

# 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.

### 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  $\chi_A$  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

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

form an orthonormal basis for  $L^2(\mathbb{R})$ . They give us the so-called *windowed Fourier transform*, defined by

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

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

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

**Figure 9.1.** 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

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

### 9.1. The windowed Fourier transform, and Gabor bases 241

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

$$f(x) = \begin{cases} 1 - |x| & -1 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}.$$

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  $[k/2, (k + 1)/2)$ ?  $\diamond$

To avoid Gibb's phenomenon, smoother windows are desirable. The sharp windows given by  $\chi_{[0,1]}$  and its modulated integer translates  $e^{2\pi inx} \chi_{[k,k+1]}(x)$  can be replaced by a *smooth window*  $g$  and its modulated integer translates

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

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<sup>1</sup> function*, and the basis a *Gabor basis*. In 1946, 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,

$$\{\psi_n\}_{n \in \mathbb{N}} \text{ is an o.n. basis in } L^2(\mathbb{R}) \iff \{\widehat{\psi}_n\}_{n \in \mathbb{N}} \text{ is an o.n. basis in } L^2(\mathbb{R}).$$

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:

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

**Exercise 9.3.** Verify equation (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  $L^2(\mathbb{R})$ .*

---

<sup>1</sup>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.

### 9.1. The windowed Fourier transform, and Gabor bases 243

**Example 9.4.** Since  $g = \chi_{[0,1]}$  generates a Gabor basis, so does its Fourier transform

$$\widehat{g}(\xi) = (\chi_{[0,1]})^\wedge(\xi) = e^{-i\pi\xi} \frac{\sin(\pi\xi)}{\pi\xi} = e^{-i\pi\xi} \operatorname{sinc}(\xi).$$

This is an example of a differentiable window  $\widehat{g}$ , in contrast with our first window which was not even continuous. However, unlike  $g$ ,  $\widehat{g}$  is not compactly supported.  $\diamond$

Which other functions  $g \in L^2(\mathbb{R})$  generate a Gabor basis? Can we find a Gabor function that is simultaneously smooth and compactly supported? The limitations of the Gabor analysis are explained by the following result.

**Theorem 9.5** (Balian–Low<sup>2</sup> Theorem). *If  $g \in L^2(\mathbb{R})$ , and  $\{g_{n,k}\}_{n,k \in \mathbb{Z}}$  is an orthonormal basis, then either*

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

**Exercise 9.6.** (*Examples 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  $g(x) = e^{-i\pi x} \operatorname{sinc} x$ .  $\diamond$

A proof of Theorem 9.5 can be found in [Dau92, p. 108, Theorem 4.1.1]. The theorem implies that a Gabor window or bell cannot be simultaneously compactly supported and smooth. The first example,  $g(x) = \chi_{[0,1]}(x)$ , is perfectly localized in time but is not even continuous, and the second example,  $g(x) = e^{-i\pi x} \operatorname{sinc}(x)$ , is the opposite. In particular the slow decay of the sinc function reflects the lack of smoothness of the characteristic function  $\chi_{[0,1]}$ . This phenomenon is an incarnation of *Heisenberg’s Uncertainty Principle*. However, if the exponentials are replaced by appropriate sines and cosines one can obtain Gabor-type bases with smooth bell functions. These are the so-called *local sine and cosine bases*, described by Coifman and Meyer [CMe], but first discovered by Malvar [Malv]. For a very good discussion see the text by Hernández and Weiss [HW].

**Exercise 9.7.** Show that the Balian–Low Theorem implies that a Gabor function cannot be both smooth and compactly supported.  $\diamond$

<sup>2</sup>Named for Roger Balian, a French physicist (1933–), and Francis E. Low, an 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

$$(9.3) \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \quad j, k \in \mathbb{Z},$$

forms an orthonormal basis of  $L^2(\mathbb{R})$ .

**Exercise 9.8.** (*Fourier Transform of a Wavelet*) Suppose  $\psi \in L^2(\mathbb{R})$ . Use the time–frequency dictionary to compute  $\widehat{\psi_{j,k}}(\xi)$ .  $\diamond$

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

The *orthogonal wavelet transform* is given by

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

and the following *reconstruction formula* holds in the  $L^2$ -sense,

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

The oldest example of a wavelet basis is the *Haar basis*, introduced by Alfréd Haar in 1910 [**Haa**]. He succeeded in finding an orthonormal basis on  $L^2([0, 1])$  that, unlike the trigonometric basis, provides uniform convergence of its partial sums for continuous functions.

**Example 9.9.** (*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  $\{h_{j,k}(x) = 2^{j/2}h(2^jx - k)\}_{j,k \in \mathbb{Z}}$  is an orthonormal basis for  $L^2(\mathbb{R})$ . We will study this example in depth in Section 9.3.  $\diamond$

**Exercise 9.10.** 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:  $\int h_{j,k} = 0$ .  $\diamond$

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  $C^\infty$  Fourier transform (decay in space implies smoothness on Fourier side).

**Exercise 9.11.** Find the Fourier transform of the Haar function  $h(x)$ .  $\diamond$

**Example 9.12.** (*The Shannon Basis*) At the other end of the spectrum, one finds the *Shannon basis*. Let  $\psi$  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  $L^2(\mathbb{R})$ .  $\diamond$

**Exercise 9.13.** 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<sup>3</sup> modulo  $2^j$  with the interval  $[0, 2^j)$ . The collection of double-paned windows  $F_j$ , provide a partition of  $\mathbb{R} \setminus \{0\}$ . On each double-paned window, the trigonometric functions

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

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

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

The Shannon wavelets are perfectly localized in frequency, therefore not in space. The compact support on frequency side translates into smoothness ( $C^\infty$ ) of the Shannon wavelet. Therefore the Shannon wavelet provides an example of a  $C^\infty$  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<sup>4</sup> in a fundamental paper in wavelet theory [Dau88]. However it is impossible to construct a  $C^\infty$  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 10.3.

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.

---

<sup>4</sup>Ingrid Daubechies, 1954–, Belgian mathematician.

### 9.3. Haar analysis

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

For these to make sense we have to first define the dyadic intervals, and describe their geometry. We then define the expectation (or averaging) and difference operators, and reduce the question of completeness of the Haar system to understanding the limiting behaviour of the averaging operators. We prove the required limit results for continuous and compactly supported functions, finally an approximation argument coupled with some uniform bounds give the desired result.

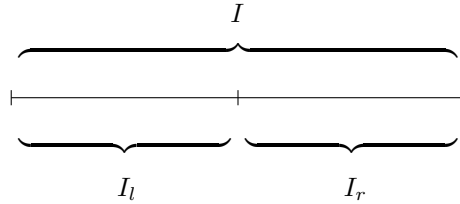
**9.3.1. The dyadic intervals.** Intervals of the form

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

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  $j^{\text{th}}$  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 9.2. Clearly,  $I = I_l \cup I_r$ . We denote by  $|I|$  the length of the interval  $I$ .



**Figure 9.2.** Parent dyadic interval  $I$  and its children  $I_l$  and  $I_r$ .

**Exercise 9.14.** (*Dyadic Intervals Are Nested Or Disjoint*) Show that if  $I, J \in \mathcal{D}$ , then exactly one of the following is true:  $I \cap J = \emptyset$ , or  $I \subseteq J$ , or  $J \subseteq I$ .  $\diamond$

Given  $J \in \mathcal{D}$ , we denote by  $\mathcal{D}(J)$  the collection of dyadic intervals that are contained in  $J$ .

Given  $x \in \mathbb{R}$  and  $j \in \mathbb{Z}$ , there is a unique interval  $I \in \mathcal{D}_j$  such that  $x \in I$ . We denote this unique interval by  $I_j(x)$ .

**9.3.2. The Haar basis.** Associated to each dyadic interval  $I$  there is a *Haar function*  $h_I$  defined by

$$h_I(x) := \frac{1}{|I|^{1/2}} [\chi_{I_r}(x) - \chi_{I_l}(x)].$$

**Figure 9.3.** 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  $h_{[0, 2]}(x) = \frac{1}{\sqrt{2}} [\chi_{[1, 2]}(x) - \chi_{[0, 1]}(x)]$ .

Figure 9.3 shows the graphs of two Haar functions.

**Exercise 9.15.** Let  $I_{j,k} = [k2^{-j}, (k+1)2^{-j}]$ . Show that

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

$\diamond$

In Exercise 9.10, you were asked to check that  $\{h_{j,k}\}_{j,k \in \mathbb{Z}}$  is an orthonormal set in  $L^2(\mathbb{R})$ . Exercise 9.15 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 9.14). 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  $J$ 's 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.

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

Not only the Haar system is orthonormal it is also a complete orthonormal system, hence a basis of  $L^2(\mathbb{R})$ .

**Theorem 9.16.** *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 9.16, 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<sup>5</sup>:

$$(9.4) \quad f = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_I.$$

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.

<sup>5</sup>Recall that equation (9.4) holds in the  $L^2$ -sense if  $\|f - \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_I\|_2 = 0$ .

- 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?<sup>6</sup>

If the Haar system is complete, then for any  $f \in L^2(\mathbb{R})$ , equation (9.4) 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  $\chi_{[0,1]}$ .

---

<sup>6</sup>This question was posed by Lindsay Crowl, one of the participants in the 2004 Program for Women in Mathematics held at the Institute for Advanced Study in Princeton, NJ, where we gave the lectures that lead to this book. We decided it was a perfectly natural concern, and furthermore it was very illustrative of the dangers of interchanging limiting operations!

**Exercise 9.17.** 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 9.17 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  $h_{I^n}(x) = \frac{1}{\sqrt{2^n}} [-\chi_{[0,2^{n-1})}(x) + \chi_{[2^{n-1},2^n)}(x)]$ , we can now evaluate pointwise the series:

$$\begin{aligned} \sum_{I \in \mathcal{D}} \langle \chi_{[0,1]}, h_I \rangle h_I(x) &= \sum_{n=1}^{\infty} \langle \chi_{[0,1]}, h_{I^n} \rangle h_{I^n}(x) \\ (9.5) \qquad \qquad \qquad &= \sum_{n=1}^{\infty} \frac{1}{2^n} [\chi_{[0,2^{n-1})}(x) - \chi_{[2^{n-1},2^n)}(x)]. \end{aligned}$$

We claim that the last term in equation (9.5) is equal to  $\chi_{[0,1)}(x)$  for each  $x \in \mathbb{R}$ . To justify this, let us evaluate the sum term by term for different values of  $x$ .

**For  $x < 0$ :** Clearly  $h_{I^n}(x) = 0$  for all  $n \geq 0$ , and so

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

**For  $0 \leq x < 1$ :** We have already mentioned that  $[0,1)$  sits on  $I_l^n$ , the left half of  $I^n$ , for all  $n \geq 1$ . Hence

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

**For  $1 \leq x < 2$ :** This time  $[1, 2) = I_r^1$ , and  $[1, 2) \subset I_\ell^n$  for all  $n \geq 2$ . Therefore

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

**For  $2 \leq x < 2^2$ :** Here  $[2, 2^2) \cap I^1 = \emptyset$ ,  $[2, 2^2) = I_r^2$ , and  $[2, 2^2) \subset I_\ell^n$  for all  $n \geq 3$ . Therefore

$$\sum_{n=1}^{\infty} \frac{1}{2^n} [\chi_{[0, 2^{n-1})}(x) - \chi_{[2^{n-1}, 2^n)}(x)] = -\frac{1}{2^2} + \sum_{n=3}^{\infty} \frac{1}{2^n} = 0.$$

The pattern is now clear.

