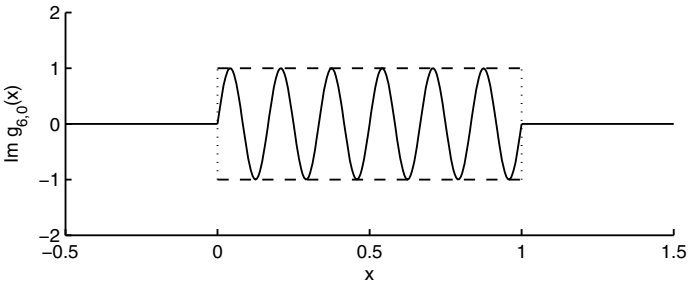

Figure 9.2. Graph (solid line) of the imaginary part $\operatorname{Im} g_{6,0}(x) = \sin(12\pi x)\chi_{[0,1)}(x)$ of the Gabor function $g_{6,0}(x) = \exp\{2\pi i 6x\}\chi_{[0,1)}(x)$, for the sharp window $g = \chi_{[0,1)}$. Dashed lines show the envelope formed by g and $-g$.

In 1946, Gabor proposed using systems of this kind in communication theory [**Gab**]. More specifically, he proposed using integer

[2]Named after Dennis Gabor (1900–1979), a Hungarian electrical engineer and inventor. He is most notable for inventing holography, for which he received the 1971 Nobel Prize in physics.

translates and modulations of the Gaussian function as a "basis" for $L^2(\mathbb{R})$, though unfortunately the Gaussian is not a Gabor function.

Lemma 9.7. *A family $\{\psi_n\}_{n\in\mathbb{N}}$ of functions is an orthonormal basis for $L^2(\mathbb{R})$ if and only if its Fourier transforms $\{\widehat{\psi_n}\}_{n\in\mathbb{N}}$ form an orthonormal basis for $L^2(\mathbb{R})$.*

Proof. Orthonormality holds on one side if and only if it holds on the other, because by the polarization formula (equation (7.24)), $\langle \psi_n, \psi_m\rangle = \langle \widehat{\psi_n}, \widehat{\psi_m}\rangle$ for all n, $m \in \mathbb{N}$. The same is true for completeness, because $f \perp \psi_n$ if and only if $\widehat{f} \perp \widehat{\psi_n}$. □

In particular, given a Gabor basis $\{g_{n,k}\}_{n,k\in\mathbb{Z}}$ as in equation (9.1), the Fourier transforms $\{\widehat{g_{n,k}}\}_{n,k\in\mathbb{Z}}$ form an *orthonormal* basis. Remarkably, they also form a *Gabor* basis. We see on closer examination that the form of the modulated integer translates in equation (9.1) is exactly what is needed here, since the Fourier transform converts translation to modulation, and vice versa.

Exercise 9.8. Use the time–frequency dictionary to show that the Fourier transforms of the Gabor basis elements are

$$\widehat{g_{n,k}}(\xi) = \widehat{g}\,(\xi - n)e^{-2\pi i k\xi} = (\widehat{g}\,)_{-k,n}(\xi). \tag{9.2}$$

◊

To sum up, we have proved the following lemma.

Lemma 9.9. *A function $g \in L^2(\mathbb{R})$ generates a Gabor basis, meaning that $\{g_{n,k}\}_{n,k\in\mathbb{Z}}$ forms an orthonormal basis in $L^2(\mathbb{R})$, if and only if $\widehat{g} \in L^2(\mathbb{R})$ generates a Gabor basis, meaning that $\{(\widehat{g})_{n,k}\}_{n,k\in\mathbb{Z}}$ forms an orthonormal basis in $L^2(\mathbb{R})$.*

Example 9.10. Since $g = \chi_{[0,1)}$ generates a Gabor basis, so does its Fourier transform

$$\widehat{g}\,(\xi) = (\chi_{[0,1)})^\wedge(\xi) = e^{-i\pi\xi}\left(\sin(\pi\xi)/\pi\xi\right) = e^{-i\pi\xi}\ \mathrm{sinc}(\xi).$$

This window $\widehat{g}$ is differentiable, in contrast to g which is not even continuous. However $\widehat{g}$ is not compactly supported, unlike g. ◊

Can we find a Gabor function that is simultaneously smooth and compactly supported? The answer is NO. The limitations of the Gabor analysis are explained by the following result.

Theorem 9.11 (Balian–Low[3] Theorem)**.** *If $g \in L^2(\mathbb{R})$ is a Gabor function, then either*

$$\int_{\mathbb{R}} x^2 |g(x)|^2 \, dx = \infty \quad \textit{or} \quad \int_{\mathbb{R}} \xi^2 |\widehat{g}\,(\xi)|^2 \, d\xi = \int_{\mathbb{R}} |g'(x)|^2 \, dx = \infty.$$

Exercise 9.12 (*Examples and Nonexamples of Balian–Low*)**.** Verify the Balian–Low Theorem for the two examples discussed so far: $g(x) = e^{2\pi i n x} \chi_{[0,1)}(x)$ and $g(x) = e^{-i\pi x} \operatorname{sinc} x$. Show also that the Gaussian $G(x) = e^{-\pi x^2}$ is not a Gabor function. ◊

A proof of Theorem 9.11 can be found in [**Dau92**, p. 108, Theorem 4.1.1]. The theorem implies that a Gabor window, or *bell*, cannot be simultaneously smooth and compactly supported. Our first example, $g(x) = \chi_{[0,1)}(x)$, is not even continuous but is perfectly localized in time, while our second example, $g(x) = e^{-i\pi x} \operatorname{sinc}(x)$, is the opposite. In particular the slow decay of the sinc function reflects the lack of smoothness of the characteristic function $\chi_{[0,1)}(x)$. This phenomenon is an incarnation of *Heisenberg's Uncertainty Principle*.

Exercise 9.13. Show that the Balian–Low Theorem implies that a Gabor function cannot be both smooth and compactly supported. ◊

However, if the exponentials are replaced by appropriate cosines and sines, one can obtain Gabor-type bases with smooth bell functions. These are the so-called *local cosine and sine bases*, first discovered by Malvar [**Malv**] and later described by Coifman and Meyer [**CMe**]. See the discussion in [**HW**] and the project in Section 9.6.

Another way to get around the Balian–Low Theorem is to accept some redundancy and use translations and modulations with less-than-integer increments, obtaining *frames* in place of orthogonal bases. Gabor would have been pleased.

There is a *continuous Gabor transform* as well, where the parameters are now real numbers instead of integers. Let g be a real and symmetric window with $\|g\|_2 = 1$. The Gabor transform is given by

$$Gf(\xi, u) = \int_{\mathbb{R}} f(x) g(x-u) e^{-2\pi i \xi x} \, dx = \langle f, g_{\xi,u} \rangle,$$

[3] Named after French physicist Roger Balian (born 1933) and American theoretical physicist Francis E. Low (1921–2007).

where $g_{\xi,u}(x) := g(x-u)e^{2\pi i\xi x}$ for u, $\xi \in \mathbb{R}$. The multiplication by the translated window localizes the Fourier integral in a neighborhood of u. The following *inversion formula* holds for $f \in L^2(\mathbb{R})$:

$$f(x) = \int_{\mathbb{R}^2} \langle f, g_{\xi,u}\rangle g_{\xi,u}(x)\, d\xi\, du. \tag{9.3}$$

These formulas are similar in spirit to the Fourier transform and the inverse Fourier transform integral formulas on $\mathbb{R}$.

Exercise 9.14. Verify the inversion formula (9.3) for f, $g \in \mathcal{S}(\mathbb{R})$, where g is a real-valued even function with $\|g\|_2 = 1$. **Hint:** Using the time–frequency dictionary, show that (9.3) is equivalent to $f(x) = \int \big(g_{\xi,0} * g_{\xi,0} * f\big)(x)\, d\xi$ and that the Fourier transform of the right-hand side is $\widehat{f}$. As usual, $*$ denotes convolution on $\mathbb{R}$. ◊

Gabor bases give partial answers to questions about localization. One problem is that the sizes of the windows are fixed. Variable widths applicable to different functions, while staying within a single basis, are the new ingredient added by wavelet analysis.

9.3. The wavelet transform

The wavelet transform involves translations (as in the Gabor basis) and scalings (instead of modulations). These translates and dilates introduce a natural zooming mechanism. The idea is to express a discrete signal in terms of its *average* value a_1 and successive levels of *detail*, d_1, d_2, ... Similarly, in archery, one first sees the entire *target* and then resolves the details of the *bull's-eye* painted on it. When using binoculars, one first locates the object (finds its average position) and then adjusts the focus until the fine details of the object jump into view. The zooming mechanism is mathematically encoded in the *multiresolution* structure of these bases; see Chapter 10.

Definition 9.15. A function $\psi \in L^2(\mathbb{R})$ is a *wavelet* if the family

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \quad \text{for } j,k \in \mathbb{Z} \tag{9.4}$$

forms an orthonormal basis of $L^2(\mathbb{R})$. If so, the basis is called a *wavelet basis*. ◊

The family of Fourier transforms of a wavelet basis is another orthonormal basis, but it is not a wavelet basis. It is generated from one function $\widehat{\psi}$ by scalings and modulations, rather than by scalings and translations.

Exercise 9.16 (*Fourier Transform of a Wavelet*)**.** Use the time–frequency dictionary (Table 7.1) to find $\widehat{\psi_{j,k}}(\xi)$, for $\psi \in L^2(\mathbb{R})$. ◊

The following reconstruction result follows immediately from the completeness of an orthonormal basis. The hard part is to identify functions ψ that are wavelets, so that the hypothesis of Proposition 9.17 holds. We tackle that issue in Chapter 10.

