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

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,

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

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 $L^2(\mathbb{R})$. ◊

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. Gabor functions $g_{n,k}$ 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*[1] *function*, and the basis a *Gabor basis*. In 1946,

[1] Named after Dennis Gabor (1900-1979), a Hungarian electrical engineer and inventor, most notable for inventing holography, for which he later received the Nobel Prize in Physics in 1971.

FIGURE 9.2. Gabor functions $g_{n,k}$ for some values of n and k, and for g a smooth window.

Gabor considered systems of this type and proposed using them in communication theory [**Gab**].

The polarization formula (7.27) 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). ♢

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. ♢

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[2] Theorem). *For $g \in L^2(\mathbb{R})$, if $\{g_{n,k}\}_{n,k\in\mathbb{Z}}$ is an orthonormal basis, then either*

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

EXERCISE 9.6. (*Examples of Balian–Low*) Verify the Balian–Low Theorem for the two examples discussed so far: $g(x) = e^{2\pi inx}\chi_{[0,1)}(x)$ and $g(x) = e^{-i\pi x}\operatorname{sinc} x$. ♢

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**]. See also Project 12.8

EXERCISE 9.7. Show that the Balian–Low Theorem implies that a Gabor function cannot be both smooth and compactly supported. ♢

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 given by

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

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 one. **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. Where $*$ denotes convolution on $\mathbb{R}$. ♢

---

[2]Named after for Roger Balian, a French physicists (1933-), and Francis E. Low, and American theoretical physicist (1921-2007).

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 $\widehat{\psi_{j,k}}(\xi)$. ◊

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. ◊

EXERCISE 9.11. Show that $\{h_{j,k}\}$ is an orthonormal set:

$$\langle h_{j,k}, h_{j',k'}\rangle = \begin{cases} 1, & \text{if } j = j' \text{ and } k = k'; \\ 0, & \text{otherwise.} \end{cases}$$

First show that the functions $h_{j,k}$ have zero integral: $\int h_{j,k} = 0$. ◊

An important part of wavelet theory is the search for smoother wavelets. The Haar function is discontinuous and perfectly localized in time, therefore not perfectly localized in frequency. However the compact support of the Haar function translates into C^∞ Fourier transform (decay in space implies smoothness on Fourier side).

EXERCISE 9.12. Find the Fourier transform of the Haar function $h(x)$. ◊

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 $L^2(\mathbb{R})$. $\lozenge$

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 (7.27). It can be seen that on Fourier side we are dealing with a windowed 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[3] modulo 2^j with the interval $[0,2^j)$. The collection of double-paned windows F_j, provide a partition of $\mathbb{R}\backslash\{0\}$. On each double-paned window, the trigonometric functions

$$\frac{1}{2^{j/2}}e^{2\pi ikx2^{-j}}\chi_{F_j}(x), \quad k\in\mathbb{Z},$$

form an orthonormal basis of $L^2(F_j)$. $\lozenge$

The Shannon wavelets are perfectly localized in frequency, therefore not in space. The compact support on frequency side translates into smoothness (C^∞) of the Shannon wavelet. Therefore the Shannon wavelet provides an example of a C^∞ wavelets that does not have compact support.

Can one find compactly supported wavelets that are smooth? Compactly supported wavelets with arbitrary (but finite) smoothness were constructed by I. Daubechies[4] in a fundamental paper in wavelet theory [**Dau88**]. However it is impossible to construct a C^∞ and compactly supported wavelet.

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 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 between length of the filter (hence length of the support) and smoothness. 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 ($\int\psi=0$).

---

[3]The intervals $[-1/2,0)\cup[1/2,1)$ are not congruent modulo 1 to the interval $[0,1)$.

[4]Ingrid Daubechies, 1954–, Belgian mathematician.

Let

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

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 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*[5] *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

[5]Named after Alberto Calderón, an Argentinian mathematician (1920-1998).

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.