**For  $2^k \leq x < 2^{k+1}$ :** We have  $[2^k, 2^{k+1}) \cap I^n = \emptyset$  for  $n < k$ ,  $[2^k, 2^{k+1}) = I_r^k$ , and  $[2^k, 2^{k+1}) \subset I_\ell^n$  for all  $n \geq k+1$ . Therefore

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

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

$$f_N(x) := \sum_{n=1}^N \frac{1}{2^n} [\chi_{[0, 2^{n-1})}(x) - \chi_{[2^{n-1}, 2^n)}(x)],$$

to  $\chi_{[0, 1)}(x)$ .

**Exercise 9.18.** 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 9.19.** Verify that  $f_N \rightarrow \chi_{[0, 1)}$  uniformly on  $\mathbb{R}$ . ◇

It is clear that we cannot interchange limit and integral despite having the uniform convergence of  $f_N$ ,

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

But wait, doesn't uniform convergence guarantee the interchange of the limit and the integral? NO, only when integrating on a compact set, and the line is not compact.

**Exercise 9.20.** (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

$$\|f_N - \chi_{[0,1]}\|_{L^2(\mathbb{R})} \rightarrow 0.$$

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  $\mathbb{R}$  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.  $\lim_{n \rightarrow \infty} \int f_n = \int f$ . 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 9.21.** Verify that if  $g : \mathbb{R} \rightarrow [0, \infty)$  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*<sup>7</sup> that will help us to understand the zooming properties of the Haar basis.

<sup>7</sup>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 234.

The *expectation* operators  $P_j : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ ,  $j \in \mathbb{Z}$  are nothing more than averages over dyadic intervals at generation  $j$ :

$$(9.6) \quad P_j f(x) := \frac{1}{|I|} \int_I f(t) dt,$$

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 9.4 shows the graph of a specific function  $f$  together with the graph of  $P_j f$ . In this example  $j = -1$ .

**Exercise 9.22.** 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 9.4.** Graphs of  $f$  and  $P_{-1}f$ .

**Figure 9.5.** Graphs of  $f$  and  $Q_{-1}f$ .

**Figure 9.6.** Graphs of  $f$ ,  $P_{-1}f$ , and  $P_0f$ .

As  $j \rightarrow \infty$ , 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*  $Q_j : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ ,  $j \in \mathbb{Z}$  are given by

$$(9.7) \quad Q_j f(x) := P_{j+1}f(x) - P_j f(x).$$

These operators  $Q_j$  encode the information necessary to go from the approximation  $P_j f$  at resolution  $j$  of  $f$  to the better approximation  $P_{j+1}f$  at resolution  $j + 1$ . Figure 9.5 shows the graph of the same function  $f$  as in Figure 9.4, together with the graph of  $Q_j f$ , again

for  $j = -1$ . Figure 9.6 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 9.23.** 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

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

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

$$(9.8) \quad \langle f, h_I \rangle h_I(x) = \begin{cases} \frac{1}{2}(m_{I_r} f - m_{I_l} f), & \text{if } x \in I_r; \\ -\frac{1}{2}(m_{I_r} f - m_{I_l} f), & \text{if } x \in I_l. \end{cases}$$

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:

$$(9.9) \quad m_I f = \frac{m_{I_l} f + m_{I_r} f}{2}.$$

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 (9.8) we conclude that

$$Q_j f(x) = \langle f, h_I \rangle h_I(x) \quad \text{for } x \in I \in \mathcal{D}_j.$$

This proves the lemma.  $\square$

**Exercise 9.24.** Verify the averaging property in equation (9.9).  $\diamond$

**9.3.4. Completeness of the Haar system.** The Haar system of functions is complete if, for all  $f \in L^2(\mathbb{R})$ , we have

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

By Lemma 9.23, this condition is equivalent to the condition

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

A telescoping series argument shows that

$$(9.10) \quad P_N f(x) - P_M f(x) = \sum_{M \leq j < N} (P_{j+1} f(x) - P_j f(x)) = \sum_{M \leq j < N} Q_j f(x).$$

Therefore, verifying completeness of the Haar system reduces to checking that

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

where all the above equalities hold in the  $L^2$ -sense.

We need only prove the following theorem.

**Theorem 9.25.** For  $f \in L^2(\mathbb{R})$ ,

$$(9.11) \quad \lim_{M \rightarrow -\infty} \|P_M f\|_2 = 0$$

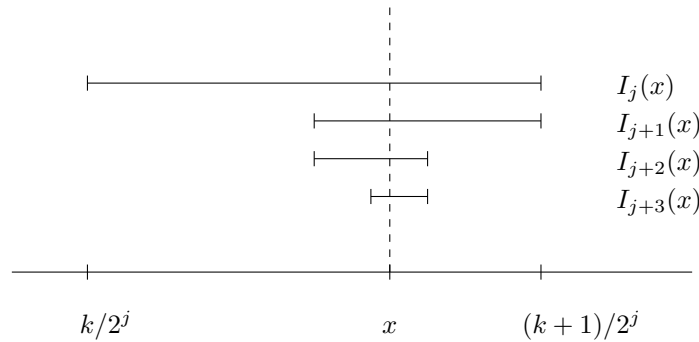
and

$$(9.12) \quad \lim_{N \rightarrow \infty} \|P_N f - f\|_2 = 0.$$

**Exercise 9.26.** Use Theorem 9.25 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.)  $\diamond$

Recall that  $I_j(x)$  is the unique dyadic interval in  $\mathcal{D}_j$  that contains  $x$ . Figure 9.7 shows part of the tower of dyadic intervals  $\{I_j(x)\}_{j \in \mathbb{Z}}$  shrinking to  $\{x\}$  as  $j$  goes to infinity. Equation (9.12) 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:

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



**Figure 9.7.** Part of the tower of nested dyadic intervals  $\dots \supset I_j(x) \supset I_{j+1}(x) \supset I_{j+2}(x) \supset I_{j+3}(x) \supset \dots$  containing the point  $x$ , with  $k/2^j \leq x \leq (k+1)/2^j$ . Here  $j$  and  $k$  are integers.

It turns out that the convergence also holds *almost everywhere*<sup>8</sup> (a.e.). This is the content of the celebrated *Lebesgue Differentiation Theorem* in  $\mathbb{R}$ .

**Theorem 9.27** (Lebesgue Dominated Convergence Theorem on  $\mathbb{R}$ ). *If  $f \in L^1_{loc}(\mathbb{R})$  (locally integrable), then*

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

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  $h \rightarrow 0$ . For a proof see [SS2, Chapter 3].

**Exercise 9.28.** Verify the *Lebesgue Differentiation Theorem* for continuous functions, and show that for continuous functions the pointwise convergence holds everywhere. That is, for all  $x \in \mathbb{R}$ ,

$$(9.13) \quad \lim_{[a,b] \rightarrow \{x\}} \frac{1}{b-a} \int_a^b f(t) dt = f(x).$$

Here the intervals  $[a, b]$  need not be dyadic. ◇

Theorem 9.25 is a consequence of the following lemmata.

**Lemma 9.29.** *For every function  $f \in L^2(\mathbb{R})$  and every integer  $j$ , the operators  $P_j$  are uniformly bounded in  $L^2(\mathbb{R})$ . More precisely,*

$$\|P_j f\|_2 \leq \|f\|_2.$$

**Lemma 9.30.** *If  $g$  is continuous and has compact support on the interval  $[-K, K]$ , then Theorem 9.25 holds.*

**Lemma 9.31.** *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.

---

<sup>8</sup>In other words, the convergence holds except on a set of measure zero.

**Proof of Theorem 9.25.** By Lemma 9.31, 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 9.30, we can choose  $N$  large enough so that for all  $j > N$ ,

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

Now apply the triangle inequality, and use Lemma 9.29 to conclude that

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

This proves equation (9.11).

Similarly, by Lemma 9.30, we can choose  $N$  large enough that for all  $j > N$ ,

$$\|P_jg - g\|_2 \leq \varepsilon/2.$$

Now apply the triangle inequality, and use Lemma 9.29 to conclude that

$$\|P_jf - f\|_2 \leq \|P_jg - g\|_2 + \|P_jh - h\|_2 \leq \frac{\varepsilon}{2} + 2\|h\|_2 \leq \varepsilon.$$

This proves equation (9.12).  $\square$

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 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 9.29. In one case the operators converge to the zero operator, as  $j \rightarrow -\infty$ , and in the other case they converge to the identity operator, as  $j \rightarrow \infty$ .

Here is a precise statement of the Uniform Boundedness Principle, from [Sch, Chapter III].

**Theorem 9.32** (Uniform Boundedness Principle). *Let  $W$  be a family of bounded linear operators  $T : X \rightarrow Y$  from a Banach space  $X$  into*

a normed space  $Y$ , such that for each  $x \in X$ ,

$$\sup_{T \in W} \|Tx\|_Y < \infty.$$

Then the operators are uniformly bounded. That is, there exists a constant  $C > 0$  such that for all  $T \in W$ , and all  $x \in X$ ,

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

In particular if the operators  $\{T_n\}_{n \geq 1}$  are uniformly bounded, that is  $\|T_n x\|_Y \leq C\|x\|_X$ , and  $\lim_{n \rightarrow \infty} \|T_n x\|_Y = 0$ , for all  $x$  on a dense subset  $A \subset X$ , then  $\lim_{n \rightarrow \infty} \|T_n x\|_Y = 0$  for all  $x \in X$ . Why? Consider a point  $x \in X$ . There are points  $x_m \in A$  such that  $\lim_{m \rightarrow \infty} \|x_m - x\|_X = 0$ , and for those points we know that  $\|T_n x_m\|_Y \rightarrow 0$  as  $n \rightarrow \infty$  for all  $m$ . Now,  $x = (x - x_m) + x_m$ , the operators are linear, hence  $T_n x = T_n(x - x_m) + T_n(x_m)$ . We can estimate the norm of  $T_n x$  using the triangle inequality and the uniform boundedness:

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

Both terms on the right hand side can be made arbitrarily small. First, by choosing  $M$  large enough we can ensure that for a given  $\varepsilon$ ,

$$\|x - x_m\|_Y < \varepsilon/2C, \quad \text{for all } m > M.$$

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

$$\lim_{n \rightarrow \infty} \|T_n x\|_Y = 0.$$

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 9.33.** Let  $X$  be the Banach space of all real-valued continuous functions of period  $2\pi$  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  $N^{\text{th}}$  partial Fourier sum of  $f$ . Recall that  $S_N f = D_N * f$

where  $D_N$  denotes the periodic Dirichlet kernel (see Chapter 4). Verify that if  $C_N > 0$  is such that

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

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

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

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

**Proof of Lemma 9.29.** For  $x \in I \in \mathcal{D}_j$ ,

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

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

Now integrate over the interval  $I$  to obtain

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

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

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

The lemma is proved.  $\square$

**Proof of Lemma 9.30.** 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

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

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 (9.11) 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

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

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

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

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 (9.12) for continuous functions with compact support.  $\square$

**Proof of Lemma 9.31.** 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  $\delta$  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,

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

We have shown (Lemma 9.30) that the step functions can approximate continuous functions with compact support in the  $L^2$ -norm. Lemma 9.31 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 9.34.** (*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 9.35.** 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? ◇

## Zooming properties of wavelets, and applications

We highlight the *zooming* properties of the Haar system, and how they can be mathematically encoded in the so-called *multiresolution analysis* (MRA). The MRA provides a framework for the construction of most wavelets, this is the celebrated Mallat's Theorem. We discuss how to implement the wavelet transform via filter banks. We present, in a very informal manner, competing attributes we would like the wavelets to have, and a by no means exhaustive catalog of wavelets. We briefly discuss wavelet packets and two-dimensional wavelets used in image processing, as well as the use of wavelet decompositions in compression and denoising of images and signals.

