Short Wavelets and Matrix Dilation Equations

Gilbert Strang and Vasily Strela
Department of Mathematics
Massachusetts Institute of Technology
Cambridge MA 02139
gs@math.mit.edu

Abstract

Scaling functions and orthogonal wavelets are created from the coefficients of a lowpass and highpass filter (in a two-band orthogonal filter bank). For "multifilters" those coefficients are *matrices*. This gives a new block structure for the filter bank, and leads to multiple scaling functions and wavelets. Geronimo, Hardin, and Massopust constructed two scaling functions that have extra properties not previously achieved. The functions Φ_1 and Φ_2 are *symmetric* (linear phase) and they have *short support* (two intervals or less), while their translates form an orthogonal family. For any single function Φ, apart from Haar's piecewise constants, those extra properties are known to be impossible. The novelty is to introduce 2 by 2 matrix coefficients while retaining orthogonality.

This note derives the properties of Φ_1 and Φ_2 from the matrix dilation equation that they satisfy. Then our main step is to construct associated wavelets: *two wavelets for two scaling functions.* The properties were derived in [1] from the iterated interpolation that led to Φ_1 and Φ_2. One pair of wavelets was found earlier by direct solution of the orthogonality conditions (using Mathematica). Our construction is in parallel with recent progress by Hardin and Geronimo, to develop the underlying algebra from the matrix coefficients in the dilation equation — in another language, to build the 4 by 4 paraunitary polyphase matrix in the filter bank. The short support opens new possibilities for applications of filters and wavelets near boundaries.

EDICS SP 2.4.4

Acknowledgements

We thank Jeff Geronimo and Doug Hardin for sharing their important ideas, and Chris Heil for help in displaying the wavelets. The authors are grateful for the support of the National Science Foundation and of INTEVEP.

Introduction

This note begins with two functions. They were constructed by Geronimo, Hardin, and Massopust and they are not wavelets. They are *scaling functions* Φ_1 and Φ_2 from which wavelets are created. Our purpose is to carry out that final step of the construction, and produce wavelets with new properties.

We believe that these properties will be useful in applications. They cannot be achieved by any ordinary wavelet W. Two wavelets are needed, at least. This is the novelty, and the source of new possibilities. A "multifilter" may also be a new idea in filter design, and we start by explaining this connection.

Normally, a filter bank has a single lowpass filter. The incoming signal, a vector $\mathbf{x}$ in discrete time, is filtered and downsampled (decimated) by 2. The output is $y_j = \sum c_k x_{2j-k}$. The other filter will execute parallel steps with coefficients d_k. This is the two-band analysis bank. A perfect reconstruction synthesis bank recovers $\mathbf{x}$ from the two downsampled outputs.

To reach scaling functions and wavelets, we iterate the lowpass filter. The continuous limit of $y_j = \sum c_k x_{2j-k}$ is the *dilation equation* for the fixed point — the scaling function $\Phi(t)$:

$$\Phi(t) = \sum c_k \Phi(2t-k).$$

Convergence to this limit is encouraged (but not guaranteed) by a zero of the frequency response $\sum c_k z^{-k}$ at $z=-1$ (or $\xi = \pi$). A zero of order p also means that every polynomial of lower degree is a combination of the functions $\Phi(t-k)$. Then the wavelet

$$W(t) = \sum d_k \Phi(2t-k)$$

has p vanishing moments [3–4]. The wavelets $W(2^j t - k)$ are mutually orthogonal if the original c_k and d_k came from a nondegenerate orthogonal filter bank. Thus wavelet theory can be described as the iterated limit of filter bank theory.

Now we introduce a "*multifilter*". It has two or more lowpass filters. The purpose of this multiplicity is to achieve the properties that are described below — linear phase with a short, orthogonal filter bank. To describe the multifilter, the coefficients c_k and d_k become matrices. In our case they are 2 by 2, so that four signals come out from the analysis bank in Figure 1. Each has been downsampled only by 2. The extra work is acceptable when the multifilters are short, and the compensation is in the properties of Φ and W.

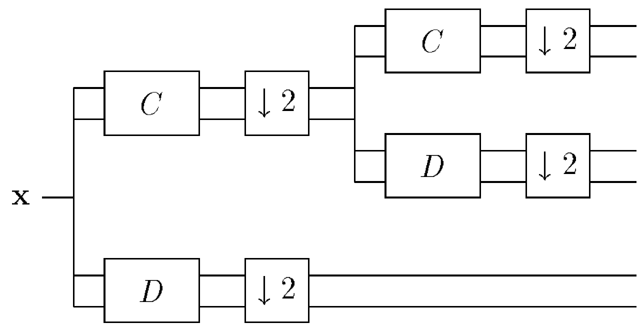

Figure 1. A bank of multifilters with the lowpass filter iterated once.