Proposition 9.17. *If $\psi \in L^2(\mathbb{R})$ is a wavelet, then the following* reconstruction formula *holds in the L^2 sense:*

$$f(x) = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle\, \psi_{j,k}(x) \qquad \text{for all } f \in L^2(\mathbb{R}).$$

Definition 9.18. The *orthogonal wavelet transform* is the map $W : L^2(\mathbb{R}) \to \ell^2(\mathbb{Z}^2)$ that assigns to each function in $L^2(\mathbb{R})$ the sequence of its wavelet coefficients:

$$Wf(j,k) := \langle f, \psi_{j,k}\rangle = \int_{\mathbb{R}} f(x)\overline{\psi_{j,k}(x)}\, dx. \qquad \text{◊}$$

The earliest known wavelet basis is the *Haar basis* on $L^2([0,1))$, introduced by Alfréd Haar in 1910 [**Haa**]. For the Haar basis, unlike the trigonometric basis, the partial sums for continuous functions converge uniformly.

Example 9.19 (*The Haar Wavelet*)**.** The *Haar wavelet* $h(x)$ on the unit interval is given by

$$h(x) := -\chi_{[0,1/2)}(x) + \chi_{[1/2,1)}(x).$$

See Figure 9.3. The family $\{h_{j,k}(x) := 2^{j/2}h(2^j x - k)\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$, as we will see in Section 9.4. ◊

Exercise 9.20. Show that $\{h_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal set; that is, verify that $\langle h_{j,k}, h_{j',k'}\rangle = 1$ if $j = j'$ and $k = k'$, and $\langle h_{j,k}, h_{j',k'}\rangle = 0$ otherwise. First show that the functions $h_{j,k}$ have zero integral: $\int h_{j,k} = 0$. ◊

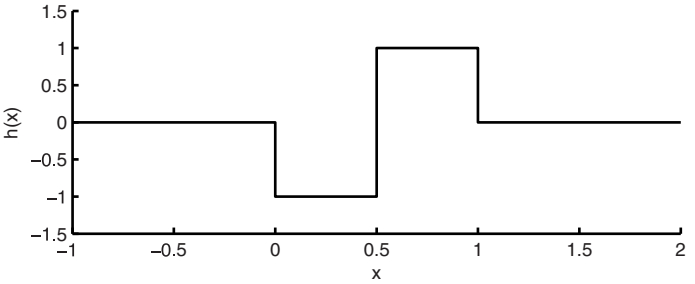

Figure 9.3. The Haar wavelet $h(x)$.

The reader may wonder how an arbitrary function in $L^2(\mathbb{R})$, which need not have zero integral, can be written in terms of the Haar functions, which all have zero integral. To resolve this puzzle, see Section 9.4.3 and the project in Section 9.7.

An important part of wavelet theory is the search for smoother wavelets. The Haar wavelet is discontinuous. It is also perfectly localized in time and therefore not perfectly localized in frequency. At the other end of the spectrum, one finds the *Shannon wavelet*[4] which is localized in frequency but not in time.

Example 9.21 (*The Shannon Wavelet*)**.** Let ψ be given on the Fourier side by

$$\widehat{\psi}(\xi) := e^{2\pi i\xi}\chi_{[-1,-1/2)\cup[1/2,1)}(\xi).$$

The family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$. The function ψ is the *Shannon wavelet*, and the corresponding basis is the *Shannon basis*. ♢

Exercise 9.22. Show that the Shannon functions $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ form an orthonormal set and each has zero integral. Furthermore, they are a basis. **Hint:** Work on the Fourier side using the polarization formula (7.24). On the Fourier side we are dealing with a win-

[4]Named after the same Shannon as the sampling theorem in Section 8.5.2.

dowed Fourier basis, with *double-paned windows* $F_j := [-2^j, -2^{j-1}) \cup [2^{j-1}, 2^j)$ of combined length 2^j, for $j \in \mathbb{Z}$, that are congruent[5] modulo 2^j to the interval $[0, 2^j)$. The collection of double-paned windows F_j forms a partition of $\mathbb{R} \setminus \{0\}$. On each double-paned window, the trigonometric functions $2^{-j/2} e^{2\pi i k x 2^{-j}} \chi_{F_j}(x)$, for $k \in \mathbb{Z}$ and a fixed j, form an orthonormal basis of $L^2(F_j)$. Now use Lemma A.51. $\Diamond$

The Shannon wavelet is perfectly localized in frequency, therefore not in time. Compact support on the frequency side translates into smoothness (C^∞) of the Shannon wavelet on the time side. Thus the Shannon wavelet is an example of a C^∞ wavelet without compact support. Can one find compactly supported wavelets that are smooth? YES. In a fundamental paper on wavelet theory [**Dau88**], I. Daubechies constructed compactly supported wavelets with arbitrary (but finite) smoothness; they are in C^k. However, it is impossible to construct wavelets that are both compactly supported and C^∞.

We can associate to most wavelets a sequence of numbers, known as a *filter*, and a companion *scaling function*.

A *finite impulse response (FIR) filter* has only finitely many nonzero entries. If a wavelet has a finite impulse response filter, then the wavelet and its scaling function are compactly supported. The converse is false; see the project in Subsection 10.5.2.

One of Daubechies' key contributions was her discovery of the family of wavelets later named after her, each of which has finitely many nonzero coefficients h_k and wavelet and scaling functions ψ and ϕ (Figure 9.4) having preselected regularity (meaning smoothness, or degree of differentiability) and compact support. As noted in [**Mey**, Chapter 1], eighty years separated Haar's work and its natural extension by Daubechies. IT'S AMAZING! Something as irregular as the spiky function in Figure 9.4 is a wavelet! No wonder it took so long to find these continuous, compactly supported wavelets.

[5] A subset A of $\mathbb{R}$ is *congruent* to an interval I if for each point $x \in A$ there exists a unique integer k such that $x + k|I|$ is in I and for each point $y \in I$ there is a unique $x \in A$ and $k \in \mathbb{Z}$ so that $y = x + k|I|$. Here is a nonexample: The set $[-1/2, 0) \cup [1/2, 1)$ is not congruent modulo 1 to the interval $[0, 1)$.

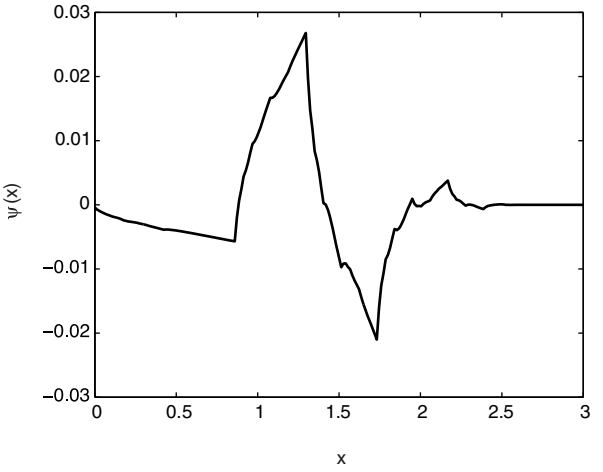

Figure 9.4. The Daubechies wavelet function ψ, for the *db*2 wavelet. The shape is reminiscent of the profile of the Sydney Opera House.

The more derivatives a wavelet has, the longer the filter and the longer the support of the wavelet. The shorter the filter, the better for implementation, so there is a trade-off. Filter bank theory and the practicability of implementing FIR filters opened the door to widespread use of wavelets in applications. See Section 10.3.4 and Chapter 11.

One can develop the theory of wavelets in $\mathbb{R}^N$ or in $\mathbb{C}^N$ via linear algebra, in the same way that we built a finite Fourier theory and introduced the discrete Haar basis in Chapter 6. See **[Fra]**. There is a *Fast Haar Transform* (FHT) (Section 6.7), and it generalizes to a *Fast Wavelet Transform* (FWT) (Chapter 11), which has been instrumental in the success of wavelets in practice.

As in the Fourier and Gabor cases, there is also a *continuous wavelet transform*. It uses continuous translation and scaling parameters, $u \in \mathbb{R}$ and $s > 0$, and a family of *time–frequency atoms* that is obtained from a normalized wavelet $\psi \in L^2(\mathbb{R})$ with zero average

($\int \psi = 0$) by shifting by u and rescaling by s. Let

$$\psi_{s,u}(x) = \frac{1}{\sqrt{s}}\,\psi\left(\frac{x-u}{s}\right), \quad \text{so} \quad (\psi_{s,u})^\wedge(\xi) = \sqrt{s}\,e^{-2\pi i u\xi}\,\widehat{\psi}\,(s\xi).$$

The *continuous wavelet transform* is then defined by

$$Wf(s,u) = \langle f, \psi_{s,u}\rangle = \int_{\mathbb{R}} f(x)\,\overline{\psi_{s,u}(x)}\,dx.$$

If ψ is real-valued and localized near 0 with spread 1, then $\psi_{s,u}$ is localized near u with spread s. The wavelet transform measures the variation of f near u at scale s (in the discrete case, $u = k2^{-j}$ and $s = 2^{-j}$). As the scale s goes to zero (j goes to infinity), the decay of the wavelet coefficients characterizes the regularity of f near u. Also, if $\widehat{\psi}$ is localized near 0 with spread 1, then $\widehat{\psi_{s,u}}$ is localized near 0 with spread $1/s$. That is, the Heisenberg boxes of the wavelets are rectangles of area 1 and dimensions $s \times 1/s$. (See Section 8.5.3.)