### 10.1. Multiresolution analyses (MRAs)

An *orthogonal multiresolution analysis* is a collection of closed subspaces  $\{V_j\}_{j \in \mathbb{Z}}$  of  $L^2(\mathbb{R})$  such that

- (1)  $\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset L^2(\mathbb{R})$  (nested)
- (2)  $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$  (trivial intersection),

- (3)  $\bigcup_{j \in \mathbb{Z}} V_j$  is dense in  $L^2(\mathbb{R})$  (density in  $L^2(\mathbb{R})$ ),
- (4)  $f(x) \in V_j$  if and only if  $f(2x) \in V_{j+1}$  (scaling property),
- (5)  $f(x) \in V_0$  if and only if  $f(x - k) \in V_0$  for any  $k \in \mathbb{Z}$  (shift invariance),
- (6) There exists a *scaling function*  $\varphi \in V_0$  such that its integer translates,  $\{\varphi(x - k)\}_{k \in \mathbb{Z}}$ , form an orthonormal basis for  $V_0$ .

Note that the scaling function  $\varphi$  completely determines the nested subspaces:  $V_0$  by property (6), and the scaling property (4) allows us to “move up” and “down” the scale of subspaces. The goal is to identify suitable scaling functions that will generate an orthogonal MRA, not every  $\varphi$  will work. But once one such function  $\varphi$  has been identified, then appropriate translates and dilates of it will form an orthonormal basis for  $V_j$ . More precisely, given the scaling function  $\varphi$ , we denote its integer translates and dyadic dilates with subscripts  $j, k$ , as we did for the wavelet  $\psi$ :

$$(10.1) \quad \varphi_{j,k} := 2^{j/2} \varphi(2^j x - k).$$

**Exercise 10.1.** Show that  $\{\varphi_{j,k}\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $V_j$ .  
 $\diamond$

Given an  $L^2$ -function  $f$ , let  $P_j f$  be the *orthogonal projection of  $f$  onto  $V_j$* :

$$(10.2) \quad P_j f := \sum_{k \in \mathbb{Z}} \langle f, \varphi_{j,k} \rangle \varphi_{j,k}.$$

The function  $P_j f$  is an approximation to the original function at scale  $2^{-j}$ . More precisely, it is the *best approximation* in the subspace  $V_j$  to  $f$ . See Theorem 5.36, and Section 9.3.3.

The approximation subspaces are nested, and so  $P_{j+1} f$  is a better approximation to  $f$  than  $P_j f$  is, or at least an equally good approximation. How do we go from the approximation  $P_j f$  to the better approximation  $P_{j+1} f$ ? Define the *difference operator*  $Q_j$  by

$$Q_j f := P_{j+1} f - P_j f.$$

To recover  $P_{j+1} f$  we add  $Q_j f$  to  $P_j f$ . It is clear that  $P_{j+1} = P_j + Q_j$ , and that the orthogonal projection  $P_j(P_{j+1} f)$  onto  $V_j$  of  $P_{j+1} f$  is

orthogonal to their difference  $P_{j+1}f - P_j(P_{j+1}f)$ . After observing that  $P_j(P_{j+1}f) = P_jf$  for all  $f \in L^2(\mathbb{R})$  we conclude that  $Q_jf$  is orthogonal to  $P_jf$ . This defines  $Q_j$  as the orthogonal projection onto a closed subspace of  $L^2(\mathbb{R})$ , denoted  $W_j$ , which we call the *detail subspace* at scale  $2^{-j}$ . The space  $W_j$  is the *orthogonal complement* of  $V_j$  in  $V_{j+1}$ . This means that  $V_j \perp W_j$ , and if  $f \in V_{j+1}$ , there exist unique  $g \in V_j$  and  $h \in W_j$  such that  $f = g + h$ . In fact  $g = P_jf$  and  $h = Q_jf$ . We use the notation already introduced at the end of Chapter 5 to denote the direct sum of two orthogonal subspaces:

$$V_{j+1} = V_j \oplus W_j.$$

**Exercise 10.2.** Show that  $P_j(P_{j+1}f) = P_jf$  for all  $f \in L^2(\mathbb{R})$ . Moreover,  $P_j(P_n f) = P_jf$  for all  $n \geq j$ .  $\diamond$

**Exercise 10.3.** Show that for all  $n < j$

$$V_j = V_n \oplus W_n \oplus W_{n-1} \oplus \cdots \oplus W_{j-2} \oplus W_{j-1}.$$

Show also that if  $j \neq k$  then  $W_j \perp W_k$ . Hence we get an orthogonal decomposition of each subspace  $V_j$  in terms of the less accurate approximation space  $V_n$  and the detail subspaces  $W_k$  at intermediate resolutions  $n \leq k < j$ .  $\diamond$

The nested subspaces  $\{V_j\}$  define an orthogonal MRA, the density condition (3) holds, and therefore the detail subspaces give an orthogonal decomposition of  $L^2(\mathbb{R})$ :

$$(10.3) \quad L^2(\mathbb{R}) = \overline{\bigoplus_{j \in \mathbb{Z}} W_j}.$$

We show in Section ?? that the scaling function  $\varphi$  determines a *wavelet*  $\psi$  such that  $\{\psi(x - k)\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $W_0$ . The detail subspace  $W_j$  is a dilation of  $W_0$ , therefore the function

$$\psi_{j,k} = 2^{j/2} \psi(2^j x - k),$$

is in  $W_j$ , and the family  $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$  forms an orthonormal basis for  $W_j$ . The orthogonal projection  $Q_j$  onto  $W_j$  is given by

$$Q_j f = \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}.$$

**Exercise 10.4.** Show that the detail subspace  $W_j$  is a dilation of  $W_0$ , that is it obeys the same scale invariance property (4) that the approximation subspaces  $V_j$  satisfy.  $\diamond$

The collection of functions  $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$  forms a wavelet basis for  $L^2(\mathbb{R})$ . This result is Mallat's Theorem, which we state here and prove in Section ??.

**Theorem 10.5** (Mallat). *Given an MRA with scaling function  $\varphi$ , there is a wavelet  $\psi \in L^2(\mathbb{R})$  such that for each  $j$ , the family  $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $W_j$ . Hence the family  $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$  is an orthonormal basis for  $L^2(\mathbb{R})$ .*

**Exercise 10.6.** Given an orthogonal MRA with scaling function  $\varphi$ , show that if  $\{\psi_{0,k}\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $W_0$ , then  $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$  is an orthonormal basis for  $W_j$ . Show that the two-parameter family  $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$  forms an orthonormal basis for  $L^2(\mathbb{R})$ .  $\diamond$

**Example 10.7.** (*The Haar MRA*) The characteristic function  $\varphi(t) = \chi_{[0,1]}(t)$  is the scaling function of an orthogonal MRA, namely the *Haar MRA*. The subspace  $V_j$  corresponds to step functions with steps on the intervals  $[k2^{-j}, (k+1)2^{-j})$ , and they have all the properties listed. The Haar function is the wavelet  $\psi$  in Mallat's theorem. We return to this example in Section 10.2.

In Section 9.3.3 we defined the expectation and difference operator for the Haar basis. These operators coincide with the ones associated with the *Haar MRA*.  $\diamond$

Are there other MRAs? Yes, there are.

**Example 10.8.** (*The Shannon MRA*) We met the Shannon wavelet in Section 9.2. The scaling function is defined on the Fourier side by

$$\widehat{\varphi}(\xi) = \chi_{[-1/2, 1/2)}(\xi),$$

The subspaces  $V_j$  consist of those functions that are band limited to the interval  $[-2^{j-1}, 2^{j-1})$ ; that is, those functions  $f$  such that the support of their Fourier transform is contained on the interval  $[-2^{j-1}, 2^{j-1})$ . The subspaces  $W_j$  consist of functions that are band limited to the double-paned window  $[-2^j, -2^{j-1}) \cup [2^{j-1}, 2^j)$ .  $\diamond$

**Exercise 10.9.** Verify that the subspaces  $V_j$  defined in Example 10.8 do generate an MRA. Verify that the subspace  $W_j$  is the orthogonal complement of  $V_j$  in  $V_{j+1}$ .  $\diamond$

While there are wavelets that do not come from an MRA, these are rare. If the wavelet has compact support then it does come from an MRA. For most applications compactly supported wavelets are desirable and sufficient. Finally, the conditions in the definition of the MRA are not independent. For full accounts of all these issues and more, consult the book by Hernández and Weiss [HW, Chapter 2] and the book by Wojtaszczyk [Woj, Chapter 2].

## 10.2. The Haar multiresolution analysis

Before considering how to construct wavelets from scaling functions associated to an orthogonal MRA, we revisit the Haar multiresolution analysis.

The scaling function for the Haar MRA is the characteristic function of the unit interval,

$$\varphi(x) = \begin{cases} 1, & \text{for } 0 \leq x < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

The subspace  $V_0$  is the closure in  $L^2(\mathbb{R})$  of the linear span of the integer translates of the Haar scaling function  $\varphi$ ,

$$V_0 := \overline{\text{span}(\{\varphi(\cdot - k)\}_{k \in \mathbb{Z}})}.$$

It consists of piecewise constant functions with jumps only at the integers, and such that the sequence of coefficients lies in  $\ell_2(\mathbb{Z})$ , and therefore the function  $f$  lies in  $L^2(\mathbb{R})$ . More precisely,

$$V_0 = \left\{ f = \sum_{k \in \mathbb{Z}} a_k \varphi_{0,k} : \sum_{k \in \mathbb{Z}} |a_k|^2 < \infty \right\}.$$

Similarly, the subspace

$$V_j := \overline{\text{span}(\{\varphi_{j,k}\}_{k \in \mathbb{Z}})}$$

is the subspace consisting of piecewise constant functions in  $L^2(\mathbb{R})$  with jumps only at the integer multiples of  $2^{-j}$ , that is,

$$V_j = \left\{ f = \sum_{k \in \mathbb{Z}} a_k \varphi_{j,k} : \sum_{k \in \mathbb{Z}} |a_k|^2 < \infty \right\}.$$

The orthogonal projection  $P_j f$  onto  $V_j$  is the piecewise constant function with jumps at the integer multiples of  $2^{-j}$ , whose value on the interval  $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$  is given by the integral average of  $f$  over the interval  $I_{j,k}$ . To go from  $P_j f$  to  $P_{j+1} f$ , we add the difference  $Q_j f$  (the expectation operators  $P_j$  and the difference operators  $Q_j$  were defined in Section 9.3.3.) We showed in Lemma 9.23 that  $Q_j f$  coincides with

$$Q_j f = \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k},$$

where  $\psi$  is the Haar wavelet, defined by

$$\psi(x) = h(x) = \begin{cases} -1, & \text{for } 0 \leq x < 1/2; \\ 1, & \text{for } 1/2 \leq x < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Hence  $Q_j$  is the orthogonal projection onto  $W_j$ , the closure of the linear span of  $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ . The subspace

$$W_j = \overline{\text{span}(\{\psi_{j,k}\}_{k \in \mathbb{Z}})},$$

consists of piecewise constant functions in  $L^2(\mathbb{R})$  with jumps only at integer multiples of  $2^{-(j+1)}$ , and average 0 between integer multiples of  $2^{-j}$ .

