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 10

Zooming properties of wavelets, and applications

We highlight the *zooming* properties of the Haar system, and how they can be mathematically encoded in the so-called *multiresolution analysis* (MRA). The MRA provides a framework for the construction of most wavelets, this is the celebrated Mallat's Theorem that we describe in detail in Section 10.3. We discuss how to implement the wavelet transform via filter banks. We present, in a very informal manner, competing attributes we would like the wavelets to have, and a by no means exhaustive catalog of wavelets. We briefly discuss wavelet packets and two-dimensional wavelets used in image processing, as well as the use of wavelet decompositions in compression and denoising of images and signals.

10.1. Multiresolution analyses (MRAs)

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

(1) $\cdots \subset \mathbf{V}_{-2} \subset \mathbf{V}_{-1} \subset \mathbf{V}_0 \subset \mathbf{V}_1 \subset \mathbf{V}_2 \subset \cdots \subset L^2(\mathbb{R})$ (nested)

(2) $\bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$ (trivial intersection),

(3) $\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j$ is dense in $L^2(\mathbb{R})$ (density in $L^2(\mathbb{R})$),

(4) $f(x) \in \mathbf{V}_j$ if and only if $f(2x) \in \mathbf{V}_{j+1}$ (scaling property),

(5) $f(x) \in \mathbf{V}_0$ if and only if $f(x-k) \in \mathbf{V}_0$ for any $k \in \mathbb{Z}$ (shift invariance),

(6) There exists a *scaling function* $\varphi \in \mathbf{V}_0$ such that its integer translates, $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$, form an orthonormal basis for $\mathbf{V}_0$.

Given the scaling function φ, we denote its integer translates and dyadic dilates with subscripts j, k, as we did for the wavelet ψ:

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

EXERCISE 10.1. Show that $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $\mathbf{V}_j$. ◊

Given an L^2-function f, let $P_j f$ be the *orthogonal projection of f onto* $\mathbf{V}_j$:

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

The function $P_j f$ is an approximation to the original function at scale 2^{-j}. More precisely, it is the *best approximation* in the subspace $\mathbf{V}_j$ to f. See Theorem 5.36, 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_jf to the better approximation $P_{j+1}f$? Define the *difference operator* Q_j by

$$Q_jf := P_{j+1}f - P_jf.$$

To recover $P_{j+1}f$ we add Q_jf to P_jf. It is clear that $P_{j+1} = P_j + Q_j$, and that the orthogonal projection onto $\mathbf{V}_j$ of $P_{j+1}f$ is orthogonal to the 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 $\mathbf{W}_j$, which we call the *detail subspace* at scale 2^{-j}. The space $\mathbf{W}_j$ is the *orthogonal complement* of $\mathbf{V}_j$ in $\mathbf{V}_{j+1}$. This means that $\mathbf{V}_j \perp \mathbf{W}_j$, and if $f \in \mathbf{V}_{j+1}$, there exist unique $g \in \mathbf{V}_j$ and $h \in \mathbf{W}_j$ such that $f = g + h$. In fact $g = P_jf$ and $h = Q_jf$. We use the notation already introduced at the end of Chapter 5 to denote the direct sum of two orthogonal subspaces:

$$\mathbf{V}_{j+1} = \mathbf{V}_j \oplus \mathbf{W}_j.$$

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

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

$$\mathbf{V}_j = \mathbf{V}_n \oplus \mathbf{W}_n \oplus \mathbf{W}_{n-1} \oplus \cdots \oplus \mathbf{W}_{j-2} \oplus \mathbf{W}_{j-1}.$$

Show also that if $j \neq k$ then $\mathbf{W}_j \perp \mathbf{W}_k$. Hence we get an orthogonal decomposition of each subspace $\mathbf{V}_j$ in terms of the less accurate approximation space $\mathbf{V}_n$ and the detail subspaces $\mathbf{W}_k$ at intermediate resolutions $n \leq k < j$.

The nested subspaces $\{\mathbf{V}_j\}$ define an MRA. Therefore the detail subspaces give an orthogonal decomposition of $L^2(\mathbb{R})$:

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} \mathbf{W}_j. \tag{10.3}$$

We show in Section 10.3 that the scaling function φ determines a *wavelet* ψ such that $\{\psi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $\mathbf{W}_0$. The detail subspace $\mathbf{W}_j$ is a dilation of $\mathbf{W}_0$, therefore the function

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

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

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

Exercise 10.4. Show that the detail subspace $\mathbf{W}_j$ is a dilation of $\mathbf{W}_0$, that is it obeys the same scale invariance property (4) that the approximation subspaces V_j satisfy.

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 10.3.

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

EXERCISE 10.6. Given an orthogonal MRA with scaling function φ, show that if $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $\mathbf{W}_0$, then $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $\mathbf{W}_j$. Show that the two-parameter family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis for $L^2(\mathbb{R})$. $\Diamond$

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 $\mathbf{V}_j$ corresponds to step functions with steps on the intervals $[k2^{-j}, (k+1)2^{-j})$, and they have all the properties listed. The Haar function is the wavelet ψ in Mallat's theorem. We return to this example in Section 10.2.

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

EXERCISE 10.9. Verify that the subspaces $\mathbf{V}_j$ defined in Example 10.8 do generate an MRA. Verify that the subspace $\mathbf{W}_j$ is the orthogonal complement of $\mathbf{V}_j$ in $\mathbf{V}_{j+1}$. $\Diamond$

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

10.2. The Haar multiresolution analysis

Before considering how to construct wavelets from scaling functions associated to an MRA, we revisit the simple example that predates the main development of wavelets, namely the Haar multiresolution analysis.