Under very mild assumptions on the wavelet ψ, we obtain a reconstruction formula. Each $f \in L^2(\mathbb{R})$ can be written as

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^{+\infty} Wf(s,u)\psi_{s,u}(x)\,\frac{du\,ds}{s^2}, \tag{9.5}$$

provided that ψ satisfies *Calderón's admissibility condition* [**Cal**]:

$$C_\psi := \int_0^\infty \frac{|\widehat{\psi}\,(\xi)|^2}{\xi}\,d\xi < \infty.$$

This reconstruction formula can be traced back to the famous *Calderón reproducing formula*[6]:

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \big(\psi_{s,0} * \overline{\widetilde{\psi_{s,0}}} * f\big)(x)\,\frac{ds}{s^2}, \tag{9.6}$$

where $\widetilde{\psi}(x) = \psi(-x)$ and $*$ denotes convolution in $\mathbb{R}$.

Exercise 9.23. Show that the Calderón reproducing formula (9.6) and the reconstruction formula (9.5) are the same. Show that equation (9.6) holds for f, $\psi \in \mathcal{S}(\mathbb{R})$ such that $C_\psi < \infty$, by checking that the Fourier transform of the right-hand side coincides with $\widehat{f}$. $\Diamond$

[6] Named after Argentinian mathematician Alberto Pedro Calderón (1920–1998).

9.4. Haar analysis

In this section we show that the Haar functions form a complete orthonormal system. Verifying the orthonormality of the system reduces to understanding the geometry of the dyadic intervals. Verifying the completeness of the system reduces to understanding that the limit in $L^2(\mathbb{R})$ of the averaging operators over intervals as the intervals shrink to a point $x \in \mathbb{R}$ is the identity operator and that the limit as the intervals grow to be infinitely long is the zero operator.

We first define the dyadic intervals and describe their geometry. We define the Haar function associated to an interval and observe that the family of Haar functions indexed by the dyadic intervals coincides with the Haar basis from Example 9.19. We show that the Haar functions are orthonormal. Next, we introduce the expectation (or averaging) and difference operators, relate the completeness of the Haar system to the limiting behavior of the averaging operators, and prove the required limit results for continuous functions and compactly supported functions. Finally, an approximation argument coupled with some uniform bounds gives the desired result: the completeness of the Haar system. Along the way we mention the Lebesgue Differentiation Theorem and the Uniform Boundedness Principle and some of their applications.

9.4.1. The dyadic intervals. Given an interval $I = [a, b)$ in $\mathbb{R}$, the *left half*, or *left child*, of I is the interval $I_l := [a, (a + b)/2)$, and the *right half*, or *right child*, of I is the interval $I_r := [(a + b)/2, b)$ (Figure 9.5). Let $|I|$ denote the length of an interval I.

Given a locally integrable function f, let $m_I f$ denote the *average value*[7] *of* f *over* I:

$$m_I f := \frac{1}{|I|} \int_I f(x)\, dx. \tag{9.7}$$

Definition 9.24 (*Dyadic Intervals*)**.** The *dyadic intervals* are the half-open intervals of the form

$$I_{j,k} = [k2^{-j}, (k + 1)2^{-j}) \qquad \text{for integers } j,\ k.$$

[7]Other popular notations for $m_I f$ are $\langle f \rangle_I$ and f_I.

Let $\mathcal{D}$ denote the set of all dyadic intervals in $\mathbb{R}$, and $\mathcal{D}_j$ the set of intervals $I \in \mathcal{D}$ of length 2^{-j}, also called the j^{th} *generation*. Then $\mathcal{D} = \bigcup_{j\in\mathbb{Z}} \mathcal{D}_j$. Each $\mathcal{D}_j$ forms a partition of the real line. ◊

For instance, $[5/8, 3/4)$ and $[-16, -12)$ are dyadic intervals, while $[3/8, 5/8)$ is not. If I is a dyadic interval, then its children, grandchildren, parents, grandparents, and so on are also dyadic intervals.

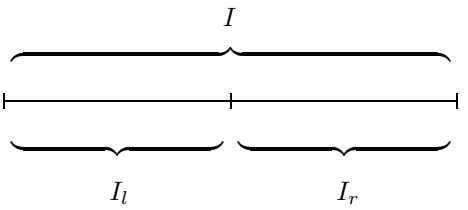

Figure 9.5. Parent interval I and its children I_l and I_r.

Each dyadic interval I belongs to a unique generation $\mathcal{D}_j$, and the next generation $\mathcal{D}_{j+1}$ contains exactly two subintervals of I, namely I_l and I_r. Given two distinct intervals I, $J \in \mathcal{D}$, either I and J are disjoint or one is contained in the other. This nestedness property of the dyadic intervals is so important that we highlight it as a lemma.

Lemma 9.25 (*Dyadic Intervals Are Nested or Disjoint*)**.** *If I, $J \in \mathcal{D}$, then exactly one of the following holds: $I \cap J = \emptyset$ or $I \subseteq J$ or $J \subsetneq I$. Moreover if $J \subsetneq I$, then J is a subset of the left or the right child of I.*

Exercise 9.26. Prove Lemma 9.25. Also show that its conclusion need not hold for nondyadic intervals. ◊

Given $x \in \mathbb{R}$ and $j \in \mathbb{Z}$, there is a unique interval in $\mathcal{D}_j$ that contains x. We denote this unique interval by $I_j(x)$. Figure 9.8 on page 244, shows several dyadic intervals $I_j(x)$ for a fixed point x; they form a *dyadic tower* containing the point x. The intervals $I_j(x)$ shrink to the set $\{x\}$ as $j \to \infty$.

9.4.2. The Haar basis. We associate to each dyadic interval I a step function h_I that is supported on I, is constant on each of I_l

and I_r, has zero integral average, and has L^2 norm equal to one. The zero integral average makes the value of h_I on the left child be the negative of the value on the right child. The L^2 normalization makes the absolute value of the function on I be exactly $1/\sqrt{|I|}$. We make the convention that the value is negative on I_l and thus positive on I_r. These considerations uniquely determine h_I.

Definition 9.27. The *Haar function associated to the interval I* is the step function h_I defined by

$$h_I(x) := (1/\sqrt{|I|})\big(\chi_{I_r}(x) - \chi_{I_l}(x)\big). \qquad \Diamond$$

The Haar wavelet h defined in Example 9.19 coincides with the Haar function $h_{[0,1)}$ associated to the unit interval $[0,1)$. The Haar wavelets $h_{j,k}$ coincide with $h_{I_{j,k}}$, where $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$. Figure 9.6 shows the graphs of two Haar functions.

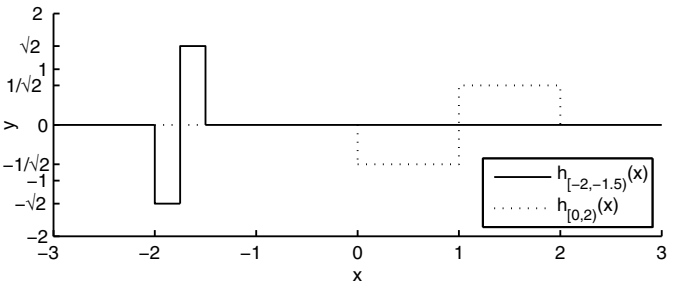

Figure 9.6. Graphs of the two Haar functions defined by $h_{[-2,-1.5)}(x) = \sqrt{2}\,[\chi_{[-1.75,-1.5)}(x) - \chi_{[-2,-1.75)}(x)]$ and $h_{[0,2)}(x) = (1/\sqrt{2})\,[\chi_{[1,2)}(x) - \chi_{[0,1)}(x)]$.

Exercise 9.28. Show that

$$h_{I_{j,k}}(x) = 2^{j/2}h(2^j x - k) = h_{j,k}(x), \qquad \text{where } h = h_{[0,1)}. \qquad \Diamond$$

By the results of Exercises 9.20 and 9.28, the set $\{h_I\}_{I\in\mathcal{D}}$ is an orthonormal set in $L^2(\mathbb{R})$. We can also prove directly that the Haar functions indexed by the dyadic intervals form an orthonormal family, using the fact that the dyadic intervals are nested or disjoint.

Lemma 9.29. *The family of Haar functions $\{h_I\}_{I\in\mathcal{D}}$ indexed by the dyadic intervals is an orthonormal family.*

Proof. Consider I, $J \in \mathcal{D}$ with $I \neq J$. Either they are disjoint or one is strictly contained in the other, by Lemma 9.25. If I and J are disjoint, then clearly $\langle h_I, h_J\rangle = 0$, because the supports are disjoint. If I is strictly contained in J, then h_J is constant on I, and in fact

$$\langle h_I, h_J\rangle = \int_I h_I(x)h_J(x)\,dx = \frac{1}{|J|^{1/2}}\int h_I(x)\,dx = 0.$$

The case $J \subsetneq I$ is similar. This proves the orthogonality of the family. For the normality, when $I = J$ we have

$$\langle h_I, h_I\rangle = \|h_I\|_{L^2(\mathbb{R})} = \int_I |h_I(x)|^2\,dx = \frac{1}{|I|}\int_I dx = 1. \qquad \square$$

The Haar system is not only orthonormal but also a complete orthonormal system and hence a basis for $L^2(\mathbb{R})$, as we now show.