Normally the d's are determined in an automatic way from the c's. The wavelet is constructed by reversing the coefficients c_k, shifting by one, and alternating signs: $W(t) = \sum(-1)^k c_{1-k}\Phi(2t-k)$. This procedure succeeds for scalar coefficients but here it fails. The orthogonality of Φ to W no longer follows from identities like $c_0c_1 - c_1c_0 = 0$, because the c_k are *matrices* – and they do not commute. A new construction is needed for two wavelets. It remains absolutely true that the invention of Φ_1 and Φ_2 is the essential step — the Φ's lead to everything. The wavelet coefficients d_k follow from linear and quadratic equations, whose solution was first computed by Mathematica. This note and the forthcoming paper [2] give two approaches to simple explicit formulas for the matrix coefficients that produce the wavelets.

Underlying the whole construction is a *paraunitary matrix*. This is a matrix polynomial $H(z)$, 4 by 4 in our problem, which is unitary on the circle $|z| = 1$. The scaling functions come from half of the matrix (containing the c's) and the wavelets come from the other half.

Properties of Φ_1 and Φ_2

We begin with the key properties of the scaling functions in Figure 2. First of all, there are two functions! By some measure this doubles the work (except these functions have so few coefficients c_k). We believe that the new properties are worth the price. These properties are stated as 1—4 and then proved.

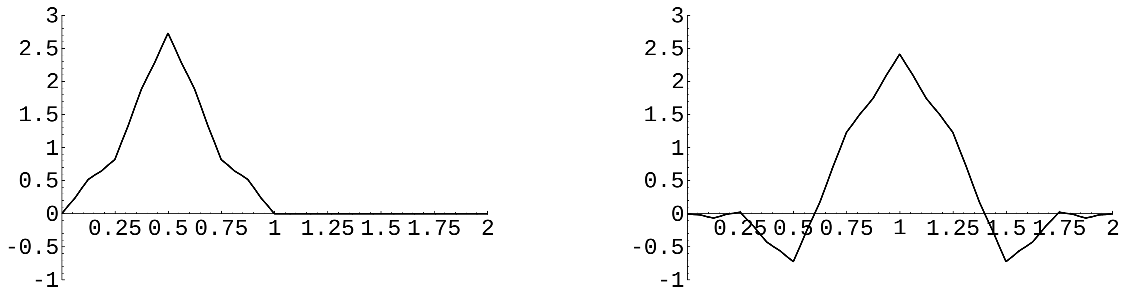

Figure 2. Scaling functions Φ_1 and Φ_2.

1. *Symmetry* Φ_1 and Φ_2 are even functions (after a shift of the origin). In the language of filters, they have linear phase. A full chapter of Daubechies [3] is devoted to this left-right symmetry — not achieved by a self-orthogonal wavelet. "It is a property of our visual system that we are more tolerant to symmetric errors than asymmetric ones."

2. *Short support* Φ_i vanishes outside the interval $[0, i]$. In problems with boundaries this is extremely valuable. A boundary condition $f(0) = 0$ can be satisfied directly — the coefficient of $\Phi_2(t-1)$ is zero in the approximation to f. With longer support we must modify the functions near boundaries.

For differential equations, finite elements are a success because they are so local. Splines have longer support. Up to now, wavelets have had the longest support of all (to achieve orthogonality). For periodic equations this presents no problem. But nonperiodic equations, and boundaries between images, are inescapable.

3. *Second-order accuracy* All wavelets can reproduce constant functions. This first-order accuracy is not good enough in practice. The step to better approximation requires that we also get linear functions right. In splines and finite elements this comes from the *hat function*, which has $H(1) = 1$ and all other $H(n) = 0$. (It is linear on intervals between integers.) Every function $at+b$ is a combination of translates $H(t-k)$, so the approximation error starts with the quadratic term and the accuracy is second-order.

The scaling functions have this property because they produce the hat function exactly:

$$\frac{1}{\sqrt{2}}\Phi_1(t) + \frac{1}{\sqrt{2}}\Phi_1(t-1) + \Phi_2(t) = H(t). \tag{1}$$

This will establish two vanishing moments for the new wavelets. They will be perpendicular to Φ_1 and Φ_2 and H — and therefore to the functions 1 and t.