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

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

The subspace $\mathbf{V}_0$ is the closure in $L^2(\mathbb{R})$ of the linear span of the integer translates of the Haar scaling function φ,

$$\mathbf{V}_0 := \overline{\operatorname{span}(\{\varphi(x-k)\}_{k\in\mathbb{Z}})}.$$

It consists of piecewise constant functions with jumps only at the integers, and such that the coefficients (possibly infinitely many of them) decay fast enough so as to belong to $\ell_2(\mathbb{Z})$, more precisely,

$$\mathbf{V}_0 = \{f = \sum_{k\in\mathbb{Z}} a_k\varphi_{0,k} : \sum_{k\in\mathbb{Z}} |a_k|^2 < \infty\}.$$

The subspace

$$\mathbf{V}_j := \overline{\operatorname{span}(\{\varphi_{j,k}\}_{k\in\mathbb{Z}})}$$

is the subspace consisting of piecewise constant functions with jumps only at the integer multiples of 2^{-j}, and such that the coefficients (possibly infinitely many of them) decay fast enough so as to belong to $\ell_2(\mathbb{Z})$ to ensure that we stay in $L^2(\mathbb{R})$, more precisely,

$$\mathbf{V}_j = \{f = \sum_{k\in\mathbb{Z}} a_k\varphi_{j,k} : \sum_{k\in\mathbb{Z}} |a_k|^2 < \infty\}.$$

The orthogonal projection P_jf onto $\mathbf{V}_j$ is the piecewise constant function with jumps at the integer multiples of 2^{-j}, whose value on the interval $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ is given by the integral average of f over the interval $I_{j,k}$. To go from P_jf to $P_{j+1}f$, we add the difference Q_jf (the expectation operators P_j and the difference operators Q_j were defined in Section 9.3.3.) We showed in Lemma 9.25 that Q_jf coincides with

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

where ψ is the Haar wavelet, defined by

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

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

$$\mathbf{W}_j = \overline{\operatorname{span}(\{\psi_{j,k}\}_{k\in\mathbb{Z}})},$$

consists of piecewise constant functions 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_jf at resolution j as successive approximations to the original signal $f \in L^2(\mathbb{R})$. These approximations are the orthogonal projections P_jf onto the approximation spaces $\mathbf{V}_j$. The *details*, necessary to move from level j to the next level $(j+1)$, are encoded in the Haar coefficients at level j, more precisely in the orthogonal projections Q_jf onto the detail subspaces $\mathbf{W}_j$. Starting at a low resolution level, we can obtain better and better resolution by adding the details at the subsequent levels. As $j \to \infty$, the resolution is increased. The steps get smaller (length 2^{-j}), and the approximation converges to f in L^2-norm (this is the content of equation (9.15) in Theorem 9.27), and a.e. (Lebesgue Differentiation Theorem). Clearly the subspaces are nested, that is, $\mathbf{V}_j \subset \mathbf{V}_{j+1}$, and their intersection is the trivial subspace containing just the zero function (this is equation (9.14) in Theorem 9.27). Lo and behold, we have shown that the Haar scaling function generates an 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 $\mathbf{V}_{-n}$ and a finest scale $\mathbf{V}_0$, truncate the chain to

$$\mathbf{V}_{-n} \subset \cdots \subset \mathbf{V}_{-2} \subset \mathbf{V}_{-1} \subset \mathbf{V}_0,$$

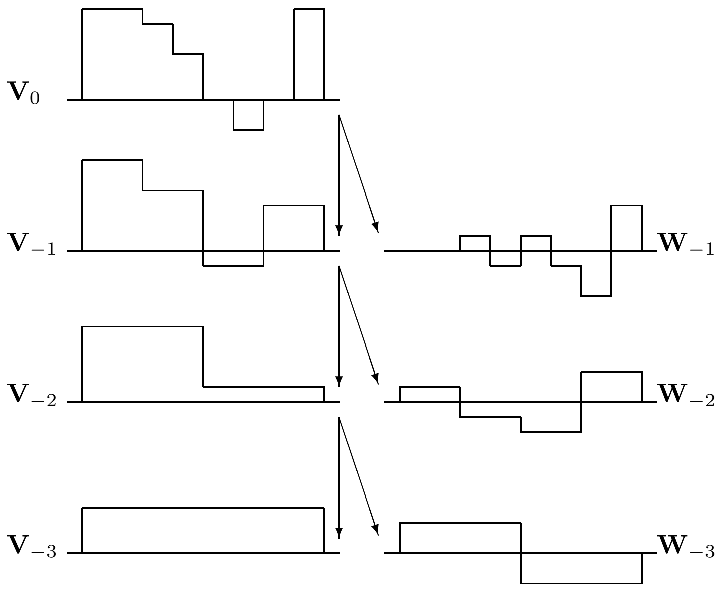


FIGURE 10.1. A wavelet decomposition: $\mathbf{V}_0 = \mathbf{V}_{-3} \oplus \mathbf{W}_{-3} \oplus \mathbf{W}_{-2} \oplus \mathbf{W}_{-1}$.

and obtain

$$\mathbf{V}_0 = \mathbf{V}_{-n} \oplus \mathbf{W}_{-n} \oplus \mathbf{W}_{-n+1} \oplus \mathbf{W}_{-n+2} \oplus \cdots \oplus \mathbf{W}_{-1}. \tag{10.4}$$

We will go through the decomposition process in the text using vectors. It is also enlightening to look at the graphical version in Figure 10.1.

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 $\mathbf{V}_0$. The convention we are using throughout the example is that the vector $v = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7] \in \mathbb{C}^8$ represents the step function $f(x) = a_j$ for $j \le x < j + 1$, $j = 0, 1, \ldots 7$, $f(x) = 0$ otherwise.

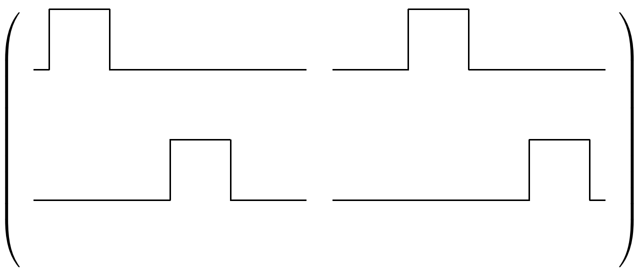
EXERCISE 10.10. Use the above convention to describe the scaling function in $\mathbf{V}_0$, $\mathbf{V}_{-1}$, $\mathbf{V}_{-2}$, $\mathbf{V}_{-3}$ supported on $[0, 8)$ and the Haar functions in $\mathbf{W}_{-1}$, $\mathbf{W}_{-2}$, $\mathbf{W}_{-3}$ supported on $[0, 8)$.

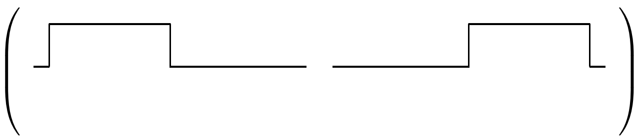
To construct the projection onto $\mathbf{V}_{-1}$ we average pairs of values, obtaining

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

The difference is in $\mathbf{W}_{-1}$, so we have

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

$\mathbf{V}_{-1} = \text{span}$ 

$\mathbf{V}_{-2} = \text{span}$ 

$\mathbf{V}_{-3} = \text{span}$ 

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

By repeating this process, we obtain

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

For the example in Figure 10.1, 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 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 $\mathbb{C}^8$.