Theorem 9.30. *The Haar functions $\{h_I\}_{I\in\mathcal{D}}$ form an orthonormal basis for $L^2(\mathbb{R})$.*

In an N-dimensional space, an orthonormal set of N vectors is automatically a basis. Thus in Chapter 6 to show that the Haar vectors form a basis for $\mathbb{C}^N$, it was enough to show that there are N orthonormal Haar vectors. In infinite-dimensional space we have to do more than simply counting an orthonormal set. To prove Theorem 9.30, we must make sure the set is *complete*. In other words, for all $f \in L^2(\mathbb{R})$, the identity

$$f = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I \tag{9.8}$$

must hold in the L^2 sense[8]. We must show that each $L^2(\mathbb{R})$ function f can be written as a (possibly infinite) sum of Haar functions, weighted by coefficients given by the inner products of f with h_I. (See for example the lower figure on the cover, where a simple signal is broken into a weighted sum of Haar functions.)

An alternative proof shows that the only function in $L^2(\mathbb{R})$ orthogonal to all Haar functions is the zero function (see Theorem A.41).

[8]Recall that equation (9.8) holds in the L^2 sense if $\|f - \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I\|_2 = 0$.

We will show that both arguments boil down to checking some limit properties of the expectation operators defined in Section 9.4.4.

9.4.3. Devil's advocate. Before we prove the completeness of the Haar system, let us play devil's advocate.

- First, for the function $f(x) \equiv 1$ we have $\langle f, h_I \rangle = \int h_I = 0$ for all $I \in \mathcal{D}$. Thus f is orthogonal to all the Haar functions, so how can the Haar system be complete? Are we contradicting Theorem 9.30? No, because f is not in $L^2(\mathbb{R})$!
- Second, how can functions that all have zero integral (the Haar functions) reconstruct functions that do not have zero integral?[9]

If the Haar system is complete, then for each $f \in L^2(\mathbb{R})$, equation (9.8) holds. Integrating on both sides and interchanging the sum and the integral, we see that

$$\text{“} \int_{\mathbb{R}} f(x)\,dx = \int_{\mathbb{R}} \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_I(x)\,dx = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle \int_{\mathbb{R}} h_I(x)\,dx = 0 \text{”}.$$

The last equality holds because the Haar functions have integral zero. It seems that all functions in $L^2(\mathbb{R})$ must themselves have integral zero. But that is not true, since for example $\chi_{[0,1)}$ belongs to $L^2(\mathbb{R})$ yet has integral one. What's wrong? Perhaps the Haar system is not complete after all. Or is there something wrong in our calculation? The Haar system *is* complete; it turns out that what is illegal above is the interchange of sum and integral. See the project in Section 9.7.

9.4.4. The expectation and difference operators, P_j and Q_j. We introduce two important *operators*[10] that will help us to understand the zooming properties of the Haar basis.

[9] This question was posed by Lindsay Crowl, a student in the 2004 Program for Women in Mathematics, where we gave the lectures that led to this book. It was a natural concern and a good example of the dangers of interchanging limit operations!

[10] An *operator* is a mapping from a space of functions into another space of functions. The input is a function and so is the output. The Fourier transform is an example of an operator.

Definition 9.31. The *expectation* operators $P_j : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $j \in \mathbb{Z}$, act by taking averages over dyadic intervals at generation j:

$$P_j f(x) := \frac{1}{|I_j|} \int_{I_j} f(t)\, dt,$$

where $I_j = I_j(x)$ is the unique interval of length 2^{-j} that contains x. $\Diamond$

The new function $P_j f$ is a step function that is constant on each dyadic interval $I \in \mathcal{D}_j$ in the j^{th} generation. Furthermore the value of the function $P_j f$ on an interval $I \in \mathcal{D}_j$ is the integral average of the function f over the interval I: for each $x \in I$,

$$P_j f(x) \equiv m_I f := \frac{1}{|I|} \int_I f(y)\, dy. \tag{9.9}$$

Figure 9.7 (upper plot) shows the graph of a particular function f, together with the graphs of $P_j f$ and $P_{j+1} f$, for $j = -1$.

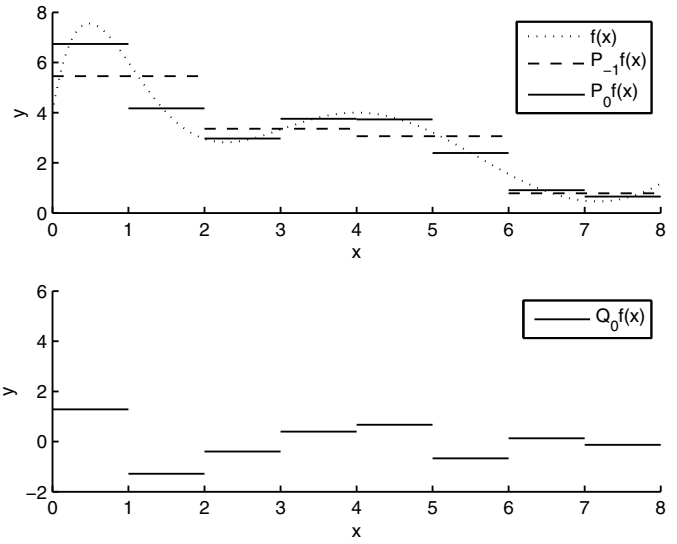

Figure 9.7. Graphs of f, $P_{-1} f$, $P_0 f$, and $Q_0 f$. We have used the function $f(x) = 4 + x(x - 1.5)(x - 4)^2(x - 9) \times e^{-x/2.5}/12$.

Exercise 9.32. Verify that $P_j f(x) = \sum_{I \in \mathcal{D}_j} m_I f \, \chi_I(x)$. ◊

As $j \to \infty$, the length $|I| = 2^{-j}$ of the steps goes to zero, and we expect $P_j f$ to be a better and better approximation of f. We make this idea precise in Section 9.4.5 below.

Definition 9.33. The *difference operators* $Q_j : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $j \in \mathbb{Z}$, are given by $Q_j f(x) := P_{j+1} f(x) - P_j f(x)$. ◊

These operators Q_j encode the information necessary to go from the approximation $P_j f$ at resolution j of f to the better approximation $P_{j+1} f$ at resolution $j+1$. Figure 9.7 (lower plot) shows the graph of $Q_j f$ for $j = 0$.

Notice that when we superimpose the pictures of $P_{j+1} f$ and $P_j f$, the averages at the coarser scale j seem to be sitting exactly halfway between the averages at the finer scale $j + 1$, so that $Q_j f$ seems to be a linear combination of the Haar functions at scale j. Lemma 9.35 makes this fact precise. It is implied by the following useful relationship between integral averages on nested dyadic intervals.

Exercise 9.34. Show that $m_I f = (m_{I_l} f + m_{I_r} f)/2$. In words, the integral average over a parent interval is the average of the integral averages over its children. ◊

Lemma 9.35. *For $f \in L^2(\mathbb{R})$, $Q_j f(x) = \sum_{I \in \mathcal{D}_j} \langle f, h_I \rangle h_I(x)$.*

Proof. By definition of h_I and since $|I| = 2|I_r| = 2|I_l|$, we have

$$\begin{aligned} \langle f, h_I \rangle h_I(x) &= \frac{\sqrt{|I|}}{2} \Big(\frac{1}{|I_r|} \int_{I_r} f - \frac{1}{|I_l|} \int_{I_l} f \Big) h_I(x) \\ &= \sqrt{|I|}/2 \, (m_{I_r} f - m_{I_l} f) \, h_I(x). \end{aligned}$$

Since $h_I(x) = 1/\sqrt{|I|}$ for $x \in I_r$ and $h_I(x) = -1/\sqrt{|I|}$ for $x \in I_l$, we conclude that if $x \in I$, then

$$\langle f, h_I \rangle h_I(x) = \begin{cases} (m_{I_r} f - m_{I_l} f)/2, & \text{if } x \in I_r; \\ -(m_{I_r} f - m_{I_l} f)/2, & \text{if } x \in I_l. \end{cases} \tag{9.10}$$

On the other hand, if $x \in I \in \mathcal{D}_j$, then $P_j f(x) = m_I f$; if $x \in I_r$, then $P_{j+1} f(x) = m_{I_r} f$; and if $x \in I_l$, then $P_{j+1} f(x) = m_{I_l} f$. Hence

$$Q_j f(x) = \begin{cases} m_{I_r} f - m_I f, & \text{if } x \in I_r; \\ m_{I_l} f - m_I f, & \text{if } x \in I_l. \end{cases}$$

The averaging property in Exercise 9.34 implies that

$$m_{I_r} f - m_I f = (m_{I_r} f - m_{I_l} f)/2 = m_I f - m_{I_l} f.$$

We conclude that $Q_j f(x) = \sum_{I \in \mathcal{D}_j} \langle f, h_I \rangle h_I(x)$ for $x \in I \in \mathcal{D}_j$. □

9.4.5. Completeness of the Haar system. To prove that the Haar system of functions is complete, we must show that for all $f \in L^2(\mathbb{R})$, we have

$$f(x) = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_I(x).$$

By Lemma 9.35, this condition is equivalent to the condition

$$f(x) = \lim_{M,N \to \infty} \sum_{-M \le j < N} Q_j f(x).$$

A telescoping series argument shows that

$$\begin{aligned} P_N f(x) - P_M f(x) &= \sum_{M \le j < N} (P_{j+1} f(x) - P_j f(x)) \\ &= \sum_{M \le j < N} Q_j f(x). \end{aligned} \tag{9.11}$$

Therefore, verifying completeness reduces to checking that

$$f(x) = \lim_{N \to \infty} P_N f(x) - \lim_{M \to -\infty} P_M f(x),$$

where all the above equalities hold in the L^2 sense. It suffices to prove the following theorem.