We can view the averages  $P_j f$  at resolution  $j$  as successive approximations to the original signal  $f \in L^2(\mathbb{R})$ . These approximations are the orthogonal projections  $P_j f$  onto the approximation spaces  $V_j$ . The *details*, necessary to move from level  $j$  to the next level  $(j+1)$ , are encoded in the Haar coefficients at level  $j$ , more precisely in the orthogonal projections  $Q_j f$  onto the detail subspaces  $W_j$ . Starting at a low resolution level, we can obtain better and better resolution by adding the details at the subsequent levels. As  $j \rightarrow \infty$ , the resolution is increased. The steps get smaller (length  $2^{-j}$ ), and the approximation converges to  $f$  in  $L^2$ -norm (this is the content of equation (9.12) in Theorem 9.25), and a.e. (Lebesgue Differentiation

Theorem). Clearly the subspaces are nested, that is,  $V_j \subset V_{j+1}$ , and their intersection is the trivial subspace containing just the zero function (this is equation (9.11) in Theorem 9.25). Lo and behold, we have shown that the Haar scaling function generates an orthogonal MRA.

We give an example of how to decompose a function into its projections onto the Haar subspaces. We have borrowed this example from [MP].

In practice, we select a coarsest scale  $V_{-n}$  and a finest scale  $V_0$ , truncate the chain to

$$V_{-n} \subset \cdots \subset V_{-2} \subset V_{-1} \subset V_0,$$

and obtain

$$(10.4) \quad V_0 = V_{-n} \oplus W_{-n} \oplus W_{-n+1} \oplus W_{-n+2} \oplus \cdots \oplus W_{-1}.$$

In the example we choose  $n=3$ , so the coarser scale correspond to intervals of length  $2^3 = 8$ , and the finest to intervals of length one. We will go through the decomposition process in the text using vectors. It is also enlightening to look at the graphical version in Figure 10.1.

We begin with a vector of  $8 = 2^3$  “samples” of a function, which we assume to be the average value of the function on 8 intervals of length 1, so that our function is supported on the interval  $[0, 8]$ . For our example, we choose the vector

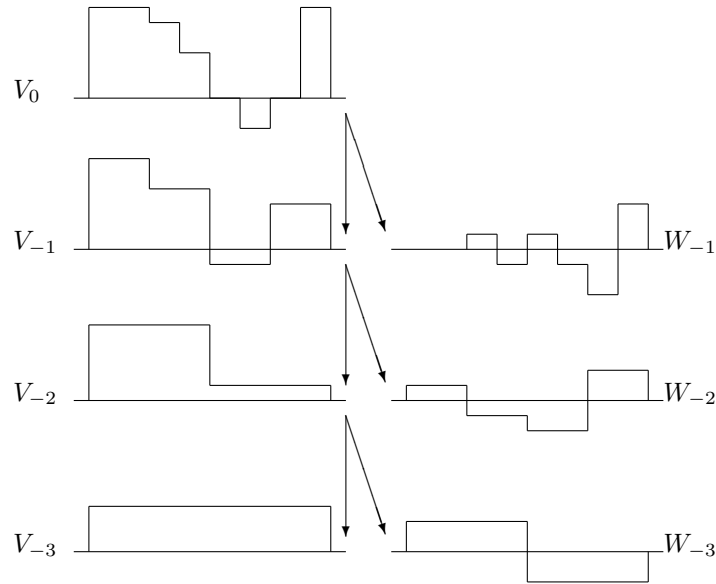
$$v_0 = [6, 6, 5, 3, 0, -2, 0, 6]$$

to represent our function in  $V_0$ . The convention we are using throughout the example is that the vector  $v = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7] \in \mathbb{C}^8$  represents the step function  $f(x) = a_j$  for  $j \leq x < j + 1$ ,  $j = 0, 1, \dots, 7$ ,  $f(x) = 0$  otherwise.

**Exercise 10.10.** Use the above convention to describe the scaling function in  $V_0, V_{-1}, V_{-2}, V_{-3}$  supported on  $[0, 8)$  and the Haar functions in  $W_{-1}, W_{-2}, W_{-3}$  supported on  $[0, 8)$ .  $\diamond$

To construct the projection onto  $V_{-1}$  we average pairs of values, obtaining

$$v_{-1} = [6, 6, 4, 4, -1, -1, 3, 3].$$



**Figure 10.1.** A wavelet decomposition:  $V_0 = V_{-3} \oplus W_{-3} \oplus W_{-2} \oplus W_{-1}$ .

The difference  $v_0 - v_{-1} = w_{-1}$  is in  $W_{-1}$ , so we have

$$w_{-1} = [0, 0, 1, -1, 1, -1, -3, 3].$$

By repeating this process, we obtain the projections of  $v_0$  onto  $V_{-2}$ ,  $W_{-2}$ ,  $V_{-3}$  and  $W_{-3}$ ,

$$\begin{aligned} v_{-2} &= [5, 5, 5, 5, 1, 1, 1, 1], \\ w_{-2} &= [1, 1, -1, -1, -2, -2, 2, 2], \\ v_{-3} &= [3, 3, 3, 3, 3, 3, 3, 3], \quad \text{and} \\ w_{-3} &= [2, 2, 2, 2, -2, -2, -2, -2]. \end{aligned}$$

For the example in Figure 10.1, the scaling function subspaces are shown in Figure 10.2 and the wavelet subspaces are shown in Figure 10.3.

To compute the coefficients of the expansion (10.2), we must compute the inner product  $\langle f, \varphi_{j,k} \rangle$  for the function (10.1). In terms of

$$\begin{aligned}
 V_{-1} &= \text{span} \left( \begin{array}{c} \left( \begin{array}{cc} \text{[Step function 1]} & \text{[Step function 2]} \end{array} \right) \\ \left( \begin{array}{cc} \text{[Step function 3]} & \text{[Step function 4]} \end{array} \right) \end{array} \right) \\
 V_{-2} &= \text{span} \left( \begin{array}{c} \left( \begin{array}{cc} \text{[Step function 5]} & \text{[Step function 6]} \end{array} \right) \end{array} \right) \\
 V_{-3} &= \text{span} \left( \begin{array}{c} \left( \begin{array}{c} \text{[Step function 7]} \end{array} \right) \end{array} \right)
 \end{aligned}$$

**Figure 10.2.** The scaling function subspaces used in Figure 10.1.

$$\begin{aligned}
 W_{-1} &= \text{span} \left( \begin{array}{c} \left( \begin{array}{cc} \text{[Wavelet function 1]} & \text{[Wavelet function 2]} \end{array} \right) \\ \left( \begin{array}{cc} \text{[Wavelet function 3]} & \text{[Wavelet function 4]} \end{array} \right) \end{array} \right) \\
 W_{-2} &= \text{span} \left( \begin{array}{c} \left( \begin{array}{cc} \text{[Wavelet function 5]} & \text{[Wavelet function 6]} \end{array} \right) \end{array} \right) \\
 W_{-3} &= \text{span} \left( \begin{array}{c} \left( \begin{array}{c} \text{[Wavelet function 7]} \end{array} \right) \end{array} \right)
 \end{aligned}$$

**Figure 10.3.** The wavelet subspaces used in Figure 10.1.

our vectors, we have for example

$$\langle f, \varphi_{0,3} \rangle = \langle [6, 6, 5, 3, 0, -2, 0, 6], [0, 0, 0, 1, 0, 0, 0, 0] \rangle = 3$$

and

$$\langle f, \varphi_{1,1} \rangle = \langle [6, 6, 5, 3, 0, -2, 0, 6], [0, 0, 1/\sqrt{2}, 1/\sqrt{2}, 0, 0, 0, 0] \rangle = 8/\sqrt{2}.$$

**Exercise 10.11.** Verify that if  $f, g$  are functions in  $V_0$  supported on  $[0, 8)$ , described according to our convention by the vectors  $v, w \in \mathbb{C}^8$ , then  $\langle f, g \rangle_{L^2(\mathbb{R})} = v \cdot w$ , where  $v \cdot w$  denotes the inner product in  $\mathbb{C}^8$ .  
 $\diamond$

The scaling function  $\varphi$  satisfies the *two-scale recurrence equation*

$$(10.5) \quad \varphi(t) = \varphi(2t) + \varphi(2t - 1).$$

Therefore  $\varphi_{j,k} = (\varphi_{j+1,2k} + \varphi_{j+1,2k+1})/\sqrt{2}$ , and so

$$\langle f, \varphi_{j,k} \rangle = \frac{1}{\sqrt{2}}(\langle f, \varphi_{j+1,2k} \rangle + \langle f, \varphi_{j+1,2k+1} \rangle).$$

Thus we can also compute

$$\langle f, \varphi_{1,1} \rangle = \frac{1}{\sqrt{2}}(5 + 3).$$

The coefficients  $\langle f, \varphi_{j,k} \rangle$  for fixed  $j$  are called the *averages* of  $f$  at scale  $j$ , and denoted  $a_{j,k}$ .

Similarly, the wavelet satisfies the *two-scale difference equation*,

$$(10.6) \quad \psi(t) = \varphi(2t) - \varphi(2t - 1),$$

and thus we can recursively compute

$$\langle f, \psi_{j,k} \rangle = \frac{1}{\sqrt{2}}(\langle f, \varphi_{j+1,2k} \rangle - \langle f, \varphi_{j+1,2k+1} \rangle).$$

**Exercise 10.12.** Verify that the two-scale recurrence equation (10.5), and the two-scale difference equation (10.6) hold for the Haar scaling and wavelet.  
 $\diamond$

The coefficients  $\langle f, \psi_{j,k} \rangle$  for fixed  $j$  are called the *differences* or *details* of  $f$  at scale  $j$ , and denoted  $d_{j,k}$ . Evaluating the whole set of Haar coefficients  $d_{j,k}$  and averages  $a_{j,k}$  requires  $2(N-1)$  additions and  $2N$  multiplications. The discrete wavelet transform can be performed using a similar *cascade algorithm* with complexity  $N$ , where  $N$  is the

number of data points. Let us remark that an arbitrary change of basis in  $N$ -dimensional space requires multiplication by an  $N \times N$  matrix, hence *a priori* one requires  $N^2$  multiplications. This is the same algorithm as the fast Haar transform we discussed in Chapter 6 in the language of matrices.

### 10.3. Mallat's algorithm revisited

In this section we give a glimpse of how we would implement the wavelet transform once an MRA is at our disposal, in a similar way to the implementation for the Haar functions.

It turns out that all one needs for computations are the so-called *filter coefficients*, and not the scaling and wavelet functions. These filter coefficients are finite sequences of numbers if and only if the scaling function is *compactly supported*, as it is the case of the Haar scaling function. In this case, the corresponding wavelet will also be compactly supported.

Given an orthogonal MRA, the scaling function  $\varphi$  satisfies the *scaling equation*, for some set of filter coefficients  $h = \{h_k\}$  such that  $\sum_k |h_k|^2 < \infty$ :

$$(10.7) \quad \varphi(t) = \sum_{k \in \mathbb{Z}} h_k \varphi_{1,k}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \varphi(2t - k).$$

In the case of the Haar MRA, we have  $h_0 = h_1 = 1/\sqrt{2}$ , and all other coefficients vanish. The sequence  $h = \{h_k\}$  is the so-called *low-pass filter*. We assume that the low-pass filter has finite length  $L$ . (It turns out that such a filter always has even length,  $L = 2M$  say. The *refinement mask* is given by  $H(\xi) = (1/\sqrt{2}) \sum h_k e^{-2\pi i k \xi}$ , which can be viewed as a 1-periodic function, of the frequency variable  $\xi$ , whose Fourier coefficients are  $\widehat{H}(n) = h_{-n}/\sqrt{2}$ .)