The scaling function φ satisfies the *two-scale recurrence equation*

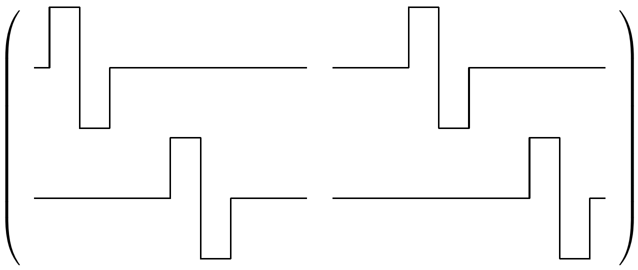
$$\varphi(t) = \varphi(2t) + \varphi(2t-1). \tag{10.5}$$

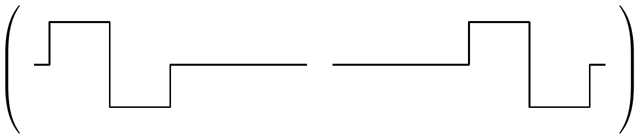
Therefore $\varphi_{j,k} = (\varphi_{j-1,2k} + \varphi_{j-1,2k+1})/\sqrt{2}$, and so

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

Thus we can also compute

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

$\mathbf{W}_{-1} = \text{span}$ 

$\mathbf{W}_{-2} = \text{span}$ 

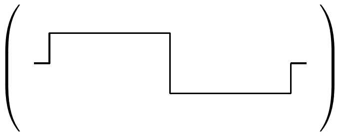
$\mathbf{W}_{-3} = \text{span}$ 

FIGURE 10.3. The wavelet subspaces used in Figure 10.1.

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*,

$$\psi(t) = \varphi(2t) - \varphi(2t-1), \tag{10.6}$$

and thus we can recursively compute

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

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.

The coefficients $\langle f, \psi_{j,k} \rangle$ for fixed j are called the *differences* or *details* of f at scale j, and denoted $d_{j,k}$. Evaluating the whole set of Haar coefficients $d_{j,k}$ and averages $a_{j,k}$ requires $2(N-1)$ additions and $2N$ multiplications. The discrete wavelet transform can be performed using a similar *cascade algorithm* with complexity N, where N is the number of data points. Let us remark that an arbitrary change of basis in N-dimensional space requires multiplication by an $N \times N$ matrix, hence *a priori* one requires N^2 multiplications. This is the same algorithm as the fast Haar transform we discussed in Chapter 6 in the language of matrices.

10.3. From MRA to wavelets: Mallat's Theorem

In this section we show, given an orthogonal MRA with scaling function φ, how to find the corresponding wavelet.

First let us find necessary and sufficient conditions on the Fourier side that guarantee that the family of integer translates of an square-integrable function forms an orthonormal family.

LEMMA 10.13. *Take $f \in L^2(\mathbb{R})$. The family $\{f_{0,k} = \tau_k f\}_{k\in\mathbb{Z}}$ of integer translates of f is orthonormal if and only if for almost every $\xi \in \mathbb{R}$*

$$\sum_{n\in\mathbb{Z}} |\widehat{f}(\xi+n)|^2 = 1.$$

PROOF. First note that by a change of variable,

$$\langle \tau_k f, \tau_m f\rangle = \langle \tau_{k-m} f, f\rangle.$$

The orthonormality of the family of integer translates is equivalent to

$$\langle \tau_k f, f\rangle = \delta_k, \quad \text{for all } k \in \mathbb{Z}.$$

Recall that the Fourier transform preserves inner products, the Fourier transform of $\tau_k f$ is the modulation of $\widehat{f}$, and the function $e^{-2\pi i k\xi}$ has period one. Therefore for all $k \in \mathbb{Z}$

$$\begin{aligned}
\delta_k &= \langle \widehat{\tau_k f}, \widehat{f}\rangle \\
&= \int_{\mathbb{R}} e^{-2\pi i k\xi} |\widehat{f}(\xi)|^2 \, d\xi \\
&= \sum_{n\in\mathbb{Z}} \int_n^{n+1} e^{-2\pi i k\xi} |\widehat{f}(\xi)|^2 \, d\xi \\
&= \int_0^1 e^{-2\pi i k\eta} \sum_{n\in\mathbb{Z}} |\widehat{f}(\eta+n)|^2 \, d\eta.
\end{aligned}$$

The last equality is obtained by performing on each integral the change of variable $\eta = \xi - n$ that maps the interval $[n, n+1)$ onto the unit interval $[0,1)$. This identity says that the periodic function of period one given by

$$F(\eta) = \sum_{n\in\mathbb{Z}} |\widehat{f}(\eta+n)|^2$$

has k^{th} Fourier coefficient equal to the Kronecker delta δ_k. Therefore it must be equal to one almost everywhere, which is exactly what we set out to prove. □

Let φ be the scaling function of an orthogonal MRA. Then its integer translates form an orthonormal family. Therefore, by Lemma 10.13, for almost every $\xi \in \mathbb{R}$

$$\sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi+n)|^2 = 1. \tag{10.7}$$

Also, $\varphi \in \mathbf{V}_0 \subset \mathbf{V}_1$, and the functions $\varphi_{1,k}(x) = \sqrt{2}\varphi(2x-k)$, for $k \in \mathbb{Z}$, form an orthonormal basis for $\mathbf{V}_1$. This means that the following *scaling equation* holds, for some coefficients $\{h_k\}$ such that $\sum_k |h_k|^2 < \infty$:

$$\varphi(x) = \sum_{k\in\mathbb{Z}} h_k \varphi_{1,k}(x) = \sqrt{2} \sum_{k\in\mathbb{Z}} h_k \varphi(2x-k). \tag{10.8}$$

On the Fourier side, the scaling equation reads

$$\widehat{\varphi}(\xi) = H(\xi/2)\widehat{\varphi}(\xi/2), \tag{10.9}$$

where $H(\xi) = (1/\sqrt{2}) \sum h_k e^{-2\pi i k\xi}$ is a function of period 1, called the *refinement mask*.

For convenience we assume that H is a trigonometric polynomial, so that all but finitely many of the coefficients $\{h_k\}$ vanish. There are multiresolution analyses whose refinement masks are trigonometric polynomials, for example the Haar MRA. In fact, for applications, these are the most useful, and they correspond to MRAs with compactly supported scaling functions $\boxed{\text{ref: Daubechies?}}$.

EXERCISE 10.14. Check that the scaling equation on the Fourier side is given by equation (10.9). ◊

We can now deduce a necessary property that the refinement mask H must satisfy, namely the so-called *quadrature mirror filter* (QMF) property.