4. *Orthogonality* The translates $\Phi_1(t-k)$ and $\Phi_2(t-k)$ are all mutually orthogonal. It is this property which the hat function lacks and the new functions offer. Notice that the product $H(t)H(t-1)$ involves $(\Phi_1(t-1))^2$ from equation (1). Its integral is not zero and $H(t)$ is not orthogonal to its translates. A direct orthogonalization using Gram-Schmidt, or in frequency using $\hat{H}(\xi)$, would destroy the short support of the hat function. By enlarging the space to include Φ_1 as well as H we have a shift-invariant space with a local orthogonal basis.

Recall from Daubechies [3] that no single function $\Phi(t)$ can possess these properties 1—4. For the sake of 3 and 4 we give up 1 and 2. (Symmetry is possible by changing to biorthogonal wavelets.) The new possibility is to find W_1 and W_2 to go with Φ_1 and Φ_2, so that the translates of *all four* are mutually orthogonal. It seems miraculous that there is also orthogonality to all dilations $W_i(2^j t - k)$. This family $W_{jk}(t)$ is an orthogonal basis for $L^2(\mathbf{R})$, exactly as in the usual case of a single wavelet.

We will connect orthogonality to a unitary block Toeplitz matrix. It is a band matrix so the Φ's have compact support. The blocks include the coefficients from the dilation equation — which is fundamental.

5. *Matrix dilation equation* The coefficients c_k are 2 by 2 matrices, multiplying vectors of scaling functions:

$$\begin{bmatrix} \Phi_1(t) \\ \Phi_2(t) \end{bmatrix} = \sum_0^3 c_k \begin{bmatrix} \Phi_1(2t-k) \\ \Phi_2(2t-k) \end{bmatrix}. \tag{2}$$

In the original "fractal function" construction [1], this dilation equation emerged at the end. It is our starting point in the verification of 1 — 4. The four matrices c_k are displayed in the top rows of equation (19) below, with d's in the bottom rows. The c's multiply powers of z in the polynomial $P(z) = \frac{1}{2}\sum c_k z^k$:

$$P(z) = \frac{1}{20}\begin{bmatrix} 6+6z & 8\sqrt{2} \\ (-1+9z+9z^2-z^3)/\sqrt{2} & -3+10z-3z^2 \end{bmatrix}. \tag{3}$$

Buried in this matrix polynomial are the good properties of the Φ's — the solutions to the dilation equation (2).

The Dilation Equation Implies 1 — 3

To speed up the verification of the first three properties we introduce the sum

$$S(t) = \frac{1}{\sqrt{2}}\Phi_1(t) + \frac{1}{\sqrt{2}}\Phi_1(t-1).$$

The dilation equation for Φ_1 can be read from 6, 6 and $8\sqrt{2}$ in the first row of $20P(z)$:

$$\Phi_1(t) = \frac{1}{10}[6\Phi_1(2t) \; + \; 6\Phi_1(2t-1) \; + \; 8\sqrt{2}\Phi_2(2t)].$$

To find the dilation equation for S, replace t by $t-1$ and add the new equation to this one. Division by $\sqrt{2}$ produces

$$S(t) = \frac{1}{10}[6S(2t) + 6S(2t-2) + 8\Phi_2(2t) + 8\Phi_2(2t-2)]. \tag{4}$$

The dilation equation for Φ_2 comes from the second row of $P(z)$. Removing Φ_1 in favor of S, the equation is

$$\Phi_2(t) = \frac{1}{10}[-S(2t) + 10S(2t-1) - S(2t-2) - 3\Phi_2(2t) + 10\Phi_2(2t-1) - 3\Phi_2(2t-2)]. \tag{5}$$

Before verifying symmetry and short support we point to the matrix polynomial that goes with the new dilation equations (4) and (5). The original $P(z)$ is replaced by

$$Q(z) = \frac{1}{20}\begin{bmatrix} 6+6z^2 & 8+8z^2 \\ -1+10z-z^2 & -3+10z-3z^2 \end{bmatrix}.$$

Note in passing that Q is related to P by a *two-scale similarity transformation*. An ordinary change of basis $u = Mv$ yields a matrix $M^{-1}PM$ that is similar to P. In our case the change from Φ_1 to S divides the first column of P by $(1+z)/\sqrt{2}$. A similarity would multiply row 1 by the same quantity. A two-scale similarity multiplies instead by $(1+z^2)/\sqrt{2}$. The result is $M^{-1}(z^2)P(z)M(z)$. This is the $Q(z)$ that we found directly.