Not all wavelets have compact support, we already encountered one such example, the Shannon wavelet. However, for applications it is a most desirable property, since compactly supported wavelets correspond to FIR filters.

The existence of a solution for the scaling equation can be expressed in the language of fixed-point theory. Given a low-pass filter  $H$ , define a transformation  $T$  by  $T\varphi(t) := \sqrt{2} \sum_k h_k \varphi(2t - k)$ . Does

$T$  have a fixed point? If yes, then the fixed point is a solution to the scaling equation. However, we do not pursue this argument here; instead we use Fourier analysis.

Observe that on Fourier side the scaling equation (10.7) becomes

$$\widehat{\varphi}(\xi) = H(\xi/2)\widehat{\varphi}(\xi/2).$$

We can iterate this formula to obtain

$$(10.8) \quad \widehat{\varphi}(\xi) = \left(\prod_{j=0}^N H(\xi/2^j)\right) \widehat{\varphi}(\xi/2^N).$$

If  $\widehat{\varphi}$  is continuous at  $\xi = 0$ ,  $\widehat{\varphi}(0) \neq 0$ , and the infinite product converges, then there is a solution to the scaling equation, namely on Fourier side,

$$(10.9) \quad \widehat{\varphi}(\xi) = \left(\prod_{j=0}^{\infty} H(\xi/2^j)\right) \widehat{\varphi}(0).$$

To obtain orthonormality of the set  $\{\varphi_{0,k}\}_{k \in \mathbb{Z}}$  we must have  $|\widehat{\varphi}(0)| = 1$ , and one usually normalizes to  $\widehat{\varphi}(0) = \int \varphi(t)dt = 1$ . This normalization happens to be useful in numerical implementations of the wavelet transform.

The conditions on the filter  $H$  that guarantee the existence of a solution  $\varphi$  to the scaling equation are now well understood (consult [HW] for more details). For example, the fact that the infinite product  $\prod_{j=0}^{\infty} H\left(\frac{\xi}{2^j}\right)$  must converge for each  $\xi$ , in particular for  $\xi = 0$ , forces  $H(0) = 1$ , or equivalently  $\sum_{k=0}^{L-1} h_k = \sqrt{2}$ . In Lemma ?? we showed that the orthonormality of the integer shifts of the scaling function implies that

$$(10.10) \quad |H(\xi)|^2 + |H(\xi + 1/2)|^2 = 1.$$

Equation (10.10) is known in the engineering community as a *quadrature mirror filter* (QMF) condition, necessary for exact reconstruction for a pair of filters. The QMF condition together with  $H(0) = 1$  implies that  $H(1/2) = 0$ , which explains the name *low-pass filter*: low frequencies near  $\xi = 0$  are kept, while high frequencies near  $\xi = 1/2$  are removed (filtered out).

The wavelet  $\psi$  we are seeking is an element of  $W_0 \subset V_1$ . Therefore it is also a superposition of the basis elements  $\{\varphi_{1,k}\}_{k \in \mathbb{Z}}$  of  $V_1$ . So

there are coefficients  $\{g_k\}$  such that  $\sum_{k \in \mathbb{Z}} |g_k|^2 < \infty$ , and

$$(10.11) \quad \psi(t) = \sum_{k \in \mathbb{Z}} g_k \varphi_{1,k}(t).$$

Define the *high-pass filter*  $g = \{g_k\}$  by

$$(10.12) \quad g_k = (-1)^{k-1} \overline{h_{1-k}}, \quad G(\xi) = \frac{1}{\sqrt{2}} \sum_{k=0}^{L-1} g_k e^{2\pi i k \xi}.$$

With this choice of filter, the function  $\psi$  given by equation (10.11) is Mallat's wavelet. It suffices to observe that with this choice,  $\widehat{\psi}(\xi) = G(\xi/2)\widehat{\varphi}(\xi/2)$  with  $G(\xi) := (1/\sqrt{2}) \sum g_k e^{-2\pi i k \xi} = e^{2\pi i \xi} \overline{H(\xi + 1/2)}$ . See Exercise 10.13.

If the high-pass filter  $G$  is itself a QMF, in other words if

$$(10.13) \quad |G(\xi)|^2 + |G(\xi + 1/2)|^2 = 1,$$

then it can be verified that for each scale  $j$ , the wavelets  $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$  form an orthonormal basis for  $W_j$ . Furthermore, the orthogonality between  $W_j$  and  $V_j$  implies that for all  $\xi$ ,

$$(10.14) \quad H(\xi)\overline{G(\xi)} + H(\xi + 1/2)\overline{G(\xi + 1/2)} = 0.$$

This relation together with the knowledge that  $H(0) = 1$ , and  $H(\pm 1/2) = 0$ , implies that  $G(0) = 0$  (equivalently  $\sum g_k = 0$ ) and  $G(\pm 1/2) = 1$ , which explains the name *high-pass filter*: high frequencies near  $\xi = \pm 1/2$  are kept, while low frequencies near  $\xi = 0$  are removed.

Given a low-pass filter  $H$  associated to an MRA we will always construct the associated high-pass filter  $G$  according to formula (10.12).

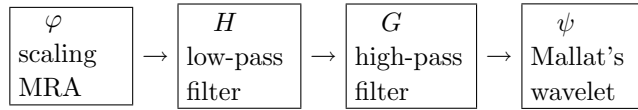
**Exercise 10.13.** Check that with this choice,  $G(\xi) = e^{2\pi i \xi} \overline{H(\xi + 1/2)}$ , and that if  $H$  satisfies a QMF condition, so does  $G$ . Check that equation (10.14) holds. Verify that on the Fourier side

$$\widehat{\psi}(\xi) = G(\xi/2)\widehat{\varphi}(\xi/2),$$

and that for each  $m \in \mathbb{Z}$ ,  $\psi_{0,m}$  is orthogonal to  $\varphi_{0,k}$  for all  $k \in \mathbb{Z}$ . That is,  $V_0 \perp W_0$ . A similar calculation shows that  $V_j \perp W_j$ .  $\diamond$

Mallat's theorem provides a mathematical algorithm for constructing the wavelet from the MRA and the scaling function via

the filter coefficients. After discussing some examples we will see in the next section that Mallat’s theorem also provides a numerical algorithm that can be implemented very successfully, namely the *cascade algorithm* or *fast wavelet transform*. We represent the cascade algorithm schematically as follows.



**Example 10.14.** (*Haar Revisited*) The characteristic function of the unit interval  $\chi_{[0,1]}$  generates an orthogonal MRA, namely the Haar MRA. The non-zero low-pass filter coefficients are  $h_0 = h_1 = 1/\sqrt{2}$ ; hence the non-zero high-pass coefficients are  $g_0 = -1/\sqrt{2}$  and  $g_1 = 1/\sqrt{2}$ . Therefore the Haar wavelet is  $\psi(t) = \varphi(2t - 1) - \varphi(2t)$ . The refinement masks are

$$H(\xi) = \frac{1 + e^{2\pi i \xi}}{2}, \quad G(\xi) = \frac{e^{-2\pi i \xi} - 1}{2}.$$

Compute the infinite product  $\prod_{j=1}^{\infty} H(\xi/2^j)$  directly in this example, and compare it with  $\widehat{\varphi}(\xi)$  (they should coincide).  $\diamond$

**Example 10.15.** (*Shannon Revisited*) This time  $\widehat{\varphi}(\xi) = \chi_{[-1/2, 1/2]}(\xi)$  generates an orthogonal MRA. It follows from Exercise ?? that

$$H(\xi) = \chi_{[-1/2, -1/4] \cup [1/4, 1/2]}(\xi).$$

Hence by Exercise 10.13,

$$G(\xi) = e^{2\pi i \xi} H(\xi + 1/2) = e^{2\pi i \xi} \chi_{[-1/4, 1/4]}(\xi)$$

(recall that we are viewing  $H(\xi)$  and  $G(\xi)$  as periodic functions on the unit interval), and

$$\widehat{\psi}(\xi) = e^{\pi i \xi} \chi_{\{1/2 < |\xi| \leq 1\}}(\xi).$$

$\diamond$

**Example 10.16.** (*Daubechies Wavelets*) For each integer  $N \geq 1$  there is an orthogonal MRA that generates a compactly and minimally supported wavelet, such that the length of the support is  $2N$ , and the filters have  $2N$  taps. They are denoted in MATLAB by  $dbN$ .

The wavelet  $db1$  is the Haar wavelet. The coefficients corresponding to  $db2$  are

$$h_0 = \frac{1 + \sqrt{3}}{4\sqrt{2}}, \quad h_1 = \frac{3 + \sqrt{3}}{4\sqrt{2}}, \quad h_2 = \frac{3 - \sqrt{3}}{4\sqrt{2}}, \quad h_3 = \frac{1 - \sqrt{3}}{4\sqrt{2}}.$$

◇

Figure 10.4 shows the graphs of the Daubechies scaling and wavelet functions  $db2$ ,  $db4$ , and  $db6$ .

**Exercise 10.17.** Check that the  $db2$  filter is a QMF, and that  $h_0 + h_1 + h_2 + h_3 = \sqrt{2}$ . ◇

It turns out that for finite filters  $H$ , the conditions (10.10) and  $\sum_k h_k = \sqrt{2}$  are sufficient to guarantee the existence of a solution  $\varphi$  to the scaling equation. For infinite filters an extra decay assumption is necessary. However, it is not sufficient to guarantee the orthonormality of the integer shifts of  $\varphi$ . But, if for example  $\inf_{|\xi| \leq 1/4} |H(\xi)| > 0$  is also true, then  $\{\varphi_{0,k}\}$  is an orthonormal set in  $L^2(\mathbb{R})$ . For more details, see for example [Fra, Chapter 5].

## 10.4. Cascade algorithm and filter banks

A *cascade algorithm*, similar to the one described for the Haar basis, can be implemented to provide a fast wavelet transform. Given the approximation coefficients  $\{a_{J,k}\}_{k=0,1,\dots,N-1}$ ,  $N = 2^J$  “scaled samples” of the function  $f$  defined on the interval  $[0, 1]$ , namely

$$a_{J,k} := \langle f, \varphi_{J,k} \rangle.$$

Then the coarser approximation and detail coefficients  $a_{j,k} := \langle f, \varphi_{j,k} \rangle$  and  $d_{j,k} := \langle f, \psi_{j,k} \rangle$  for scales  $j < J$  can be calculated in order  $LN$  operations, where  $L$  is the length of the filter, and  $N$  is the number of samples, that is the number of the coefficients in the finest approximation scale. Let  $a_j$  denote the sequence  $\{a_{j,k}\}$  and  $d_j$  the sequence  $\{d_{j,k}\}$ .

In order to see why this is possible, let us consider the simpler case of calculating  $a_0$  and  $d_0$  given  $a_1$ . The scaling equation connects

$\varphi_{0,\ell}$  to  $\{\varphi_{1,m}\}$ :

$$\begin{aligned}
 \varphi_{0,\ell}(x) &= \varphi(x - \ell) = \sum_{k \in \mathbb{Z}} h_k \varphi_{1,k}(x - \ell) \\
 &= \sum_{k \in \mathbb{Z}} h_k \sqrt{2} \varphi(2x - (2\ell + k)) \\
 &= \sum_{m \in \mathbb{Z}} h_{m-2\ell} \varphi_{1,m} \\
 &= \sum_{m \in \mathbb{Z}} \overline{\tilde{h}_{2\ell-m}} \varphi_{1,m},
 \end{aligned}$$

where  $\tilde{h}_k := \overline{h_{-k}}$ .