LEMMA 10.15. *Given an orthogonal MRA with scaling function φ, and corresponding refinement mask H that is a trigonometric polynomial, then*

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

PROOF. Insert equation (10.9) into equation (10.7), obtaining

$$\begin{aligned} 1 &= \sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi + n)|^2 \\ &= \sum_{n\in\mathbb{Z}} |H((\xi + n)/2)|^2 |\widehat{\varphi}((\xi + n)/2)|^2. \end{aligned}$$

Now separate the sum over the odd and even integers, use the fact that H has period one to factor it out from the sum, and use (twice) equation (10.7), which holds for almost every point ξ:

$$\begin{aligned} 1 &= |H(\xi/2)|^2 \sum_{k\in\mathbb{Z}} |\widehat{\varphi}(\xi/2 + k)|^2 + |H(\xi/2 + 1/2)|^2 \sum_{k\in\mathbb{Z}} |\widehat{\varphi}((\xi + 1)/2 + k)|^2 \\ &= |H(\xi/2)|^2 + |H(\xi/2 + 1/2)|^2. \end{aligned}$$

Equality holds almost everywhere. Since H is a trigonometric polynomial, H is continuous and so equality must hold everywhere. □

Note that if $f \in \mathbf{V}_1$, the same argument we used for φ shows that there must be a function of period one, $m_f(\xi) \in L^2([0,1))$, such that

$$\widehat{f}(\xi) = m_f(\xi/2)\widehat{\varphi}(\xi/2). \tag{10.10}$$

In this notation, the refinement mask $H = m_\varphi$.

We now describe the functions in $\mathbf{W}_0$, the orthogonal complement of $\mathbf{V}_0$ in $\mathbf{V}_1$, where all the subspaces correspond to an orthogonal MRA with scaling function φ, and refinement mask a trigonometric polynomial H.

LEMMA 10.16. *A function $f \in \mathbf{W}_0$ if and only if there is a function $v(\xi)$ of period one such that*

$$\widehat{f}(\xi) = e^{\pi i \xi} v(\xi) \overline{H(\xi/2 + 1/2)} \widehat{\varphi}(\xi/2).$$

PROOF. First, $f \in \mathbf{W}_0$ if and only if $f \in \mathbf{V}_1$ and $f \perp \mathbf{V}_0$. The fact that $f \in \mathbf{V}_1$ allows us to reduce the problem to showing that

$$m_f(\xi) = e^{2\pi i \xi} \sigma(\xi) \overline{H(\xi + 1/2)},$$

where $\sigma(\xi)$ is some function with period $1/2$. (Then $v(\xi) = \sigma(\xi/2)$ will have period one.)

The orthogonality $f \perp V_0$ is equivalent to $\langle f, \varphi_{0,k}\rangle = 0$ for all $k \in \mathbb{Z}$.

A calculation similar to the one in the proof of Lemma 10.13 shows that

$$\begin{aligned} 0 &= \langle \widehat{f}, \widehat{\varphi_{0,k}} \rangle \\ &= \int_{\mathbb{R}} \widehat{f}(\xi) e^{2\pi i k \xi} \overline{\widehat{\varphi}(\xi)}\, d\xi \\ &= \int_{\mathbb{R}} e^{2\pi i k \xi} m_f(\xi/2) \widehat{\varphi}(\xi/2) \overline{H(\xi/2) \widehat{\varphi}(\xi/2)}\, d\xi \\ &= \int_{\mathbb{R}} e^{2\pi i k \xi} m_f(\xi/2) \overline{H(\xi/2)} |\widehat{\varphi}(\xi/2)|^2\, d\xi. \end{aligned}$$

In the third line we have used equations (10.10) and (10.9).

At this point we want to use the same trick we used in Lemma 10.13: break the integral over $\mathbb{R}$ into the sum of integrals over the intervals $[n, n+1)$, change variables to the unit interval, and use the periodicity of the exponential to get

$$\begin{aligned} 0 &= \sum_{n\in\mathbb{Z}} \int_n^{n+1} e^{2\pi i k \xi} m_f(\xi/2) \overline{H(\xi/2)} |\widehat{\varphi}(\xi/2)|^2\, d\xi \\ &= \int_0^1 e^{2\pi i k \xi} \sum_{n\in\mathbb{Z}} m_f((\xi+n)/2) \overline{H((\xi+n)/2)} |\widehat{\varphi}((\xi+n)/2)|^2\, d\xi. \end{aligned}$$

We would also like to take advantage of the periodicity of the functions m_f and H. These are function of period one, but we are adding half an integer. If we separate the last sum, as we did in Lemma 10.15, into the sums over the odd and even integers, we will win. In fact we obtain that for all $k \in \mathbb{Z}$

$$\begin{aligned} 0 &= \int_0^1 e^{2\pi i k \xi} [m_f(\xi/2) \overline{H(\xi/2)} \sum_{m\in\mathbb{Z}} |\widehat{\varphi}(\xi/2+m)|^2 \\ &\quad + m_f(\xi/2+1/2) \overline{H(\xi/2+1/2)} \sum_{m\in\mathbb{Z}} |\widehat{\varphi}((\xi+1)/2+m)|^2]\, d\xi \\ &= \int_0^1 e^{2\pi i k \xi} \left[m_f(\xi/2) \overline{H(\xi/2)} + m_f(\xi/2+1/2) \overline{H(\xi/2+1/2)} \right] d\xi, \end{aligned}$$

where the last equality is a consequence of applying equation (10.7) twice. This time the 1-periodic function (see Exercise 10.17)

$$F(\xi) = m_f(\xi/2) \overline{H(\xi/2)} + m_f(\xi/2+1/2) \overline{H(\xi/2+1/2)},$$

has been shown to have Fourier coefficients identically equal to zero. Hence it must be the zero function almost everywhere:

$$m_f(\xi) \overline{H(\xi)} + m_f(\xi+1/2) \overline{H(\xi+1/2)} = 0 \quad \text{a.e.} \tag{10.11}$$