Theorem 9.36. *For $f \in L^2(\mathbb{R})$, we have*

$$\lim_{M \to -\infty} \|P_M f\|_2 = 0 \qquad \textit{and} \tag{9.12}$$

$$\lim_{N \to \infty} \|P_N f - f\|_2 = 0. \tag{9.13}$$

Exercise 9.37. Use Theorem 9.36 to show that if $f \in L^2(\mathbb{R})$ is orthogonal to all Haar functions, then f must be zero in $L^2(\mathbb{R})$. ◊

Aside 9.38. Before proving Theorem 9.36, we pause to develop the Lebesgue Differentiation Theorem on $\mathbb{R}$. Here $I_j(x)$ is the unique dyadic interval in $\mathcal{D}_j$ that contains x, as shown in Figure 9.8. Equation (9.13) says that given $x \in \mathbb{R}$, the averages $P_N f$ of the function f over the dyadic intervals $\{I_j(x)\}_{j\in\mathbb{Z}}$ converge to $f(x)$ in the L^2 sense as the intervals shrink: $\lim_{j\to\infty} \frac{1}{|I_j(x)|} \int_{I_j(x)} f(t)\,dt = f(x)$. In fact, the convergence also holds *almost everywhere*[11] (a.e.).

Theorem 9.39 (Lebesgue Differentiation Theorem on $\mathbb{R}$)**.** *Suppose f is locally integrable: $f \in L^1_{\text{loc}}(\mathbb{R})$. Let I denote any interval, dyadic or not, that contains x. Then*

$$\lim_{x\in I, |I|\to 0} \frac{1}{|I|} \int_I f(y)\,dy = f(x) \quad \textit{for a.e. } x \in \mathbb{R}.$$

For a proof see [**SS05**, Chapter 3]. In Chapter 4 we stated the special case with intervals $[x-h, x+h]$ centered at x, as $h \to 0$. ◊

Exercise 9.40. Prove the *Lebesgue Differentiation Theorem* for continuous functions, and show that for continuous functions the pointwise convergence holds everywhere. That is, for all $x \in \mathbb{R}$ and intervals $[a, b]$ containing x, $\lim_{[a,b]\to\{x\}} \frac{1}{b-a} \int_a^b f(t)\,dt = f(x)$. ◊

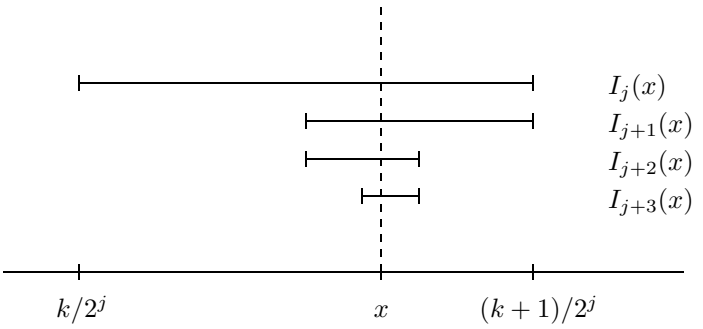

Figure 9.8. The tower $\cdots \supset I_j(x) \supset I_{j+1}(x) \supset I_{j+2}(x) \supset I_{j+3}(x) \supset \cdots$ of nested dyadic intervals containing the point x, with $k/2^j \le x \le (k+1)/2^j$. Here j and k are integers.

[11] In other words, the convergence holds except possibly on a set of measure zero.

Theorem 9.36 is a consequence of the following three lemmas.

Lemma 9.41. *The operators P_j are* uniformly bounded *in $L^2(\mathbb{R})$. Specifically, for each function $f \in L^2(\mathbb{R})$ and for every integer j,*

$$\|P_j f\|_2 \le \|f\|_2.$$

Lemma 9.42. *If g is continuous and has compact support on the interval $[-K, K]$, then Theorem* 9.36 *holds for g.*

Lemma 9.43. *The continuous functions with compact support are* dense *in $L^2(\mathbb{R})$. In other words, given $f \in L^2(\mathbb{R})$, for any $\varepsilon > 0$ there exist functions g and h, such that $f = g + h$, where g is a continuous function with compact support on an interval $[-K, K]$ and $h \in L^2(\mathbb{R})$ with small L^2 norm, $\|h\|_2 < \varepsilon$.*

We first deduce the theorem from the lemmas and then after some comments prove the lemmas.

Proof of Theorem 9.36. By Lemma 9.43, given $\varepsilon > 0$, we can decompose f as $f = g + h$, where g is continuous with compact support on $[-K, K]$ and $h \in L^2(\mathbb{R})$ with $\|h\|_2 < \varepsilon/4$. By Lemma 9.42, we can choose N large enough so that for all $j > N$,

$$\|P_{-j} g\|_2 \le \varepsilon/2.$$

The expectation operators P_j are linear operators, so $P_j(g + h) = P_j g + P_j h$. Now by the Triangle Inequality and Lemma 9.41 we have

$$\|P_{-j} f\|_2 \le \|P_{-j} g\|_2 + \|P_{-j} h\|_2 \le \frac{\varepsilon}{2} + \|h\|_2 \le \varepsilon.$$

This proves equation (9.12). Similarly, by Lemma 9.42, we can choose N large enough that for all $j > N$,

$$\|P_j g - g\|_2 \le \varepsilon/2.$$

Now using the Triangle Inequality (twice) and Lemma 9.41,

$$\|P_j f - f\|_2 \le \|P_j g - g\|_2 + \|P_j h - h\|_2 \le \frac{\varepsilon}{2} + 2\|h\|_2 \le \varepsilon.$$

This proves equation (9.13). □

Exercise 9.44. Show that the operators P_j are linear. ◊

We have twice used a very important principle from functional analysis: *If a sequence of linear operators is uniformly bounded on a Banach space and the sequence converges to a bounded operator on a dense subset of the Banach space, then it converges on the whole space to the same operator.* For us the Banach space is $L^2(\mathbb{R})$, the dense subset is the set of continuous functions with compact support, the linear operators are P_j, and the uniform bounds are provided by Lemma 9.41. The operators converge to the zero operator as $j \to -\infty$, and they converge to the identity operator as $j \to \infty$.

A related principle is the *Uniform Boundedness Principle*, or the *Banach–Steinhaus Theorem*, from functional analysis. See for example **[Sch**, Chapter III]. This principle gives sufficient conditions for a family of operators to be uniformly bounded. A beautiful application of the Uniform Boundedness Principle is to show the existence of a real-valued continuous periodic function whose Fourier series diverges at a given point x_0. See Aside 9.48.

We now prove the three lemmas that implied Theorem 9.36, which in turn gave us the completeness of the Haar system.

Proof of Lemma 9.41. We can estimate for $x \in I \in \mathcal{D}_j$,

$$\begin{aligned} |P_j f(x)|^2 &= \Big| \frac{1}{|I|} \int_I f(t)\,dt \Big|^2 \le \frac{1}{|I|^2} \Big(\int_I 1^2\,dt \Big) \Big(\int_I |f(t)|^2\,dt \Big) \\ &= \frac{1}{|I|} \int_I |f(t)|^2\,dt. \end{aligned}$$

The inequality is a consequence of the Cauchy–Schwarz Inequality.

Now integrate over the interval I to obtain

$$\int_I |P_j f(x)|^2\,dx \le \int_I |f(t)|^2\,dt,$$

and sum over all intervals in $\mathcal{D}_j$ (this is a disjoint family that covers the whole line!):

$$\begin{aligned} \int_{\mathbb{R}} |P_j f(x)|^2\,dx &= \sum_{I \in \mathcal{D}_n} \int_I |P_j f(x)|^2\,dx \le \sum_{I \in \mathcal{D}_n} \int_I |f(t)|^2\,dt \\ &= \int_{\mathbb{R}} |f(t)|^2\,dt, \end{aligned}$$

as required. We have used Lemma A.51 twice here. □

Proof of Lemma 9.42. Suppose that the function g is continuous and that it is supported on the interval $[-K, K]$. If j is large enough that $K < 2^j$ and if $x \in [0, 2^j) \in \mathcal{D}_{-j}$, then $|P_{-j}g(x)| = \frac{1}{2^j}\int_0^K |g(t)|dt$, and applying the Cauchy–Schwarz Inequality, we get

$$|P_{-j}g(x)| \le \frac{1}{2^j}\Big(\int_0^K 1^2\,dt\Big)^{1/2}\Big(\int_0^K |g(t)|^2\,dt\Big)^{1/2} \le \frac{\sqrt{K}}{2^j}\|g\|_2.$$

The same inequality holds for $x < 0$. Also, if $|x| \ge 2^j$, then $P_{-j}g(x) = 0$, because the interval in $\mathcal{D}_j$ that contains x is disjoint with the support of g. We can now estimate the L^2 norm of $P_{-j}g$:

$$\|P_{-j}g\|_2^2 = \int_{-2^j}^{2^j} |P_{-j}g(x)|^2\,dx \le \frac{1}{2^{2j}}K\|g\|_2^2\int_{-2^j}^{2^j} 1\,dx = 2^{-j+1}K\|g\|_2^2.$$

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 (9.12) for continuous functions with compact support.