We can now compute  $a_{0,\ell}$  in terms of  $a_1$ :

$$\begin{aligned}
 a_{0,\ell} &= \langle f, \varphi_{0,\ell} \rangle \\
 &= \sum_{m \in \mathbb{Z}} \tilde{h}_{2\ell-m} \langle f, \varphi_{1,m} \rangle \\
 &= \sum_{m \in \mathbb{Z}} \tilde{h}_{2\ell-m} a_{1,m} \\
 &= \tilde{h} * a_1(2\ell),
 \end{aligned}$$

where  $\tilde{h} = \{\tilde{h}_k\}$  is the conjugate flip of the low-pass filter  $h$ .

Similarly, for the detail coefficients  $d_{0,\ell}$ , the two-scale equation connects  $\psi_{0,\ell}$  to  $\{\varphi_{1,m}\}$ . The only difference is that the filter coefficients are  $\{g_k\}$  instead of  $\{h_k\}$ , so we obtain

$$\psi_{0,\ell}(x) = \sum_{m \in \mathbb{Z}} \tilde{g}_{2\ell-m} \varphi_{1,m},$$

and consequently,

$$d_{0,\ell} = \sum_{m \in \mathbb{Z}} \tilde{g}_{2\ell-m} a_{1,m} = \tilde{g} * a_1(2\ell),$$

where  $\tilde{g} = \{\tilde{g}_k\}$  is the conjugate flip of the high-pass filter  $g$ .

**Exercise 10.18.** Verify that for all  $j \in \mathbb{Z}$ ,

$$\begin{aligned}
 \varphi_{j,k} &= \sum_{m \in \mathbb{Z}} h_{m-2\ell} \varphi_{j+1,m} \\
 \psi_{j,k} &= \sum_{m \in \mathbb{Z}} g_{m-2\ell} \varphi_{j+1,m}
 \end{aligned}$$

◇

As a consequence of Exercise 10.18: we obtain for all  $j \in \mathbb{Z}$

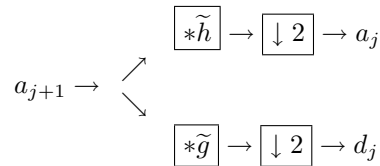
$$\begin{aligned} a_{j,k} &= \sum_{n \in \mathbb{Z}} \tilde{h}_{2k-n} a_{j+1,n} = \tilde{h} * a_{j+1}(2k), \\ d_{j,k} &= \sum_{n \in \mathbb{Z}} \tilde{g}_{2k-n} a_{j+1,n} = \tilde{g} * a_{j+1}(2k), \end{aligned}$$

We can compute the approximation and detail coefficients at a rougher scale ( $j$ ) by convolving (circular convolution) the approximation coefficients at the finer scale ( $j+1$ ) with the low and high-pass filters  $\tilde{h}$  and  $\tilde{g}$ , and *down-sampling* by a factor of two. More precisely the *down-sampling operator* takes an  $N$ -vector and maps it into a vector half as long by discarding the odd entries,

$$Ds(n) = s(2n).$$

The down-sampling operator is denoted by the symbol  $\downarrow 2$ .

In electrical engineering terms, we have just described the *analysis phase* of a *subband filtering scheme*. We can represent the analysis phase schematically as follows.



Another useful operation is *up-sampling*, which is the right inverse of down-sampling. The *up-sampling operator* takes an  $N$ -vector and maps it to a vector twice as long, by intertwining zeros:

$$Us(n) = \begin{cases} s(n/2), & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

The up-sampling operator is denoted by the symbol  $\uparrow 2$ .

**Exercise 10.19.** Compute the Fourier transform for the up-sampling and down-sampling operators in finite-dimensional space. ◇

**Exercise 10.20.** Verify that given a vector  $s \in \mathbb{C}^N$  then  $DU s = s$ , but  $UD s$  is not always  $s$ . For which vectors  $s$  is  $UD s = s$ ? ◇

The reconstruction of the “samples” at level  $j+1$  from the samples and details at the previous level  $j$  is also an order  $N$  algorithm.

We will carefully analyze how to get from the coarse approximation and detail coefficients  $a_j$  and  $d_j$  to the approximation coefficients in the finer scale  $a_{j+1}$ . The calculation is based on the equation  $P_{j+1}f = P_j f + Q_j f$ , which is equivalent to

$$(10.15) \quad \sum_{m \in \mathbb{Z}} a_{j+1,m} \varphi_{j+1,m} = \sum_{\ell \in \mathbb{Z}} a_{j,\ell} \varphi_{j,\ell} + \sum_{\ell \in \mathbb{Z}} d_{j,\ell} \psi_{j,\ell}.$$

We have formulas that express  $\varphi_{j,\ell}$  and  $\psi_{j,\ell}$  in terms of  $\{\varphi_{j+1,m}\}_{m \in \mathbb{Z}}$ ; see Exercise 10.18. Inserting those formulas in the right hand side (RHS) of the equality (10.15), and collecting all terms that are multiples of  $\varphi_{j+1,m}$ , we get

$$\begin{aligned} \text{RHS} &= \sum_{\ell \in \mathbb{Z}} a_{j,\ell} \sum_{m \in \mathbb{Z}} h_{m-2\ell} \varphi_{j+1,m} + \sum_{\ell \in \mathbb{Z}} d_{j,\ell} \sum_{m \in \mathbb{Z}} g_{m-2\ell} \varphi_{j+1,m} \\ &= \sum_{m \in \mathbb{Z}} \left[ \sum_{\ell \in \mathbb{Z}} h_{m-2\ell} a_{j,\ell} + g_{m-2\ell} d_{j,\ell} \right] \varphi_{j+1,m}. \end{aligned}$$

This calculation, together with equation (10.15), implies that

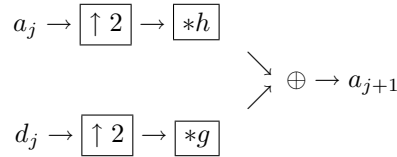
$$a_{j+1,m} = \sum_{\ell \in \mathbb{Z}} h_{m-2\ell} a_{j,\ell} + g_{m-2\ell} d_{j,\ell}.$$

Rewriting in terms of convolutions and up-samplings, we obtain

$$(10.16) \quad a_{j+1,m} = h * Ua_j(m) + g * Ud_j(m),$$

where here we first up-sample the approximation and detail coefficients to restore the right dimensions, then convolve with the filters, and finally add the outcomes.

The synthesis phase of the *subband filtering scheme* can be represented schematically as follows.



**Exercise 10.21.** Verify (10.16). ◇

The process of computing the coefficients can be represented by the following tree or *pyramid scheme*.

$$\begin{array}{ccccccc}
 a_J & \rightarrow & a_{J-1} & \rightarrow & a_{J-2} & \rightarrow & a_{J-3} & \cdots \\
 & & \searrow & & \searrow & & \searrow & \\
 & & d_{J-1} & & d_{J-2} & & d_{J-3} & \cdots
 \end{array}$$

The reconstruction of the signal can be represented with another pyramid scheme, as follows:

$$\begin{array}{ccccccc}
 a_n & \rightarrow & a_{n+1} & \rightarrow & a_{n+2} & \rightarrow & a_{n+3} & \cdots \\
 & & \nearrow & & \nearrow & & \nearrow & \\
 d_n & & d_{n+1} & & d_{n+2} & & \cdots &
 \end{array}$$

Note that once the low-pass filter  $H$  is chosen, everything else—high-pass filter  $G$ , scaling function  $\varphi$ , wavelet  $\psi$ , and MRA—is completely determined. In practice one never computes the values of  $\varphi$  and  $\psi$ . All the manipulations are performed with the filters  $G$  and  $H$ , even if they involve calculating quantities associated to  $\varphi$  or  $\psi$ , like moments or derivatives. However, to help our understanding, we might want to produce pictures of the wavelet and scaling functions from the filter coefficients.

**Example 10.22.** The cascade algorithm can be used to produce very good approximations for both  $\psi$  and  $\varphi$ , and this is how pictures of the wavelets and the scaling functions are obtained. For the scaling function  $\varphi$ , it suffices to observe that  $a_{1,k} = \langle \varphi, \varphi_{1,k} \rangle = h_k$  and  $d_{j,k} = \langle \varphi, \psi_{j,k} \rangle = 0$  for all  $j \geq 1$  (the first because of the scaling equation, the second because  $V_0 \subset V_j \perp W_j$  for all  $j \geq 1$ ), that is what we need to initialize and iterate as many times as we wish (say  $n$  times) the synthesis phase of the filter bank,

$$H \rightarrow \boxed{\uparrow 2} \rightarrow \boxed{*H} \rightarrow \cdots \rightarrow \boxed{\uparrow 2} \rightarrow \boxed{*H} \rightarrow \{\langle \varphi, \varphi_{n+1,k} \rangle\}.$$

The output after  $n$  iterations is the set of approximation coefficients at scale  $j = n + 1$ . After multiplying by a scaling factor, one can make precise the statement that

$$\varphi(k2^{-j}) \sim 2^{-j/2} \langle \varphi, \varphi_{j,k} \rangle.$$



The following substitute for Plancherel holds: (*Riesz basis property*) there exist  $A, B > 0$  such that for all  $f \in L^2(\mathbb{R})$ ,

$$\sum_{j,k} |\langle f, \psi_{j,k}^* \rangle|^2 \sim \|f\|_2^2 \sim \sum_{j,k} |\langle f, \psi_{j,k} \rangle|^2,$$

where  $A \approx B$  means that there exist constants  $c, C > 0$  such that  $cA \leq B \leq CA$ . The relative size of the similarity constants  $c, C$  becomes important for numerical calculations; it is related to the *condition number* of a matrix.

**Exercise 10.24.** Consider two linearly independent vectors in  $\mathbb{R}^2$ ,  $\vec{v}_1 = (x_1, y_1)$ ,  $\vec{v}_2 = (x_2, y_2)$ . They form a basis of  $\mathbb{R}^2$ . Find a pair of so-called *dual vectors*  $\vec{v}_1^*, \vec{v}_2^* \in \mathbb{R}^2$ , such that given any vector  $\vec{u} \in \mathbb{R}^2$ ,

$$\vec{u} = \langle \vec{u}, \vec{v}_1^* \rangle \vec{v}_1 + \langle \vec{u}, \vec{v}_2^* \rangle \vec{v}_2.$$

Show that  $\vec{u} = \langle \vec{u}, \vec{v}_1^* \rangle \vec{v}_1 + \langle \vec{u}, \vec{v}_2^* \rangle \vec{v}_2$ . Find positive numbers  $c$  and  $C$  such that

$$c \sum_{j=1}^2 |\langle \vec{u}, \vec{v}_j^* \rangle|^2 \leq |\vec{u}|^2 \leq C \sum_{j=1}^2 |\langle \vec{u}, \vec{v}_j^* \rangle|^2.$$

Draw a picture explaining the orthogonality properties among these vectors. The numbers  $c, C$  should be related to the angle between the vectors  $\vec{v}_1$  and  $\vec{v}_2$ .  $\diamond$

A general filter bank is a sequence of convolutions and other simple operations such as up- and down-sampling. The study of such banks is an entire subject in engineering called *multirate signal analysis*, or *subband coding*. The term *filter* is used to denote a convolution operator because such operator can cut out various frequencies if the corresponding Fourier multiplier vanishes (or is very small) at those frequencies. Consult Strang and Nguyen's book [SN] for more information.

Filter banks can be implemented quickly because of the Fast Fourier Transform. Remember that circular convolution becomes multiplication on the Fourier side, that is a diagonal matrix (need only  $N$  products, whereas ordinary matrix multiplication requires  $N^2$  products), and to go back and forth from Fourier to time domain, we can use the Fast Fourier Transform, so the total number of operations is of the order  $N \log_2 N$ , see Chapter 6.