This equation says that for a ξ for which it holds, the vector $\vec{v} \in \mathbb{C}^2$ given by

$$\vec{v} = \big(m_f(\xi), m_f(\xi+1/2)\big),$$

must be orthogonal to the vector in $\vec{w} \in \mathbb{C}^2$ given by

$$\vec{w} = \big(H(\xi), H(\xi+1/2)\big).$$

Notice that the QMF property of H (see Lemma 10.15) ensures that $\vec{w}$ is not zero; in fact $|\vec{w}| = 1$. The vector space $\mathbb{C}^2$ over the complex numbers $\mathbb{C}$ is a two-dimensional space. Therefore the orthogonal complement of the one-dimensional subspace generated by the vector $\vec{w}$ is one-dimensional. It suffices to locate one non-zero vector $\vec{u} \in C^2$ that is orthogonal to $\vec{w}$ to completely characterize all the vectors $\vec{v}$ that are orthogonal to $\vec{w}$, in fact $\vec{v} = \lambda\vec{u}$ for $\lambda \in \mathbb{C}$. The vector

$$\vec{u} = \big(-\overline{H(\xi + 1/2)}\,, \overline{H(\xi)}\,\big)$$

is orthogonal to $\vec{w}$. Therefore $\vec{v} = \lambda\big(-\overline{H(\xi + 1/2)}\,, \overline{H(\xi)}\,\big)$, and we conclude that the 1-periodic function m_f must satisfy

$$m_f(\xi) = -\lambda(\xi)\overline{H(\xi + 1/2)}, \quad m_f(\xi + 1/2) = \lambda(\xi)\overline{H(\xi)}.$$

Therefore $\lambda(\xi)$ is a function of period one such that $-\lambda(\xi + 1/2) = \lambda(\xi)$. Equivalently, $\lambda(\xi) = e^{2\pi i\xi}\sigma(\xi)$, where $\sigma(\xi)$ has period $1/2$ (see Exercise 10.18). The conclusion is that $f \in \mathbf{W}_0$ if and only if

$$m_f(\xi) = e^{2\pi i\xi}\sigma(\xi)\overline{H(\xi + 1/2)},$$

where $\sigma(\xi)$ is a function with period $1/2$, as required. □

Exercise 10.17. Show that if H and G are periodic functions of period one, then the new function

$$F(\xi) = G(\xi/2)\overline{H((\xi/2)} + G(\xi/2 + 1/2)\overline{H(\xi/2 + 1/2)}$$

also has period one. ◊

Exercise 10.18. Show that $\lambda(\xi)$ is a function of period one such that

$$\lambda(\xi + 1/2) = -\lambda(\xi)$$

if and only if $\lambda(\xi) = e^{2\pi i\xi}\sigma(\xi)$ where $\sigma(\xi)$ has period $1/2$. ◊

We are now ready to present the wavelet ψ associated to the MRA with scaling function φ and refinement mask H, where H is a trigonometric polynomial. The function ψ we are looking for is in $\mathbf{W}_0$. Therefore on the Fourier side it must satisfy the equation

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

where

$$m_\psi(\xi) = e^{2\pi i\xi}\sigma(\xi)\overline{H(\xi + 1/2)}, \tag{10.12}$$

where $\sigma(\xi)$ is a function with period $1/2$. Furthermore, because we want the integer translates of ψ to form an orthonormal system, we can apply Lemma 10.13 to ψ and deduce a QMF property for the 1-periodic function $m_\psi(\xi)$. Namely, for almost every $\xi \in \mathbb{R}$,

$$|m_\psi(\xi)|^2 + |m_\psi(\xi + 1/2)|^2 = 1.$$

Substituting equation (10.12) into this equation implies that for almost every $\xi \in \mathbb{R}$,

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

But $\sigma(\xi)$ has period $1/2$, and H satisfies the QMF condition. We conclude that $|\sigma(\xi)| = 1$ almost everywhere.

Proof of Mallat's Theorem (Theorem 10.5). Choose the simplest possible function of period $1/2$ that has absolute value 1 all the time, namely $\sigma(\xi) \equiv 1$. Define the wavelet on the Fourier side to be

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

where G is the 1-periodic function given by

$$G(\xi) := e^{2\pi i\xi}\overline{H(\xi + 1/2)}.$$

By Lemma 10.16, $\psi \in \mathbf{W}_0$. Its integer translates are also in $\mathbf{W}_0$, because

$$\widehat{\psi_{0,k}}(\xi) = e^{-2\pi i k\xi} G(\xi/2)\widehat{\varphi}(\xi/2).$$

Furthermore G satisfies a QMF property, and so the family of integer translates of ψ is an orthonormal family in $\mathbf{W}_0$.

It remains to show that this family spans $\mathbf{W}_0$. By Lemma 10.16, there is a square-integrable function $v(\xi)$ of period one such that

$$\widehat{f}(\xi) = v(\xi)e^{\pi i\xi}\overline{H(\xi/2+1/2)}\widehat{\varphi}(\xi/2) = v(\xi)\widehat{\psi}(\xi).$$

But $v(\xi) = \sum_{k\in\mathbb{Z}} a_k e^{-2\pi i k\xi}$, where $\sum_{k\in\mathbb{Z}} |a_k|^2 < \infty$. Inserting the trigonometric expansion of v, and using the fact that modulations on Fourier side come from translations, we obtain

$$\widehat{f}(\xi) = \sum_{k\in\mathbb{Z}} a_k e^{-2\pi i k\xi}\widehat{\psi}(\xi) = \sum_{k\in\mathbb{Z}} a_k \widehat{\psi_{0,k}}(\xi).$$

Taking the inverse Fourier transform, we see that

$$f = \sum_{k\in\mathbb{Z}} a_k \psi_{0,k}.$$

That is, f belongs to the span of the integer translates of ψ. The integer translates of ψ form an orthonormal basis of $\mathbf{W}_0$. By scale invariance, the functions $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ form an orthonormal basis of $\mathbf{W}_j$. Thus the family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis of $L^2(\mathbb{R})$, as required. □

EXERCISE 10.19. Verify that if $\tau_k\psi$ is an orthonormal system, and $\psi \in \mathbf{W}_0$, then m_ψ satisfies the QMF property. ◊

10.4. Mallat's algorithm revisited

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

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

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

$$\varphi(t) = \sum_{k\in\mathbb{Z}} h_k\varphi_{1,k}(t) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k\varphi(2t-k). \tag{10.13}$$

In general this sum is not finite, but whenever it is, both the scaling function φ and the wavelet ψ are *compactly supported*. In the case of the Haar MRA, we have $h_0 = h_1 = 1/\sqrt{2}$, and all other coefficients vanish. The sequence $h = \{h_k\}$ is the so-called *low-pass filter*. We assume that the low-pass filter has finite length L. (It turns out that such a filter always has even length, $L = 2M$ say. The *refinement mask* is given by $H(\xi) = (1/\sqrt{2})\sum h_k e^{-2\pi i k\xi}$, which can be viewed as a 1-periodic function, of the frequency variable ξ, whose Fourier coefficients are $\widehat{H}(n) = h_{-n}/\sqrt{2}$.