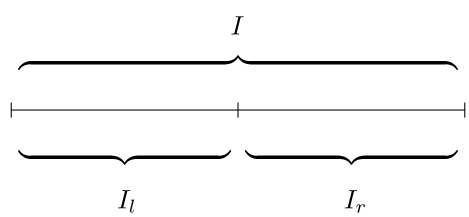

FIGURE 9.3. Parent dyadic interval I and its children I_l and I_r.

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 $J \subset I$. ♢

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

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

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

$\diamondsuit$

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 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[6]:

$$f = \sum_{I\in\mathcal{D}} \langle f, h_I\rangle h_I. \tag{9.7}$$

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.

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?[7]

---

[6]Recall that equation (9.7) holds in the L^2-sense if $\|f - \sum_{I\in\mathcal{D}}\langle f, h_I\rangle h_I\|_2 = 0$.

[7]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

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 $[0,1) \subset I_l^n = [0,2^{n-1})$. So if $x \in [0,1)$, then $h_{I^n}(x) = -1/\sqrt{2^n}$. This observation and Exercise 9.19 show that

$$\langle \chi_{[0,1]}, h_I\rangle = \begin{cases} -\frac{1}{\sqrt{2^n}}, & \text{if } I = I^n,\ n \geq 1; \\ 0, & \text{otherwise.}\end{cases}$$

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:

$$\begin{aligned}\sum_{I\in\mathcal{D}} \langle \chi_{[0,1]}, h_I\rangle h_I(x) &= \sum_{n=1}^{\infty} \langle \chi_{[0,1]}, h_{I^n}\rangle h_{I^n}(x) \\ &= \sum_{n=1}^{\infty} \frac{1}{2^n}\big[\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)\big]. \qquad (9.8)\end{aligned}$$

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

$$\sum_{n=1}^{\infty} \langle \chi_{[0,1]}, h_{I^n}\rangle h_{I^n}(x) = 0.$$

perfectly natural concern, and furthermore it was very illustrative of the dangers of interchanging limiting operations!

For $0 \le x < 1$**:** We have already mentioned that $[0,1)$ sits on I_l^n, the left half of I^n, for all $n \ge 1$. 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 $1 \le x < 2$**:** This time $[1,2) = I_r^1$, and $[1,2) \subset I_\ell^n$ for all $n \ge 2$. 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 $2 \le x < 2^2$**:** Here $[2,2^2) \cap I^1 = \emptyset$, $[2,2^2) = I_r^2$, and $[2,2^2) \subset I_\ell^n$ for all $n \ge 3$. 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 $2^k \le x < 2^{k+1}$**:** 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 $n \ge k+1$. 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,

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

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 $\mathbb{R}$ (see Theorem **??**) 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*[8] 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:

$$P_j f(x) := \frac{1}{|I|} \int_I f(t)\, dt, \tag{9.9}$$

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. ◊

FIGURE 9.5. Graphs of f and $P_{-1}f$. [***Add the formula for this f, also $n = j = -1$ adjust figure.]

[8]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 162.

FIGURE 9.6. Graphs of f and $Q_{-1}f$. [***$n = j = -1$ adjust figure]

FIGURE 9.7. Graphs of f, $P_{-1}f$, and $P_0 f$. [***$n = j = -1$ adjust figure]

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.

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

$$\langle f, h_I \rangle h_I(x) = \begin{cases} \frac{1}{2}(m_{I_r} f - m_{I_l} f), & \text{if } x \in I_r; \\ -\frac{1}{2}(m_{I_r} f - m_{I_l} f), & \text{if } x \in I_l. \end{cases} \tag{9.11}$$

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

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

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

$$Q_j f(x) = \begin{cases} \frac{1}{2}(m_{I_r} f - m_{I_l} f), & \text{if } x \in I_r; \\ -\frac{1}{2}(m_{I_r} f - m_{I_l} f), & \text{if } x \in I_l. \end{cases}$$

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. □

EXERCISE 9.26. Verify the averaging property in equation (9.12). ◊

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

$$P_N f(x) - P_M f(x) = \sum_{M \le j < N} (P_{j+1} f(x) - P_j f(x)) = \sum_{M \le j < N} Q_j f(x). \tag{9.13}$$