There are commercial and free software programs dealing with these applications. The main commercial one is the MATLAB *Wavelet Toolbox*. *Wavelab* is a free software program based on MATLAB. It was developed at Stanford. *Lastwave* is a toolbox with subroutines written in C, with “a friendly shell and display environment” according to Mallat. It was developed at the École Polytechnique. There is much more on the Internet, and you need only search on ‘wavelets’ to see the enormous amount of information, codes, and so on that is available online.

### 10.5. Design features

Most of the applications of wavelets exploit their ability to approximate functions as efficiently as possible, that is with as few coefficients as possible. For different applications one wishes the wavelets to have various properties. Some of them are competing against each other, so it is up to the user to decide which ones are most efficient for her problem. The most popular conditions are orthogonality, compact support, vanishing moments, symmetry and smoothness.

**Orthogonality:** Orthogonality allows for straightforward calculation of the coefficients (via inner products with the basis elements). It guarantees that energy is preserved. However sometimes orthogonality can be substituted by *biorthogonality*. In this case, there is an auxiliary set of dual functions that is used to compute the coefficients by taking inner products; also the energy is almost preserved.

**Compact support:** We have already stressed that compact support is important for numerical purposes, such as implementation of the FIR. In terms of detecting point singularities, it is clear that if the signal  $f$  has a singularity at  $t_0$  then if  $t_0$  is inside the support of  $\psi_{j,n}$ , the corresponding coefficient could be large. If  $\psi$  has support of length  $l$ , then at each scale  $j$  there are  $l$  wavelets interacting with the singularity (that is their support contains  $t_0$ ). The shorter the support the fewer wavelets interacting with the singularity.

We have already mentioned that compact support of the scaling function coincides with FIR. Moreover if the low-pass filter is supported on  $[N_1, N_2]$ , so is  $\varphi$ , and it is not hard to see that  $\psi$  has support of the same length  $(N_2 - N_1)$  but centered at  $1/2$ .

**Smoothness:** The regularity of the wavelet has effect on the error introduced by thresholding or quantizing the wavelet coefficients. Suppose an error  $\varepsilon$  is added to the coefficient  $\langle f, \psi_{j,k} \rangle$ . Then we add an error of the form  $\varepsilon \psi_{j,k}$  to the reconstruction. Smooth errors are often less *visible* or *audible*. Often better quality images are obtained when the wavelets are smooth. However, the smoother the wavelet, the longer the support.

There is no orthogonal wavelet that is  $C^\infty$  and has exponential decay. Therefore there is no hope of finding an orthogonal wavelet that is  $C^\infty$  and has compact support.

**Vanishing moments:** A function  $\psi$  has  $p$  vanishing moments if for all  $k = 0, 1, \dots, p - 1$ ,

$$\int_{\mathbb{R}} x^k \psi(x) dx = 0.$$

If  $\psi$  has  $p$  vanishing moments, then  $\psi$  is orthogonal to polynomials of degree  $p - 1$ . Smooth functions  $f$  have small fine-scale wavelet coefficients. More precisely, if the function to be analyzed is  $k$ -regular, it can be approximated well by a Taylor polynomial of degree  $k$ . If  $k < p$ , then the wavelets are orthogonal to that Taylor polynomial, and the coefficients are small. If  $\psi$  has  $p$  vanishing moments, then the polynomials of degree  $p - 1$  are reproduced by the scaling functions.

The constraints imposed on orthogonal wavelets imply that if  $\psi$  has  $p$  vanishing moments then its support is of length at least  $2p - 1$ . Daubechies wavelets have minimum support length for a given number of vanishing moments. So there is a trade-off between the length of the support and the number of vanishing moments. If the function has few singularities and is smooth between singularities, then we might as well take advantage of the vanishing moments. If there are many singularities, we might prefer to use wavelets with shorter supports.

**Symmetry:** It is not possible to construct compactly supported symmetric orthogonal wavelets, except for the Haar wavelets. However, symmetry is often useful for image and signal analysis. It can be obtained at the expense of one of the other properties. If we give up orthogonality, then there are compactly supported, smooth and symmetric biorthogonal wavelets. If we use *multiwavelets*<sup>1</sup>, we can construct them to be orthogonal, smooth, compactly supported and symmetric. Some wavelets have been designed to be nearly symmetric (Daubechies symmlets, for example).

## 10.6. A catalog of wavelets

We list the main wavelets and indicate their properties.

**Haar wavelet:** Perfectly localized in time, less localized in frequency, discontinuous. Symmetric. Shortest possible support, only one vanishing moment, hence not well adapted to approximating smooth functions.

**Shannon wavelet:** This wavelet does not have compact support, however it is  $C^\infty$ . It is band-limited, but its Fourier transform is discontinuous, hence  $\psi(t)$  decays like  $1/|t|$  at infinity. The Fourier transform of the Shannon wavelet,  $\widehat{\psi}(\xi)$ , is zero in a neighborhood of  $\xi = 0$ , hence all its derivatives are zero at  $\xi = 0$ , and so  $\psi$  has an infinite number of vanishing moments.

**Meyer wavelet:** This is a symmetric band-limited function whose Fourier transform is smooth, hence  $\psi(t)$  has faster decay at infinity. The scaling function is also band-limited. Hence both  $\psi$  and  $\varphi$  are  $C^\infty$ . The wavelet  $\psi$  has an infinite number of vanishing moments. (This wavelet was found by Strömberg in 1983, and it went unnoticed for several years).

---

<sup>1</sup>In this case one has more than one scaling function and more than one wavelet function, whose dilates and translates provide a basis in the MRA.

**Battle-Lemarié spline wavelets:** Polynomial splines of degree  $m$ . The wavelet  $\psi$  has  $m + 1$  vanishing moments. They don't have compact support, but they have exponential decay. Since they are polynomial splines of degree  $m$ , they are  $m - 1$  times continuously differentiable. For  $m$  odd,  $\psi$  is symmetric around  $1/2$ . For  $m$  even, it is antisymmetric around  $1/2$ . The linear spline wavelet is the Franklin wavelet. [\*\*\* Give a reference for Franklin? Ph. Franklin, *A set of continuous orthogonal functions*. Math. Annalen 100 (1928) pp.522–529. ]

**Daubechies compactly supported wavelets:** They have compact support of minimal length for any given number of vanishing moments. More precisely, if  $\psi$  has  $p$  vanishing moments, then the filters have length  $2p$  (or  $2p$  taps). For large  $p$ ,  $\varphi$  and  $\psi$  are uniformly Lipschitz  $\alpha$  of the order  $\alpha \sim 0.2p$ . They are asymmetric. When  $p = 1$  we recover the Haar wavelet.

**Daubechies symmlets:**  $p$  vanishing moments, minimum support of length  $2p$ , as symmetric as possible.

**Coiflets:**  $\psi$  has  $p$  vanishing moments, minimum support, and  $\varphi$  has  $p - 1$  moments vanishing (from the second to the  $p^{\text{th}}$  moment, never the first since  $\int \varphi = 1$ ). This extra property requires enlarging the support of  $\psi$  to length  $(3p - 1)$ . This time if we approximate a regular function  $f$  by a Taylor polynomial, then the approximation coefficients satisfy

$$2^{J/2} \langle f, \varphi_{J,k} \rangle \sim f(2^J k) + O(2^{-(k+1)J}).$$

Hence at fine scale  $J$ , the approximation coefficients are close to the signal samples.

The coiflets were constructed by Daubechies after Coifman requested them for the purpose of applications to almost diagonalization of singular integral operators.

**Mexican hat:** Has a closed formula involving second derivatives of the Gaussian:

$$\psi(t) = C(1 - t^2)e^{t^2/2},$$

where the constant is chosen to normalize it in  $L^2$ . It does not come from an MRA, and it is not orthogonal. It is appropriate for continuous wavelet transform. It has exponential decay but not compact support. According to Daubechies, this function is popular in vision analysis.

**Morlet wavelet:** Given by the closed formula

$$\psi(t) = Ce^{-t^2/2} \cos(5t).$$

It does not come from an MRA, and it is not orthogonal. It is appropriate for continuous wavelet transform. It has exponential decay but not compact support.

**Spline biorthogonal wavelets:** These are compactly supported. There are two positive integer parameters  $N, N^*$ .  $N^*$  determines the scaling function  $\varphi^*$ , it is a spline of order  $[N^*/2]$ . The other scaling function and both wavelets depend on both parameters.  $\psi^*$  is a compactly supported piecewise polynomial of order  $N^* - 1$ , which is  $C^{N^*-2}$  at the knots, and whose support gets larger as  $N$  increases.  $\psi$  also has support increasing with  $N$  and vanishing moments as well. Their regularity can differ notably. The filter coefficients are dyadic rationals, which makes them very attractive for numerical purposes. The functions  $\psi$  are known explicitly. The dual filters are very unequal in length, which could be a nuisance when performing image analysis for example.

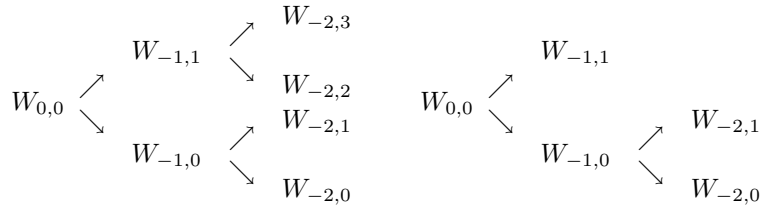
All the previous wavelets are encoded in MATLAB. One can review them and their properties online, and we encourage the reader to do so.

### 10.7. Wavelet packets

To perform the wavelet transform we iterate at the level of the low-pass filter (approximation). In principle it is an arbitrary choice; one could iterate at the high-pass filter level, or any desirable combination. The full dyadic tree gives an overabundance of information, and it corresponds to the *wavelet packets*. Each finite wavelet packet has the information for reconstruction in many different bases, including the wavelet basis. There are fast algorithms to search for the

*best basis*. The *Haar packet* includes the Haar basis and the *Walsh basis*. Wickerhauser’s book is a good source of information on this topic [Wic].

Denote the spaces by  $W_{j,n}$ , where  $j$  is the scale as before, and  $n$  determines the “frequency”.



1. Haar packet binary tree: two levels. Leaves determine Walsh basis.  
 2. Connected binary subtree. Leaves determine Haar basis.

Notice that  $W_{0,0} = V_0$ , more generally,  $W_{j,0} = V_j$ , and  $W_{j,1} = W_j$ . We also know that the spaces  $W_{j-1,2n}$  and  $W_{j-1,2n+1}$  are orthogonal and their direct sum is  $W_{j,n}$ . Therefore the leaves of every connected binary subtree<sup>2</sup> of the wavelet packet tree correspond to an orthogonal basis of the initial space.

Each of the bases encoded in the wavelet packet corresponds to a dyadic tiling of the phase plane in Heisenberg boxes of area one. They provide a much richer time–frequency analysis.