Not all wavelets have compact support, we already encoutnered 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.

We observed in (10.9) that on Fourier side the scaling equation becomes

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

We can iterate this formula to obtain

$$\widehat{\varphi}(\xi) = \left(\Pi_{j=0}^{N} H(\xi/2^j)\right)\widehat{\varphi}(\xi/2^N). \tag{10.14}$$

If $\widehat{\varphi}$ is continuous at $\xi = 0$, $\widehat{\varphi}(0) \neq 0$, and the infinite product converges, then there is a solution to the scaling equation. It turns out that to obtain orthonormality of the set $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ we must have $|\widehat{\varphi}(0)| = 1$, and one usually normalizes to $\widehat{\varphi}(0) = \int\varphi(t)dt = 1$. This normalization happens to be useful in numerical implementations of the wavelet transform.

The conditions on the filter H that guarantee the existence of a solution φ to the scaling equation are now well understood (consult [**HW**] for more details). For example, the fact that the infinite product $\Pi_{j=0}^{\infty} H\left(\frac{\xi}{2^j}\right)$ must converge for each ξ, in particular for $\xi = 0$, forces $H(0) = 1$, or equivalently $\sum_{k=0}^{L-1} h_k = \sqrt{2}$. In Lemma 10.15 we showed that the orthonormality of the integer shifts of the scaling function implies that

$$|H(\xi)|^2 + |H(\xi+1/2)|^2 = 1. \tag{10.15}$$

Equation (10.15) is known in the engineering community as a *quadrature mirror filter* (QMF) condition, necessary for exact reconstruction for a pair of filters. The QMF condition together with $H(0) = 1$ implies that $H(1/2) = 0$, which explains the name *low-pass filter*: low frequencies near $\xi = 0$ are kept, while high frequencies near $\xi = 1/2$ are removed (filtered out).

The wavelet ψ we are seeking is an element of $\mathbf{W}_0 \subset \mathbf{V}_1$. Therefore it is also a superposition of the basis elements $\{\varphi_{1,k}\}_{k\in\mathbb{Z}}$ of $\mathbf{V}_1$. So there are coefficients $\{g_k\}$ such that $\sum_{k\in\mathbb{Z}} |g_k|^2 < \infty$, and

$$\psi(t) = \sum_{k\in\mathbb{Z}} g_k\varphi_{1,k}(t). \tag{10.16}$$

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

$$g_k = (-1)^{k-1}\overline{h_{1-k}}, \qquad G(\xi) = \frac{1}{\sqrt{2}}\sum_{k=0}^{L-1} g_k e^{2\pi ik\xi}. \tag{10.17}$$

With this choice of filter, the function ψ given by equation (10.16) is Mallat's wavelet, which we constructed in Section 10.3. It suffices to observe that with this choice, $\widehat{\psi}(\xi) = G(\xi/2)\widehat{\varphi}(\xi/2)$ with $G(\xi) := (1/\sqrt{2})\sum g_k e^{-2\pi ik\xi} = e^{2\pi i\xi}\overline{H(\xi+1/2)}$. See Exercise 10.20.

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

$$|G(\xi)|^2 + |G(\xi+1/2)|^2 = 1, \tag{10.18}$$

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

$$H(\xi)\overline{G(\xi)} + H(\xi+1/2)\overline{G(\xi+1/2)} = 0. \tag{10.19}$$

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

Given a low-pass filter H associated to an MRA we will always construct the associated high-pass filter G according to formula (10.17).

Exercise 10.20. Check that with this choice, $G(\xi) = e^{2\pi i\xi}\overline{H(\xi+1/2)}$, and that if H satisfies a QMF condition, so does G. Check that equation (10.19) holds. Verify that on the Fourier side

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

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

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.

$$\boxed{\begin{matrix}\varphi\\ \text{scaling MRA}\end{matrix}} \to \boxed{\begin{matrix}H\\ \text{low-pass filter}\end{matrix}} \to \boxed{\begin{matrix}G\\ \text{high-pass filter}\end{matrix}} \to \boxed{\begin{matrix}\psi\\ \text{Mallat's wavelet}\end{matrix}}$$

Example 10.21. (*Haar Revisited*) The characteristic function of the unit interval $\chi_{[0,1]}$ generates an orthogonal MRA, namely the Haar MRA. The non-zero low-pass filter coefficients are $h_0 = h_1 = 1/\sqrt{2}$; hence the non-zero high-pass coefficients are $g_0 = -1/\sqrt{2}$ and $g_1 = 1/\sqrt{2}$. Therefore the Haar wavelet is $\psi(t) = \varphi(2t-1) - \varphi(2t)$. The refinement masks are

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

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

Example 10.22. (*Shannon Revisited*) This time $\widehat{\varphi}(\xi) = \chi_{[-1/2,1/2)}(\xi)$ generates an orthogonal MRA. It follows from Exercise 10.14 that

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

Hence by Exercise 10.20,

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

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

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

♢

EXAMPLE 10.23. (*Daubechies Wavelets*) For each integer $N \geq 1$ there is an orthogonal MRA that generates a compactly and minimally supported wavelet, such that the length of the support is $2N$, and the filters have $2N$ taps. They are denoted in MATLAB by dbN. The wavelet $db1$ is the Haar wavelet. The coefficients corresponding to $db2$ are

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

$\Diamond$

EXERCISE 10.24. Check that the $db2$ filter is a QMF, and that $h_0+h_1+h_2+h_3 = \sqrt{2}$. $\Diamond$

It turns out that for finite filters H, the conditions (10.15) and $\sum_k h_k = \sqrt{2}$ are sufficient to guarantee the existence of a solution φ to the scaling equation. For infinite filters an extra decay assumption is necessary. However, it is not sufficient to guarantee the orthonormality of the integer shifts of φ. But, if for example $\inf_{|\xi|\leq 1/4} |H(\xi)| > 0$ is also true, then $\{\varphi_{0,k}\}$ is an orthonormal set in $L^2(\mathbb{R})$. For more details, see for example [**Fra**, Chapter 5].

10.5. Cascade algorithm and filter banks

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

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

Then the coarser approximation and detail coefficients $a_{j,k} := \langle f, \varphi_{j,k}\rangle$ and $d_{j,k} := \langle f, \psi_{j,k}\rangle$ for scales $j < J$ can be calculated in order LN operations, where L is the length of the filter, 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 $\widetilde{h} = \{\widetilde{h}_k\}$ is the conjugate flip of the low-pass filter h.