We are assuming that g is continuous and that it is supported on the compact interval $[-K, K] \subset [-2^M, 2^M]$. But then g is uniformly continuous. So given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|g(y) - g(x)| < \varepsilon/\sqrt{2^{M+1}} \quad \text{whenever } |y - x| < \delta.$$

Now choose $N > M$ large enough that $2^{-j} < \delta$ for all $j > N$. Each point x is contained in a unique $I \in \mathcal{D}_j$, with $|I| = 2^{-j} < \delta$. Therefore $|y - x| \le \delta$ 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

$$\begin{aligned}\int_{\mathbb{R}} |P_jg(x) - g(x)|^2\,dx &= \int_{|x|\le 2^M} |P_jg(x) - g(x)|^2\,dx \\ &< \frac{\varepsilon^2}{2^{M+1}}\int_{|x|\le 2^M} 1\,dx = \varepsilon^2.\end{aligned}$$

Notice that if $|x| > 2^M$, then for $j > N \ge M$, $P_jg(x)$ 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 $n > N$,

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

This proves equation (9.13) for continuous functions with compact support. □

Proof of Lemma 9.43. This lemma is an example of an approximation theorem in $L^2(\mathbb{R})$. First, we choose K large enough so that the tail of f has very small L^2 norm, in other words, $\|f\chi_{\{x\in\mathbb{R}:|x|>K\}}\|_2 \leq \varepsilon/3$. Second, we recall that on compact intervals, the continuous functions are dense in $L^2([-K, K])$; see Theorem 2.75. (For example, polynomials are dense, and trigonometric polynomials are also dense, by the Weierstrass Approximation Theorem (Theorem 3.4).) Now choose g_1 continuous on $[-K, K]$ so that $\|(f - g_1)\chi_{[-K,K]}\|_2 \leq \varepsilon/3$. Third, it could happen that g_1 is continuous on $[-K, K]$, but when extended to be zero outside the interval, it is not continuous on the line. That can be fixed by giving yourself some margin at the endpoints: define g to coincide with g_1 on $[-K + \delta, K - \delta]$ and to be zero outside $[-K, K]$, and connect these pieces with straight segments, so that g is continuous on $\mathbb{R}$. Finally, choose δ small enough so that $\|g_1 - g\|_2 \leq \varepsilon/3$. Now let

$$h = f - g = f\chi_{\{x\in\mathbb{R}:|x|>K\}} + (f - g_1)\chi_{[-K,K]} + (g_1 - g)\chi_{[-K,K]}.$$

By the Triangle Inequality,

$$\|h\|_2 \leq \|f\chi_{\{x\in\mathbb{R}:|x|>K\}}\|_2 + \|(f - g_1)\chi_{[-K,K]}\|_2 + \|g_1 - g\|_2 \leq \varepsilon. \quad \square$$

We have shown (Lemma 9.42) that the step functions can approximate continuous functions with compact support in the L^2 norm. Lemma 9.43 shows that we can approximate L^2 functions by continuous functions with compact support. Therefore, we can approximate L^2 functions by step functions, in the L^2 norm. Furthermore, we can choose the steps to be dyadic intervals of a fixed generation, for any prescribed accuracy.

Exercise 9.45 (*Approximation by Step Functions*)**.** Show that continuous functions with compact support can be approximated in the

uniform norm by step functions (with compact support). Furthermore, one can choose the intervals where the approximating function is constant to be dyadic intervals of a fixed generation for any prescribed accuracy. More precisely, show that given f continuous and compactly supported on $\mathbb{R}$ and $\varepsilon > 0$, there exists $N > 0$ such that for all $j > N$ and for all $x \in \mathbb{R}$, $|P_j f(x) - f(x)| < \varepsilon$. ◊

Aside 9.46. The Lebesgue Differentiation Theorem (Exercise 9.40) holds for continuous functions, as a consequence of the Fundamental Theorem of Calculus. A related function, defined for all $x \in \mathbb{R}$ and each locally integrable function f (possibly as ∞), is the *Hardy–Littlewood maximal function*[12] denoted Mf and defined by

$$Mf(x) = \sup_{x \in I} \frac{1}{|I|} \int_I |f(t)|\, dt. \tag{9.14}$$

The supremum in the definition of $Mf(x)$ is taken over all intervals containing x. Notice that the supremum is always defined, unlike the limit in the Lebesgue Differentiation Theorem as the intervals shrink to the point x; *a priori* that limit might not exist. The Lebesgue Differentiation Theorem guarantees both that the limit does exist and that it equals the value of the function at the point x, for almost every x. Boundedness and weak boundedness results for the maximal function (see the project in Section 12.9.2) can be used to deduce the Lebesgue Differentiation Theorem; see [**SS05**, Chapter 3, Section 1].

Maximal functions appear throughout harmonic analysis as controllers for other operators T, in the sense that $\|T\| \leq \|M\|$. It is important to understand their boundedness properties. For example, all convolution operators with good radial kernels can be bounded pointwise by a constant multiple of $Mf(x)$.

Exercise 9.47. Let $k_t \in L^1(\mathbb{R})$ be defined for $t > 0$ by $k_t(x) := t^{-1}K(|y|/t)$, where $K : [0, \infty) \to [0, \infty)$ is a monotone decreasing nonnegative function. This is a family of good kernels in $\mathbb{R}$; see Section 7.8. Show that the following pointwise estimate holds:

$$|K_t * f(x)| \leq Mf(x)\, \|k_1\|_1. \qquad \text{◊}$$

[12] Named after the British mathematicians Godfrey Harold (G. H.) Hardy (1877–1947) and John Edensor Littlewood (1885–1977).

As examples of such kernels, consider the *heat kernel* and the *Poisson kernel* on $\mathbb{R}$, introduced in Section 7.8. In these examples, $K(x) = e^{-|x|^2/4}/\sqrt{4\pi}$ for the heat kernel, and $K(x) = 1/(1+|x|^2)$ for the *Poisson kernel*. A similar argument to the one that gives the Lebesgue Differentiation Theorem implies that if k_t is the heat kernel or the Poisson kernel, then $\lim_{t\to 0} k_t * f(x) = f(x)$ for almost every $x \in \mathbb{R}$ and for every $f \in L^1(\mathbb{R})$. See **[Pin**, Sections 3.5.1 and 3.5.2]. We already knew this fact for continuous functions, since in that case uniform convergence to f was proven in Chapter 7. $\Diamond$

Exercise 9.48. We outline an existence proof (but not a constructive proof) that there is a continuous periodic function whose Fourier series diverges at some point. See the project in Section 2.5 for a constructive proof.

Theorem 9.49 (Uniform Boundedness Principle)**.** *Let W be a family of bounded linear operators $T : X \to Y$ from a Banach space X into a normed space Y, such that for each $x \in X$, $\sup_{T\in W} \|Tx\|_Y < \infty$. Then the operators are uniformly bounded: there is a constant $C > 0$ such that for all $T \in W$ and all $x \in X$, $\|Tx\|_Y \le C\|x\|_X$.*

Now let X be the Banach space $C(\mathbb{T})$ of all real-valued continuous functions of period 2π with uniform norm, and let $Y = \mathbb{C}$. For $f \in C(\mathbb{T})$, define $T_N(f) := S_N f(0) \in \mathbb{C}$, where $S_N f$ is the N^{th} partial Fourier sum of f. Then $S_N f = D_N * f$ where D_N is the periodic Dirichlet kernel. Show that if $|T_N f| \le C_N \|f\|_\infty$, then necessarily $C_N \ge c\|D_N\|_{L^1(\mathbb{T})}$. But as we showed in Chapter 4, $\|D_N\|_{L^1(\mathbb{T})} \approx \log N$. It follows that the operators T_N cannot be uniformly bounded, so there must be some $f \in C(\mathbb{T})$ such that $\sup_{N\ge 0} |S_N f(0)| = \infty$. The partial Fourier sums of this function f diverge at $x = 0$. $\Diamond$

9.5. Haar vs. Fourier

We give two examples to illustrate how the Haar basis can outperform the Fourier basis. The first deals with localized data. The second is related to the unconditionality of the Haar basis in $L^p(\mathbb{R})$ for all $1 < p < \infty$, in contrast to the trigonometric basis which is not an unconditional basis for $L^p([0,1]$ except when $p = 2$. We take the opportunity to introduce some operators that are important in harmonic

analysis: the *martingale transform*, the *square function*, and *Petermichl's shift operator*[13]. We deduce the unconditionality in L^p of the Haar basis from boundedness properties of these dyadic operators.

9.5.1. Localized data. The first example is a caricature of the problem: What is the most localized "function" we could consider? The answer is the delta distribution. If we could find the Fourier series of a periodic delta distribution, we would see that it has a very slowly decreasing tail that extends well beyond the highly localized support of the delta function. However, its Haar transform is very localized; although the Haar transform still has a tail, the tail decays faster than that of the Fourier series. We try to make this impressionistic comment more precise in what follows.

Consider the following approximation of the delta distribution:

$$f_N(x) = 2^N \chi_{[0,2^{-N}]}.$$

Each of these functions has mass 1, and they converge in the sense of distributions to the delta distribution:

$$\lim_{N\to\infty} T_{f_N}(\phi) = \lim_{N\to\infty} \int f_N(x)\phi(x)\,dx = \phi(0) = \delta(\phi),$$

by the Lebesgue Differentiation Theorem for continuous functions (Exercise 9.40).

Exercise 9.50. Compute the Fourier transform of f_N. Show that if we view f_N as a periodic function on $[0,1)$, with M^{th} partial Fourier sum $S_M(f_N)(x) = \sum_{|m|\le M} \widehat{f_N}(m)\, e^{2\pi i m x}$, then the following rate of decay holds: $\|f_N - S_M(f_N)\|_{L^2([0,1))} \sim 1/\sqrt{M}$. ♢

We want to compare to the Haar decomposition, working on the interval $[0,1)$. We have to be a little careful in view of the next exercise.

Exercise 9.51. Show that the set $\{h_I\}_{I\in\mathcal{D}([0,1])}$ is not a complete set in $L^2([0,1])$. What are we missing? Can you complete the set? You must show that the set is now complete. ♢

[13]Named after German mathematician Stefanie Petermichl (born 1971).