Each of the spaces  $W_{j,n}$  is generated by the integer shifts of a wavelet function at scale  $j$  and frequency  $n$ . More precisely, let  $\omega_{j,k,n}(t) = 2^{-j/2}\omega_n(2^{-j}t - k)$ , where  $n \in \mathbb{N}$ ,  $j, k \in \mathbb{Z}$ , and

$$\begin{aligned} \omega_{2n}(t) &= \sqrt{2} \sum h_k \omega_n(2t - k), & \omega_0 &= \varphi, \\ \omega_{2n+1}(t) &= \sqrt{2} \sum g_k \omega_n(2t - k), & \omega_1 &= \psi, \end{aligned}$$

where  $\{h_k\}$  and  $\{g_k\}$  are the low and high-pass filters of the MRA. Then

$$W_{j,n} = \text{span}\{\omega_{j,k,n} : k \in \mathbb{Z}\} = \left\{ f = \sum_k a_{j,k,n} \omega_{j,k,n} : \sum_k |a_{j,k,n}|^2 < \infty \right\}.$$

<sup>2</sup>Starting from the “top” node ( $W_{0,0}$ ), we allow the node to have either two offspring or none. The same rule applies to the offspring, if there are any. The nodes that have no offspring are the *leaves*.

**Example 10.25.** (*The Walsh Functions*) For the Haar function, the corresponding wavelet packet is described by the equations

$$\begin{aligned}\omega_{2n}(t) &= \omega_n(2t) + \omega_n(2t - 1) \\ \omega_{2n+1}(t) &= \omega_n(2t) - \omega_n(2t - 1).\end{aligned}$$

The functions so obtained are the *Walsh functions*. They are step functions that play the role of the sines and cosines.  $\diamond$

**Exercise 10.26.** Identify all possible bases for the Haar packet binary tree with three levels. How many bases can you find? Sketch the corresponding phase plane diagrams. In particular, sketch the diagram for the Walsh functions.  $\diamond$

**Exercise 10.27.** Use MATLAB to plot the  $n^{\text{th}}$  Walsh function  $\omega_n(t)$  and the functions  $\sin nt$  and  $\cos nt$  for  $-2\pi \leq t \leq 2\pi$  on the same axes, for  $n = 1, 2, \dots, 5$ . Notice that the Walsh functions resemble piecewise-constant versions of the trigonometric functions, with approximately the right frequencies.  $\diamond$

**Exercise 10.28.** What are the Walsh functions in the finite-dimensional case?  $\diamond$

We have seen that a signal of length  $N = 2^J$  can be decomposed in  $2^N$  different ways, the number of binary subtrees of a complete binary tree of depth  $J$ . This is a large number, and one would like to search efficiently in the tree to obtain the best basis with respect to some criteria.

Functionals verifying an additive-type property are well suited for searches of this type. Coifman and Wickerhauser introduced a number of such functionals [CW], among them some *entropy criteria*. Given a signal  $s$  and  $(s_i)_i$  its coefficients in an orthonormal basis. The entropy  $E$  must be an additive cost function such that  $E(0) = 0$  and  $E(s) = \sum_i E(s_i)$ . There are fast algorithms that allow one to search in the wavelet packet tree for the orthonormal basis that minimizes some given entropies. (Four such algorithms are encoded in MATLAB.) Furthermore, the search can be performed in  $O(N \log(N))$  operations. This search procedure is implemented in the Wavelet Toolbox.

The wavelet packet and *cosine packet* libraries create large libraries of orthogonal bases, all of which have fast algorithms. The

Fourier and wavelet bases are particular examples in this time–frequency library; so are Gabor-like bases.

### 10.8. Two-dimensional wavelets

There is a standard procedure to construct bases in 2-D space from given bases in 1-D, the *tensor product*. In particular, given a wavelet basis  $\{\psi_{j,k}\}$  in  $L^2(\mathbb{R})$ , the family of tensor products

$$\psi_{j,k;i,n}(x,y) = \psi_{j,k}(x)\psi_{i,n}(y), \quad j,k,i,n \in \mathbb{Z},$$

is an orthonormal basis in  $L^2(\mathbb{R}^2)$ . Unfortunately we have lost the multiresolution structure. Notice that we are mixing up scales in the above process, that is the scaling parameters  $i, j$  can be anything.

**Exercise 10.29.** Show that if  $\{\psi_n\}_{n \in \mathbb{N}}$  is an orthonormal basis for a closed subspace  $V \subset L^2(\mathbb{R})$ , then the family of functions defined on  $\mathbb{R}^2$  by

$$\psi_{m,n}(x,y) = \psi_m(x)\psi_n(y),$$

is an orthonormal basis for  $V \otimes V$ . Here  $V \otimes V$  is the closure of the linear span of the functions  $\{\psi_{m,n}\}_{m,n \in \mathbb{N}}$ .  $\diamond$

**Exercise 10.30.** What would the trigonometric basis in  $L^2([0,1]^2)$  be? How about the finite-dimensional trigonometric basis, say in two dimensions?  $\diamond$

We would like to use this idea but at the level of the approximation spaces  $V_j$  in the MRA. For each scale  $j$ , the family  $\{\varphi_{j,k}\}_k$  is an orthonormal basis of  $V_j$ . Consider the tensor products  $\varphi_{j,k,n}(x,y) = \varphi_{j,k}(x)\varphi_{j,n}(y)$  of these functions. Let  $\mathcal{V}_j$  be the closure in  $L^2(\mathbb{R}^2)$  of the linear span of those functions, in other words

$$\mathcal{V}_j = V_j \otimes V_j := \left\{ f(x,y) = \sum_{n,k} a_{j,n,k} \varphi_{j,k,n}(x,y) : \sum_{n,k} |a_{j,n,k}|^2 < \infty \right\}.$$

Notice that we are not mixing scales at the level of the MRA. It is not hard to see that the spaces  $\mathcal{V}_j$  form an MRA in  $L^2(\mathbb{R}^2)$ , with scaling function

$$\varphi(x,y) = \varphi(x)\varphi(y).$$

Therefore the integer shifts  $\{\varphi(x-k, y-n) = \varphi_{0,k,n}\}_{k,n \in \mathbb{Z}}$  form an orthonormal basis of  $\mathcal{V}_0$ , consecutive approximation spaces are

connected via scaling by 2 on both variables, and the other conditions are clear.

The orthogonal complement of  $\mathcal{V}_j$  in  $\mathcal{V}_{j+1}$  is denoted by  $\mathcal{W}_j$ . The rules of arithmetic are valid for direct sums and tensor products of subspaces, namely

$$\begin{aligned}\mathcal{V}_{j+1} &= V_{j+1} \otimes V_{j+1} \\ &= (V_j \oplus W_j) \otimes (V_j \oplus W_j) \\ &= (V_j \otimes V_j) \oplus [(V_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (W_j \otimes W_j)] \\ &= \mathcal{V}_j \oplus \mathcal{W}_j.\end{aligned}$$

Thus the space  $\mathcal{W}_j$  can be viewed as the direct sum of three tensor products, namely

$$\mathcal{W}_j = (W_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (V_j \otimes W_j).$$

Therefore *three wavelets* are necessary to span the detail spaces:

$$\psi^d(x, y) = \psi(x)\psi(y), \quad \psi^v(x, y) = \psi(x)\varphi(y), \quad \psi^h(x, y) = \varphi(x)\psi(y),$$

where  $d$  stands for diagonal,  $v$  for vertical, and  $h$  for horizontal. The reason for these names is that each of the subspaces somehow favors details in those directions.

The same approach works in higher dimensions. There are  $2^n - 1$  wavelet functions and one scaling function, where  $n$  is the dimension.

**Exercise 10.31.** Describe a three-dimensional MRA. (These are useful for video compression.)  $\diamond$

This construction has the advantage that the bases are separable, implementing the fast two dimensional wavelet transform is not difficult. In fact it can be done by successively applying the one-dimensional FWT. The disadvantage is that the analysis is very axis-dependent, which might not be desirable for certain applications.

**Example 10.32.** (*The Two-Dimensional Haar Basis*) The scaling function is the characteristic function

$$\varphi(x, y) = \chi_{[0,1]^2}(x, y)$$

of the unit cube. The following pictures help us to understand the nature of the two-dimensional Haar wavelets and scaling function.

$$\begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array}$$

$$\varphi(x, y)$$

$$\begin{array}{|c|c|} \hline -1 & 1 \\ \hline 1 & -1 \\ \hline \end{array}$$

$$\psi^d(x, y)$$

$$\begin{array}{|c|c|} \hline 1 & -1 \\ \hline 1 & -1 \\ \hline \end{array}$$

$$\psi^h(x, y)$$

$$\begin{array}{|c|c|} \hline -1 & -1 \\ \hline 1 & 1 \\ \hline \end{array}$$

$$\psi^v(x, y)$$

◇

**Exercise 10.33.** Describe the two-dimensional discrete Haar MRA.

◇

There are *non-separable* two-dimensional MRA's. The most famous one corresponds to an analogue of the Haar basis. The scaling function is the characteristic function of a two-dimensional set. It turns out that the set has to be rather complicated. In fact it is a self-similar set with fractal boundary, the so-called *twin dragon*.

## 10.9. Basics of compression and denoising

One of the main goals in signal and image processing is to be able to code the information with as few data as possible, as we saw in the case of Amanda and her mother in Chapter 1. Having fewer data points allows for faster transmission and easier storage. In the presence of noise, one also wants to separate the noise from the signal, and one would like to have a basis that concentrates the signal in a few large coefficients and delegates the noise to very small coefficients.

In both the deterministic and the noisy case, the steps to follow are:

- Transform the data, find coefficients with respect to a given basis.
- Threshold the coefficients. Essentially one keeps the large ones and discards the small ones. Information is lost in this step, so perfect reconstruction is no longer possible.
- Reconstruct with the thresholded coefficients, and hope that the resulting compressed signal is a good approximation to your original signal, in other words that you have successfully denoised the original.

Wavelet bases are good for decorrelating coefficients. They are also good for denoising in the presence of *white noise*.

The crudest approach would be to use the projection into an approximation space as your compressed signal, discarding all the details after certain scale  $j$ :

$$P_j f = \sum_k \langle f, \varphi_{j,k} \rangle \varphi_{j,k}.$$

The noise is usually concentrated in the finer scales (higher frequencies!), so this approach does denoise the signal, but at the same time it removes many of the sharp features of the signal that were encoded in the finer wavelet coefficients. A more refined technique is required, called *thresholding*. There are different thresholding techniques. The most popular are *hard thresholding* (it is a keep-or-toss thresholding), and *soft thresholding* (the coefficients are attenuated following a linear scheme). There is also the issue about thresholding individual coefficients or *block-thresholding*. How to select the threshold is another issue. In the denoising case, there are some thresholding selection rules that are justified by probability theory (essentially the law of large numbers), and are used widely by statisticians. Both thresholding and threshold selection principles are encoded in MATLAB.

In traditional approximation theory there are two possible methods, *linear* approximation and *non-linear* approximation.

- *Linear approximation*: In linear approximation, one selects *a priori*  $N$  elements in the basis and projects onto the subspace generated by those elements, regardless of the function that is being approximated. It is a linear scheme:

$$P_N^l f = \sum_{n=1}^N \langle f, \psi_n \rangle \psi_n.$$

- *Non-linear approximation*: In non-linear approximation, one chooses the basis elements depending on the function. For example the  $N$  basis elements could be chosen so that the coefficients are the largest in size for the particular function. This time the chosen basis elements depend on the function being

approximated:

$$P_N^{\text{nl}} f = \sum_{n=1}^N \langle f, \psi_{n,f} \rangle \psi_{n,f}.$$

The non-linear approach has proven quite successful. There is a lot more information about these issues in [Mall98, Chapters 9 and 10]. See also Project 12.5.

$\varphi$   $\psi$   
db2  
  
db4  
  
db6

**Figure 10.4.** Graphs of The Daubechies scaling function and wavelet for filters length 4, 8, and 12. Note how the smoothness increases with filter length. Also note the small wiggles away from the main oscillation.