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

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

and consequently,

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

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

Exercise 10.25. Verify that for all $j \in \mathbb{Z}$,

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

$\Diamond$

As a consequence of Exercise 10.25: we obtain for all $j \in \mathbb{Z}$

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

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

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

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

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

$$
\begin{tikzcd}
 & \boxed{*\widetilde{h}} \arrow[r] & \boxed{\downarrow 2} \arrow[r] & a_j \\
a_{j+1} \arrow[ur] \arrow[dr] & & & \\
 & \boxed{*\widetilde{g}} \arrow[r] & \boxed{\downarrow 2} \arrow[r] & d_j
\end{tikzcd}
$$

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 10.26. Compute the Fourier transform for the up-sampling and down-sampling operators in finite-dimensional space. $\Diamond$

EXERCISE 10.27. Verify that given a vector $s \in \mathbb{C}^N$ then $DUs = s$, but UDs is not always s. For which vectors s is $UDs = s$? $\Diamond$

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

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

$$\sum_{m\in\mathbb{Z}} a_{j+1,m}\varphi_{j+1,m} = \sum_{\ell\in\mathbb{Z}} a_{j,\ell}\varphi_{j,\ell} + \sum_{\ell\in\mathbb{Z}} d_{j,\ell}\psi_{j,\ell}. \tag{10.20}$$

We have formulas that express $\varphi_{j,\ell}$ and $\psi_{j,\ell}$ in terms of $\{\varphi_{j+1,m}\}_{m\in\mathbb{Z}}$; see Exercise 10.25. Inserting those formulas in the right hand side (RHS) of the equality (10.20), and collecting all terms that are multiples of $\varphi_{j+1,m}$, we get

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

This calculation, together with equation (10.20), implies that

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

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

$$a_{j+1,m} = h * Ua_j(m) + g * Ud_j(m), \tag{10.21}$$

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

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

$$\begin{array}{lll} a_j \to \boxed{\uparrow 2} \to \boxed{*h} & & \\ & \searrow & \\ & & \oplus \to a_{j+1} \\ & \nearrow & \\ d_j \to \boxed{\uparrow 2} \to \boxed{*g} & & \end{array}$$

EXERCISE 10.28. Verify (10.21). $\Diamond$

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

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

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

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

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

EXAMPLE 10.29. The cascade algorithm can be used to produce very good approximations for both ψ and φ, and this is how pictures of the wavelets and the scaling functions are obtained. For the scaling function φ, it suffices to observe that $a_{1,k} = \langle \varphi, \varphi_{1,k} \rangle = h_k$ and $d_{j,k} = \langle \varphi, \psi_{j,k} \rangle = 0$ for all $j \geq 1$ (the first because of the scaling equation, the second because $\mathbf{V}_0 \subset \mathbf{V}_j \perp \mathbf{W}_j$ for all $j \geq 1$), that is what we need to initialize and iterate as many times as we wish (say n times) the synthesis phase of the filter bank,

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

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

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

A graph can now be plotted (at least for real-valued filters).

This is done similarly for the wavelet ψ. Notice that this time $a_{1,k} = \langle \psi, \varphi_{1,k} \rangle = g_k$ and $d_{j,k} = \langle \psi, \psi_{j,k} \rangle = 0$ for all $j \geq 1$. Now the cascade algorithm produces the approximation coefficients at scale j after $n = j - 1$ iterations:

$$G \rightarrow \boxed{\uparrow 2} \rightarrow \boxed{\ *H\ } \rightarrow \boxed{\uparrow 2} \rightarrow \boxed{\ *H\ } \rightarrow \cdots \rightarrow \boxed{\uparrow 2} \rightarrow \boxed{\ *H\ } \rightarrow \{\langle \psi, \varphi_{j,k} \rangle\}.$$

This time

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

and we can plot a reasonable approximation of ψ from these approximated samples. $\Diamond$

EXERCISE 10.30. Use the algorithm described in Example 10.29 to create a picture of the scaling and the wavelet functions corresponding to the Daubechies wavelets $db2$ after 4 iterations and after 10 iterations. The coefficients for $db2$ are listed in Example 10.23. $\Diamond$

Altogether, we have a *perfect reconstruction filter bank*:

$$\begin{tikzcd}
 & \boxed{*\widetilde{h}} \arrow[r] & \boxed{\downarrow 2} \arrow[r] & a_{j-1} \arrow[r] & \boxed{\uparrow 2} \arrow[r] & \boxed{*h} \arrow[dr] & & \\
a_j \arrow[ur] \arrow[dr] & & & & & & \oplus \arrow[r] & a_j \\
 & \boxed{*\widetilde{g}} \arrow[r] & \boxed{\downarrow 2} \arrow[r] & d_{j-1} \arrow[r] & \boxed{\uparrow 2} \arrow[r] & \boxed{*g} \arrow[ur] & &
\end{tikzcd}$$

The filter bank need not always involve the same filters in the analysis and synthesis phases. When it does, as in our case, we have an *orthogonal filter bank*. One can also obtain perfect reconstruction in more general cases. If we replace $\widetilde{h}$ by a filter $\widetilde{h}^*$, and $\widetilde{g}$ by $\widetilde{g}^*$, then we obtain a *biorthogonal filter bank* with dual filters h, h^*, g, g^*. At this point it might be worth reviewing Section 6.3 on dual basis in the finite dimensional setting.

In the case of a biorthogonal filter bank, there is an associated *biorthogonal MRA* with *dual scaling functions* φ, φ^*. Under certain conditions we can find biorthogonal wavelets ψ, ψ^* that generate a *biorthogonal wavelet basis* $\{\psi_{j,k}\}$ and *dual basis* $\{\psi^*_{j,k}\}$, so that

$$f(x) = \sum_{j,k} \langle f, \psi^*_{j,k}\rangle \psi_{j,k} = \sum_{j,k} \langle f, \psi_{j,k}\rangle \psi^*_{j,k}.$$

The following substitute for Plancherel 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 10.31. 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 $\vec{v}_2$. ◊

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.

10.6. Design features

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

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

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

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

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

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

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

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

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

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

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

10.7. A catalog of wavelets

We list the main wavelets and indicate their properties.

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

Shannon wavelet: This wavelet does not have compact support, however it is C^∞. It is band-limited, but its Fourier transform is discontinuous, hence $\psi(t)$ decays like $1/|t|$ at infinity. 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 ψ has an infinite number of vanishing moments.