The Haar basis on the interval $[0,1)$ consists of all the Haar functions indexed on dyadic subintervals of $[0,1)$ and the characteristic function $\chi_{[0,1)}$ of the interval $[0,1)$. The function $\chi_{[0,1)}$ is orthonormal to all the Haar functions selected, so it must be part of an orthonormal basis.

Exercise 9.52. Compute the Haar coefficients of f_N viewed as a function on $[0,1)$. Check that the partial Haar sum $P_M(f_N)$, defined by $P_M(f_N) := \sum_{0\le j<M} Q_j(f_N) + P_0(f_N)$, decays exponentially in M: $\|f_N - P_M(f_N)\|_{L^2([0,1))} = 2^{-M/2}$. ◊

The exponential decay rate seen in Exercise 9.52 is much better than the square-root decay in Exercise 9.50. Suppose we want to approximate F_N with an L^2 error of magnitude less than 10^{-5}. In the Fourier case we would need a partial Fourier sum of order $M \sim 10^{10}$. In the Haar case it suffices to consider $M = 10\log_2 10 < 40$. However, beware: to move from trigonometric polynomials of degree M to degree $M+1$, we just add two functions $e^{\pm 2\pi i(M+1)}$, while to move from generation M to generation $M+1$, we need 2^M Haar functions, so this calculation is deceptive.

However, the localization properties of the Haar functions allow us to worry only about the Haar functions that are supported near where the action occurs in the function. For example, if the function is constant on an interval, then the Haar coefficients corresponding to the Haar functions supported in the interval vanish, because they have the zero integral property. This is not the case with the trigonometric functions, whose support spreads over the whole real line. In the example discussed, the function f_N is supported on the interval $[0, 2^{-N})$ and is constant there. So in fact the number of Haar functions in a given generation whose support intersects the support of f_N and which are not completely inside it is just one. In both the Fourier and the Haar cases, the parameter M counts the same number of basis functions contributing to the estimate. Now the difference in the estimate of the error is dramatic when comparing the rates of convergence. The same phenomenon is observed for compactly supported wavelets other than the Haar wavelet.

See the project in Section 10.6 for more on linear and nonlinear approximations. The lesson from this example is that when there is good localization, the wavelet basis will perform better than the Fourier basis.

9.5.2. Unconditional bases and dyadic operators. The Haar basis is an *unconditional basis* for $L^p(\mathbb{R})$, for $1 < p < \infty$. See the Appendix for the formal definition. Informally, we can approximate a function in the L^p norm with an infinite linear combination of Haar functions (basis), and the order of summation doesn't matter (conditional convergence). Further, the coefficient of h_I must be $\langle f, h_I\rangle$, and we can recover the L^p norm of the function from knowledge about the *absolute value* of these coefficients, that is, using some formula involving only $|\langle f, h_I\rangle|$ for each I. No information about the sign or argument of $\langle f, h_I\rangle$ is necessary. In particular, if $f \in L^p(\mathbb{R})$ and

$$f(x) = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I(x),$$

then the new functions defined by

$$T_\sigma f(x) := \sum_{I\in\mathcal{D}} \sigma_I \langle f, h_I\rangle h_I(x), \quad \text{where } \sigma_I = \pm 1, \tag{9.15}$$

are also in $L^p(\mathbb{R})$ and their norms are comparable to that of f. We will make the statement precise in the next theorem.

Definition 9.53. For a given sequence $\sigma = \{\sigma_I\}_{I\in\mathcal{D}}$, the operator T_σ in equation (9.15) is called the *martingale transform.* ♢

Theorem 9.54. *Let $\sigma = \{\sigma_I\}_{I\in\mathcal{D}}$ be a sequence of plus and minus ones, and let T_σ be its associated martingale transform. There exist constants c, $C > 0$ depending only on $1 < p < \infty$, such that for all choices σ of signs and for all $f \in L^p(\mathbb{R})$ and for $1 < p < \infty$,*

$$c\|f\|_p \le \|T_\sigma f\|_p \le C\|f\|_p. \tag{9.16}$$

The first inequality is deduced from the second one with $c = 1/C$ after noting that $T_\sigma(T_\sigma f) = f$. The second we will deduce from similar inequalities valid for the dyadic square function.

Let us illustrate Theorem 9.54 in the case $p = 2$. We know that the Haar functions provide an orthonormal basis in $L^2(\mathbb{R})$. In

particular, Plancherel's Identity holds:

$$\|f\|_2^2 = \sum_{I\in\mathcal{D}} |\langle f, h_I\rangle|^2.$$

To compute the L^2 norm of the function, we add the squares of the absolute values of the Haar coefficients. Since each $|\sigma_I|^2 = 1$, we have

$$\|T_\sigma f\|_2 = \|f\|_2,$$

and so inequality (9.16) holds with $c = C = 1$. Therefore the martingale transform is an isometry in $L^2(\mathbb{R})$. In $L^p(\mathbb{R})$ we do not have an isometry, but we have the next best thing, which is encoded in the norm equivalence given by inequalities (9.16). We state another norm equivalence for yet another operator, the (nonlinear) *dyadic square function* S^d. It will imply Theorem 9.54.

Definition 9.55. The *dyadic square function* $S^d f$ is defined for functions $f \in L^2(\mathbb{R})$ by

$$S^d(f)(x) = \Big(\sum_{I\in\mathcal{D}} \frac{|\langle f, h_I\rangle|^2}{|I|}\chi_I(x)\Big)^{1/2}. \tag{9.17}$$ ◊

It turns out that in the case of the dyadic square function, we also have a norm equivalence in $L^p(\mathbb{R})$.

Theorem 9.56. *There exist positive constants c_p and C_p, depending only on $1 < p < \infty$, such that for all $f \in L^p(\mathbb{R})$*

$$c_p\|f\|_p \le \|S^d(f)\|_p \le C_p\|f\|_p. \tag{9.18}$$

This norm equivalence can be considered as an L^p substitute for Plancherel's Identity for Haar functions. Theorem 9.56 tells us that we can recover the L^p norm of f from the L^p norm of $S^d(f)$. In the definition of the dyadic square function only the absolute values of the Haar coefficients of f are used, and so that information is all that is required to decide whether f is in $L^p(\mathbb{R})$. The case $p = 2$ is left as an exercise. The case $p \neq 2$ can be found in [**Graf08**, Appendix C2] and is based on a beautiful probabilistic result called Khinchine's Inequality[14] that is worth exploring and understanding. See the project in Section 9.8.

[14] Named after Russian mathematician Aleksandr Yakovlevich Khinchine (1894–1959).

Exercise 9.57 (*The Dyadic Square Function Is an Isometry on* $L^2(\mathbb{R})$)**.** Verify that $\|S^d(f)\|_2 = \|f\|_2$. ◊

Exercise 9.58 (*The Dyadic Square Function in Terms of Difference Operators*)**.** Show that $\big[S^d(f)(x)\big]^2 = \sum_{j\in\mathbb{Z}} |Q_j f(x)|^2$, where Q_j is the difference operator defined in Definition 9.33. Lemma 9.35 will be useful. ◊

Proof of Theorem 9.54. The definition of the dyadic square function (9.17) implies that for all choices of signs σ,

$$S^d(f) = S^d(T_\sigma f).$$

Theorem 9.56 implies the second inequailty in the norm equivalence (9.16), because

$$c_p\|T_\sigma f\|_p \le \|S^d(T_\sigma f)\|_p = \|S^d(f)\|_p \le C_p\|f\|_p.$$

Dividing by $c_p > 0$, we conclude that $\|T_\sigma f\|_p \le C\|f\|_p$, as desired. The proof of the lower bound is similar. □

We can argue in a similar fashion to show that the following dyadic operator is bounded on $L^p(\mathbb{R})$ for all $1 < p < \infty$.

Definition 9.59. *Petermichl's shift operator* is defined for each function $f \in L^2(\mathbb{R})$ by

$$Ш f(x) = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle \frac{1}{\sqrt{2}}\big(h_{I_r}(x) - h_{I_l}(x)\big). \qquad ◊$$

The operator Ш (a letter in the Cyrillic alphabet, pronounced "sha") was introduced by Stefanie Petermichl **[Pet]** in connection with the Hilbert transform (see Chapter 12).

Exercise 9.60 (*Petermichl's Shift Operator*)**.** (i) Let $\operatorname{sgn}(I, \tilde{I}) = 1$ when I is the right daughter of $\tilde{I}$, and $\operatorname{sgn}(I, \tilde{I}) = -1$ when I is the left daughter. Show that $Ш f(x) = \sum_{I\in\mathcal{D}} \big(\operatorname{sgn}(I, \tilde{I})/\sqrt{2}\big)\langle f, h_{\tilde{I}}\rangle h_I(x)$.

(ii) Show that $\|Ш f\|_2 = \|f\|_2$.

(iii) Show that there are constants c_p, $C_p > 0$ such that for all $f \in L^p(\mathbb{R})$, $c_p\|f\|_p \le \|Ш f\|_p \le C_p\|f\|_p$. **Hint:** Calculate $SШ f$, and then argue as we did for the martingale transform in the proof of Theorem 9.54. ◊

For general wavelets we also have *averaging* and *difference operators*, P_j and Q_j, and a corresponding *square function*. The same norm equivalence (9.18) holds in $L^p(\mathbb{R})$. As it turns out, wavelet bases provide unconditional bases for a whole zoo of function spaces (Sobolev spaces[15], Hölder spaces, etc.). See for example [**Dau92**, Chapter 9].