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:

$$\lim_{j\to\infty} \frac{1}{|I_j(x)|} \int_{I_j(x)} f(t)\,dt = f(x).$$

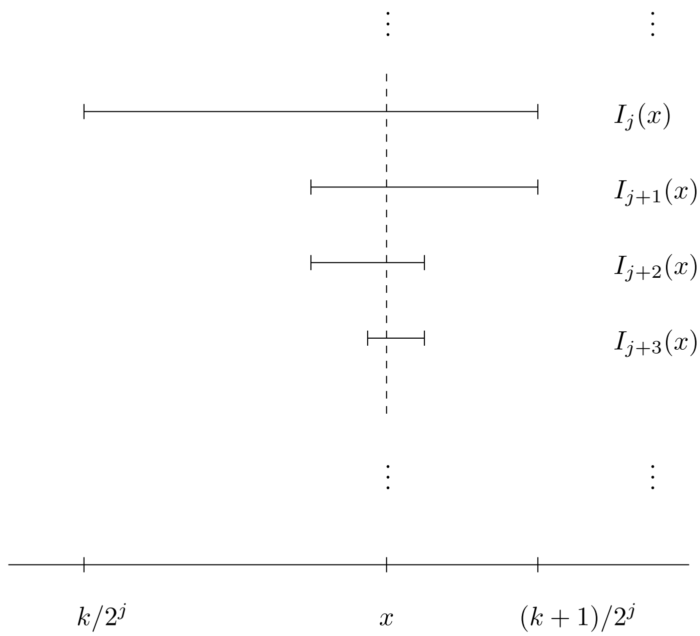

FIGURE 9.8. Part of the tower of nested dyadic intervals $\cdots \supset I_j(x) \supset I_{j+1}(x) \supset I_{j+2}(x) \supset I_{j+3}(x) \supset \cdots$ containing the point x, with $k/2^j \le x \le (k+1)/2^j$. Here j and k are integers.

It turns out that the convergence also holds *almost everywhere*[9] (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 $f \in L^1_{loc}(\mathbb{R})$ (locally integrable), 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].

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

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. ◊

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,*

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

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 $j > N$,

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

Now apply the triangle inequality, and use Lemma 9.31 to conclude that

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

This proves equation (9.14).

Similarly, by Lemma 9.32, we can choose N large enough that for all $j > N$,

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

Now apply the triangle inequality, and use Lemma 9.31 to conclude that

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

This proves equation (9.15). □

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 $j \to -\infty$, and in the other case they converge to the identity operator, as $j \to \infty$.

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 $x \in X$,

$$\|Tx\|_Y \le C\|x\|_X.$$

In particular if the operators $\{T_n\}_{n\ge1}$ are uniformly bounded, that is $\|T_n x\|_Y \le C\|x\|_X$, 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:

$$\|T_n x\|_Y \le \|T_n(x - x_m)\|_Y + \|T_n x_m\|_Y \le C\|x - x_m\|_Y + \|T_n x_m\|_Y.$$

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

$$|T_N f| \le C_N \|f\|_\infty,$$

then necessarily $C_N \ge c\|D_N\|_{L^1(\mathbb{T})}$. But we showed in Chapter 4 that $\|D_N\|_{L^1(\mathbb{T})} \approx \log N$. This implies that the operators T_N cannot be uniformly bounded, so there must exist $f \in C(\mathbb{T})$ such that

$$\sup_{N\ge0} |S_N f(0)| = \infty.$$

Therefore for this function f the partial Fourier sums do not converge at $x = 0$. $\Diamond$

PROOF OF LEMMA 9.31. For $x \in I \in \mathcal{D}_j$,

$$|P_j f(x)|^2 = \left|\frac{1}{|I|}\int_I f(t)\,dt\right|^2 \le \frac{1}{|I|^2}\left(\int_I 1^2\,dt\right)\left(\int_I |f(t)|^2\,dt\right) = \frac{1}{|I|}\int_I |f(t)|^2\,dt.$$

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. □

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

$$\begin{aligned} |P_{-j}g(x)| &= \frac{1}{2^j}\int_0^K |g(t)|dt \\ &\leq \frac{1}{2^j}\left(\int_0^K 1^2\, dt\right)^{1/2}\left(\int_0^K |g(t)|^2\, dt\right)^{1/2} \\ &\leq \frac{1}{2^j}\sqrt{K}\|g\|_2, \end{aligned}$$

where the last inequality is another application of the Cauchy–Schwarz inequality. The same inequality holds for $x < 0$. If $|x| \geq 2^j$ then $P_{-j}g(x) = 0$.