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

Battle-Lemarié spline wavelets: Polynomial splines of degree m. The wavelet ψ has $m+1$ vanishing moments. They don't have compact support, but they have exponential decay. Since they are polynomial splines of degree m, they are $m-1$

[1]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.

times continuously differentiable. For m odd, ψ is symmetric around 1/2. For m even, it is antisymmetric around 1/2. The linear spline wavelet is the Franklin wavelet. [*** Give a reference for Franklin? Ph. Franklin, *A set of continuous orthogonal functions*. Math. Annalen 100 (1928) pp.522–529.]

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

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

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

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

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

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

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

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

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

Morlet wavelet: Given by the closed formula

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

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

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

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

10.8. Wavelet packets

To perform the wavelet transform we iterate at the level of the low-pass filter (approximation). In principle it is an arbitrary choice; one could iterate at the high-pass filter level, or any desirable combination. The full dyadic tree gives an overabundance of information, and it corresponds to the *wavelet packets*. Each finite wavelet packet has the information for reconstruction in many different bases, including the wavelet basis. There are fast algorithms to search for the *best basis*. The *Haar packet* includes the Haar basis and the *Walsh basis*. Wickerhauser's book is a good source of information on this topic [**Wic**].

Denote the spaces by $\mathbf{W}_{j,n}$, where j is the scale as before, and n determines the "frequency".

$$
\begin{tikzcd}[row sep=small]
 & & \mathbf{W}_{-2,3} \\
 & \mathbf{W}_{-1,1} \arrow[ur] \arrow[dr] & \\
\mathbf{W}_{0,0} \arrow[ur] \arrow[ddr] & & \mathbf{W}_{-2,2} \\
 & & \mathbf{W}_{-2,1} \\
 & \mathbf{W}_{-1,0} \arrow[ur] \arrow[dr] & \\
 & & \mathbf{W}_{-2,0}
\end{tikzcd}
$$

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

$$
\begin{tikzcd}[row sep=small]
 & \mathbf{W}_{-1,1} & \\
\mathbf{W}_{0,0} \arrow[ur] \arrow[ddr] & & \\
 & & \mathbf{W}_{-2,1} \\
 & \mathbf{W}_{-1,0} \arrow[ur] \arrow[dr] & \\
 & & \mathbf{W}_{-2,0}
\end{tikzcd}
$$

2. Connected binary subtree.
Leaves determine Haar basis.

Notice that $\mathbf{W}_{0,0} = \mathbf{V}_0$, more generally, $\mathbf{W}_{j,0} = \mathbf{V}_j$, and $\mathbf{W}_{j,1} = \mathbf{W}_j$. We also know that the spaces $\mathbf{W}_{j-1,2n}$ and $\mathbf{W}_{j-1,2n+1}$ are orthogonal and their direct sum is $\mathbf{W}_{j,n}$. Therefore the leaves of every connected binary subtree[2] of the wavelet packet tree correspond to an orthogonal basis of the initial space.

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

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

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

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

$$
\mathbf{W}_{j,n} = \operatorname{span}\{\omega_{j,k,n} : k \in \mathbb{Z}\} = \Big\{ f = \sum_k a_{j,k,n} w_{j,k,n} : \sum_k |a_{j,k,n}|^2 < \infty \Big\}.
$$

EXAMPLE 10.32. (*The Walsh Functions*) For the Haar function, the corresponding wavelet packet is described by the equations

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

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

[2]Starting from the "top" node ($\mathbf{W}_{0,0}$), we allow the node to have either two offspring or none. The same rule applies to the offspring, if there are any. The nodes that have no offspring are the *leaves*.

EXERCISE 10.33. Identify all possible bases for the Haar packet binary tree with three levels. How many bases can you find? Draw the corresponding phase plane diagrams. In particular, draw the Walsh functions. ◊

EXERCISE 10.34. Use MATLAB to plot the n^{th} Walsh function $\omega_n(t)$ and the functions $\sin nt$ and $\cos nt$ for $-2\pi \le t \le 2\pi$ 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. ◊

EXERCISE 10.35. What are the Walsh functions in the finite-dimensional case? ◊

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

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

The wavelet packet and *cosine packet* libraries create large libraries of orthogonal bases, all of which have fast algorithms. The Fourier and wavelet bases are particular examples in this time–frequency library; so are Gabor-like bases.

10.9. Two-dimensional wavelets

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

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

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

EXERCISE 10.36. Show that if $\{\psi_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for a closed subspace $\mathbf{V} \subset L^2(\mathbb{R})$, then the family of functions defined on $\mathbb{R}^2$ by

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

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

EXERCISE 10.37. What would the trigonometric basis in $L^2([0,1]^2)$ be? How about the finite-dimensional trigonometric basis, say in two dimensions? ◊

We would like to use this idea but at the level of the approximation spaces $\mathbf{V}_j$ in the MRA. For each scale j, the family $\{\varphi_{j,k}\}_k$ is an orthonormal basis of $\mathbf{V}_j$.

Consider the tensor products $\varphi_{j,k,n}(x,y) = \varphi_{j,k}(x)\varphi_{j,n}(y)$ of these functions. Let $\mathcal{V}_j$ be the closure in $L^2(\mathbb{R}^2)$ of the linear span of those functions, in other words

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

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

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

Therefore the integer shifts $\{\varphi(x-k, y-n) = \varphi_{0,k,n}\}_{k,n\in\mathbb{Z}}$ form an orthonormal basis of $\mathcal{V}_0$, consecutive approximation spaces are connected via scaling by 2 on both variables, and the other conditions are clear.

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

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

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

$$\mathcal{W}_j = (\mathbf{W}_j \otimes \mathbf{W}_j) \oplus (\mathbf{W}_j \otimes \mathbf{V}_j) \oplus (\mathbf{V}_j \otimes \mathbf{W}_j).$$

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

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

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

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

EXERCISE 10.38. 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 10.39. (*The Two-Dimensional Haar Basis*) The scaling function is the characteristic function

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

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

$$\underset{\varphi(x,y)}{\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 1 & 1 \\ \hline\end{array}} \qquad \underset{\psi^d(x,y)}{\begin{array}{|c|c|}\hline -1 & 1 \\ \hline 1 & -1 \\ \hline\end{array}} \qquad \underset{\psi^h(x,y)}{\begin{array}{|c|c|}\hline 1 & -1 \\ \hline 1 & -1 \\ \hline\end{array}} \qquad \underset{\psi^v(x,y)}{\begin{array}{|c|c|}\hline -1 & -1 \\ \hline 1 & 1 \\ \hline\end{array}}$$

$\Diamond$

Exercise 10.40. 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*. [*** Cool. Include a reference. Point to the project too.]

10.10. Basics of compression and denoising

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

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

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

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

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

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

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

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

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

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

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

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

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