The trigonometric system is an orthonormal basis for $L^2([0,1])$. However it does not provide an unconditional basis in $L^p([0,1])$ for $p \neq 2$; see [**Woj91**, II.D.9]. There is a square function that plays the same rôle that the dyadic square function plays for the Haar basis:

$$Sf(\theta) := \Big(\sum_{j \geq 0} |\Delta_j f(\theta)|^2 \Big)^{1/2}.$$

Here $\Delta_j f$ is the projection of f onto the subspace of trigonometric polynomials of degree n where $2^{j-1} \leq |n| < 2^j$ for $j \geq 1$:

$$\Delta_j f(\theta) := \sum_{2^{j-1} \leq |n| < 2^j} \widehat{f}\,(j)\, e^{2\pi i n x} \quad \text{and} \quad \Delta_0 f(\theta) = \widehat{f}(0).$$

It is true that $\|f\|_p$ is comparable to $\|S(f)\|_p$ in the sense of inequality (9.18). We are allowed to change the signs of the Fourier coefficients of f on *the dyadic blocks* of frequency. Denoting by $\mathcal{T}_\delta f$ the function reconstructed with the modified coefficients, that is,

$$\mathcal{T}_\delta f(\theta) := \sum_{j \geq 0} \delta_j \Delta_j f(\theta), \qquad \delta_j = \pm 1,$$

we have $S(\mathcal{T}_\delta f) = S(f)$, and so their L^p norms are the same and are both equivalent to $\|f\|_p$. But there is no guarantee that the same will be true if we change some but not all signs *inside* a given dyadic block! In that case, $S(f)$ does not have to coincide with $S(\mathcal{T}_\delta f)$.

The study of square functions is known as *Littlewood–Paley theory*[16]; it is a widely used tool in harmonic analysis. For an introduction to dyadic harmonic analysis, see the lecture notes by the first author [**Per01**]. The books [**Duo**], [**Graf08**], [**Ste70**], and [**Tor**] all discuss this important topic.

[15] Named after Russian mathematician Sergei Lvovich Sobolev (1908–1989).

[16] Named after British mathematicians John Edensor Littlewood (1885–1977) and Raymond Edward Alan Christopher Paley (1907–1933).

9.6. Project: Local cosine and sine bases

Local cosine (or sine) bases are orthonormal bases for $L^2(\mathbb{R})$, consisting of functions that are both smooth and compactly supported. In this sense they could be said to defeat the Balian–Low Theorem (Theorem 9.11), which implies that the functions in a Gabor basis for $L^2(\mathbb{R})$ cannot be both smooth and compactly supported. Local cosine bases were first discovered by Malvar [**Malv**] and are sometimes called Malvar–Wilson wavelets[17]. See [**MP**, Section 2.3], [**CMe**], [**HW**, Chapter 1], and [**JMR**, Chapter 6].

(a) Understand the construction of local cosine and sine bases, using these sources or others.

(b) Why is it useful to have a basis with the properties stated above? Find some applications in the literature.

(c) Find some software implementations of local cosine bases, and explore how they work. Or write your own implementation.

(d) Clarify the relationship of local cosine and sine bases with the Balian–Low Theorem (Theorem 9.11) and Heisenberg's Uncertainty Principle (Theorem 8.44).

9.7. Project: Devil's advocate

Investigate whether the Haar series $\sum_{I\in\mathcal{D}}\langle\chi_{[0,1)},h_I\rangle h_I(x)$ for the function $\chi_{[0,1)}$ converges to $\chi_{[0,1)}$ pointwise (yes), uniformly (yes), in $L^1(\mathbb{R})$ (no), or in $L^2(\mathbb{R})$ (yes). See also Section 9.4.3.

(a) Show that $\langle\chi_{[0,1)},h_I\rangle = -1/\sqrt{2^n}$ if $I = I^n := [0,2^n)$, for $n \geq 1$. Show that for $I \in \mathcal{D}$ and for $I \neq I_n$ for every $n \geq 1$, $\langle\chi_{[0,1)},h_I\rangle = 0$. Hence show that for each $x \in \mathbb{R}$,

$$\sum_{I\in\mathcal{D}}\langle\chi_{[0,1]},h_I\rangle h_I(x) = \sum_{n=1}^{\infty}\frac{1}{2^n}\left[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\right]. \tag{9.19}$$

[17] Named after the Brazilian engineer Henrique Malvar (born 1957) and the American physicist Kenneth Wilson (born 1936). Wilson received the 1982 Nobel Prize in physics.

(b) Show that the right-hand side of equation (9.19) is equal to $\chi_{[0,1)}(x)$ for each $x \in \mathbb{R}$. Consider separately the cases $x < 0$, $0 \le x < 1$, and $2^k \le x < 2^{k+1}$ for $k \ge 0$. For instance for $2^k \le x < 2^{k+1}$,

$$\sum_{n=1}^{\infty} \frac{1}{2^n} \left[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\right] = -\frac{1}{2^k} + \sum_{n=k+1}^{\infty} \frac{1}{2^n} = 0.$$

(c) The partial Haar sums f_N of f are defined to be $f_N(x) := \sum_{n=1}^{N} \frac{1}{2^n} \left[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\right]$. Show that $f_N(x) = 1 - 2^{-N}$ if $x \in [0,1)$, $f_N(x) = -2^{-N}$ if $x \in [1,2^N)$, and $f_N(x) = 0$ otherwise. Show that $\int_{\mathbb{R}} f_N(x)\,dx = 0$ for all $N \in \mathbb{N}$. Show that $\{f_N\}_{N\in\mathbb{N}}$ converges uniformly to $\chi_{[0,1)}(x)$ on $\mathbb{R}$. Explain why we cannot interchange the limit and the integral, despite the uniform convergence of f_N. Indeed, if we interchange them, we will conclude that

$$\lim_{N\to\infty} \int_{\mathbb{R}} f_N(x)\,dx = 0 \neq 1 = \int_{\mathbb{R}} \chi_{[0,1)}(x)\,dx = \int_{\mathbb{R}} \lim_{N\to\infty} f_N(x)\,dx.$$

Hint: The real line $\mathbb{R}$ is not compact.

(d) Show that $\{f_N\}_{N\in\mathbb{N}}$ does not converge to $\chi_{[0,1)}$ in $L^1(\mathbb{R})$. Show that $\{f_N\}_{N\in\mathbb{N}}$ converges to $\chi_{[0,1)}$ in $L^2(\mathbb{R})$.

(e) The Lebesgue Dominated Convergence Theorem (see Chapter 2 and Theorem A.59) implies that the interchange of limit and integral would be valid if there were an integrable function g such that $|f_N(x)| \le g(x)$ for all $N \in \mathbb{N}$. Deduce constructively that there can be no such dominating function $g \in L^1(\mathbb{R})$. **Hint:** Calculate explicitly $g_1(x) = \sup_{N\ge 1} |f_N(x)|$. Any dominant function g must be larger than or equal to g_1. Verify that g_1 is not integrable.

9.8. Project: Khinchine's Inequality

This project deals with *Khinchine's Inequality*, which is the key to proving the norm equivalence in $L^p(\mathbb{R})$ between the square function Sf and the function f (Theorem 9.56). First one needs to get acquainted with the *Rademacher functions*[18] and to be comfortable with the fact that these functions are independent random variables. You will find enough to get you started in [**Graf08**, Appendix C] and [**Woj91**, Section I.B.8]. See also [**Ste70**, Appendix] and [**Zyg59**].

[18] Named after German mathematician Hans Adolph Rademacher (1892–1969).

(a) The Rademacher functions $\{r_n(t)\}_{n=1}^{\infty}$ are defined on $[0,1]$ by $r_n(t) = \operatorname{sgn}\big(\sin(2^n t\pi)\big)$. Show that this definition is equivalent to the following recursive definition: $r_1(t) = 1$ if $0 \le t \le 1/2$, $r_1(t) = -1$ if $1/2 < t \le 1$, and given $r_n(t)$, then $r_{n+1}(t) = 1$ on the left half and -1 on the right half of each interval where $r_n(t)$ is constant.

(b) Show that the Rademacher system is an orthonormal system but *not* a complete orthonormal system.

(c) Verify that $\{r_n(t)\}_{n=1}^{\infty}$ is a sequence of mutually independent random variables on $[0,1]$, each taking value 1 with probability $1/2$ and value -1 with probability $1/2$. The mutual independence amounts to checking that for all integrable functions f_j we have

$$\int_0^1 \prod_{j=0}^{n} f_j(r_j(t))\,dt = \prod_{j=0}^{n} \int_0^1 f_j(r_j(t))\,dt.$$

(d) Prove *Khinchine's Inequality* (or find a proof in the literature and make sure you understand it): for all square summable sequences of scalars $\{a_n\}_{n\ge1}$ and for every p with $0 < p < \infty$, we have similarity of the L^2 and L^p norms: $\big\|\sum_{n\ge1} a_n r_n\big\|_{L^p([0,1])} \sim \big\|\sum_{n\ge1} a_n r_n\big\|_{L^2([0,1])}$.

(e) Prove a similar inequality for *lacunary* sequences of trigonometric functions: If $\{n_k\}_{k\ge1}$ is a sequence of natural numbers such that $\inf_{k\ge1}(n_{k+1}/n_k) = \lambda > 1$, then for all square summable sequences of scalars $\{a_k\}_{k\ge1}$ and for every p with $0 < p < \infty$, we have $\big\|\sum_{k\ge1} a_k e^{in_k\theta}\big\|_{L^p([0,1])} \sim \big\|\sum_{k\ge1} a_k e^{in_k\theta}\big\|_{L^2([0,1])}$.