We can now estimate the L^2-norm of $P_{-j}g$:

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

By choosing N large enough, we can make $2^{-N+1}K\|g\|_2^2 < \varepsilon^2$. That is, given $\varepsilon > 0$, there is an $N > 0$ such that for all $j > N$,

$$\|P_{-j}g\|_2 \leq \varepsilon.$$

This proves equation (9.14) 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 $|y - x| \leq \delta$ for all $y \in I$, and

$$|P_j g(x) - g(x)| \leq \frac{1}{|I|}\int_I |g(y) - g(x)|\, dy \leq \frac{\varepsilon}{\sqrt{2^{M+1}}}.$$

Squaring and integrating over $\mathbb{R}$, we get

$$\int_{\mathbb{R}} |P_j g(x) - g(x)|^2 dx = \int_{-2^M}^{2^M} |P_j g(x) - g(x)|^2 dx < \frac{\varepsilon^2}{2^{M+1}} \int_{-2^M}^{2^M} 1\, dx = \varepsilon^2.$$

Notice that if $|x| > 2^M$, then for $n > N \geq M$, $P_j g(x)$ is the average over an interval $I \in \mathcal{D}_j$ that is completely outside the support of g. For such x and j, $P_j g(x) = 0$, 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 (9.15) 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 $\|f\chi_{\{x\in\mathbb{R}:|x|>K\}}\|_2 \leq \varepsilon/3$. Next we recall that on compact intervals, the continuous functions are dense in $L^2([-K,K])$; see Theorem 2.70. (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 $\|(f-g_1)\chi_{[-K,K]}\|_2 \leq \varepsilon/3$. 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

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

By the triangle inequality,

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

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 $x \in \mathbb{R}$,

$$|P_j f(x) - f(x)| < \varepsilon.$$

◊

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? ◊

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

$$S_M(f_N)(x) = \sum_{|m|\le M} \widehat{f_N}\,(m) e^{2\pi i m x},$$

then

$$\|f_N - S_M(f_N)\|_{L^2([0,1))} \sim M^{-1/2}.$$

$\diamondsuit$

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 $N = 1$. $\diamondsuit$

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 $M = 10\log_2 10 < 40$. [****careful to move from trig. poly. of degree M to degree $M+1$ we just add two functions $e^{\pm 2\pi i(M+1)}$, however to move from generation M to generation $M+1$ we need 2^M Haar functions, so this calculation might be deceptive.].

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. $\diamondsuit$

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, martingale transforms, and square functions

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,

$$c\|f\|_p \le \|T_\sigma f\|_p \le C\|f\|_p. \tag{9.18}$$

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 dyadic square function, one has the following norm equivalence in L^p:

$$c_p\|f\|_p \le \|S^d(f)\|_p \le C_p\|f\|_p. \tag{9.20}$$

It can be considered as an L^p substitute for Plancherel for Haar functions. Inequalities (9.20) tells us that we can "recover" the L^p-norm of f from the L^p-norm of $S^d(f)$, note that in the definition of the the dyadic square function only the absolute values of the Haar coefficients of f are used, hence that information is all that is required to decide whether f is in $L^p(\mathbb{R})$ or not.

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

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

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 $p \neq 2$. There is a square function that plays the same role as the dyadic square function plays for the Haar basis, 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

$$\Delta_j f(x) = \sum_{2^j \le |n| < 2^{j+1}} \widehat{f}(j) e^{2\pi i n x}.$$

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,

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

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 $L^p([0,1])$. [****simple example will be nice!]

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].

[*** Give references. Perhaps give some of the uses of LP theory. Boundedness of solutions of PDE, for example.]