The first three properties are easy from the new dilation equations:

1. *Symmetry* Replace t by $2-t$ in (4) and (5). The equations for the new functions $S(2-t)$ and $\Phi_2(2-t)$ are again (4) and (5). The solutions (known to be unique up to a scalar multiple) must be the same: $S(t) = S(2-t)$ and $\Phi_2(t) = \Phi_2(2-t)$. This is symmetry around $t = 1$.

From the definition of S, it follows that Φ_1 is symmetric around $t = \frac{1}{2}$.

2. *Short support* The dilation equations (4) and (5) have only three coefficients C_0, C_1, C_2. The right sides involve $2t-2$ but not $2t-3$. By a standard argument in [3], the solutions S and Φ_2 are supported on the interval $[0,2]$. It follows that Φ_1 is supported on $[0,1]$.

3. *Second-order accuracy* The hat function is a sum of three narrower hats:

$$H(t) = \frac{1}{2}H(2t) + H(2t-1) + \frac{1}{2}H(2t-2).$$

Add (4) and (5) to find exactly the same equation for $S + \Phi_2$. By uniqueness H must be identical with $S + \Phi_2$, because there is agreement at the point $t = 1$.

We emphasize that the approximation accuracy is decided by these scaling functions — not the wavelets. Φ_1 and Φ_2 and their translates span a "low frequency space" V_0. This contains the functions 1 and t — to which the wavelets are orthogonal. In the filter matrix, the dilation coefficients c_k give the lowpass filter. The wavelet coefficients d_k — still to be discovered, because $(-1)^k c_{1-k}$ will not work — enter the bandpass filter. The translates of the wavelets span the higher frequency space W_0.

The wavelet spaces W_0, $W_1, \ldots, W_j$ do play a part in approximation. They reduce the mesh size to $\frac{1}{2}$ and $\frac{1}{4}$ and eventually 2^{-j}. The dilation from t to $2t$ rescales V_0 to V_1 and W_0 to W_1. Frequencies are doubled. We are approximating by a more refined space, with more detail. The great contribution of Mallat's *multiresolution analysis* [3, 4] is that V_j equals $V_0 + W_0 + \ldots + W_{j-1}$. So the accuracy moves to the scale 2^{-j} by including all those wavelets. Also the basis becomes "hierarchical", and the pyramid algorithm yields a fast wavelet transform.

The finite element method [5, p. 56] provides earlier examples of two trial functions per element — corresponding to two scaling functions per interval. The purpose is the same, to reduce the support and maintain the accuracy. For finite elements, short support is crucial. The most familiar example interpolates f and f' at each meshpoint by a piecewise cubic. The basis functions are the "Hermite cubics" in Figure 3 and their translates. They are C^1 where the B-spline is C^2 — but its support is $[0,4]$. The cubic spline interpolates one number at each meshpoint, namely f itself — as in critical sampling.

This polynomial example has no orthogonality! But the cubics still satisfy a dilation equation. Goodman and Lee [7] studied *biorthogonal* wavelets, which have quite long support — as is usual when the Φ's are piecewise polynomials. It requires a more fractal construction to achieve *self*-orthogonality with short support.

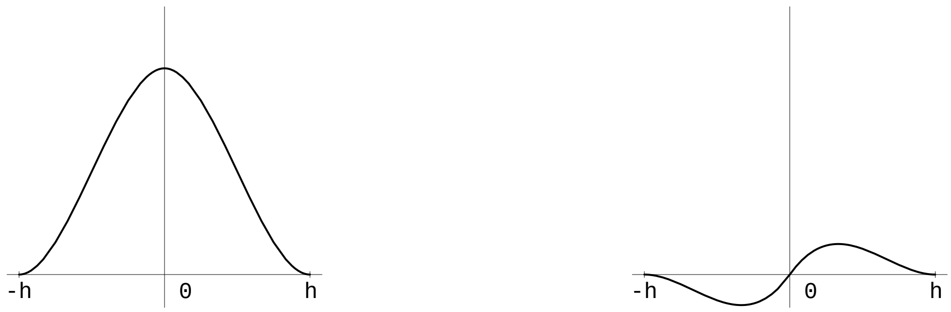

Figure 3. Two finite elements that combine into a cubic spline.

This finite element example involves cubics, where our example begins with the hat function — we are also requiring orthogonality. Before establishing that final property, we note a remarkable inhomogeneous dilation equation for the functions in V_0. It is the fixed point equation (2.2) of Geronimo, Hardin, and Massopust, restricted to one interval:

$$f(t) = \begin{cases} sf(2t) \ + \ \text{linear function}, & 0 \le t \le \frac{1}{2} \\ sf(2t-1) \ + \ \text{linear function}, & \frac{1}{2} \le t \le 1. \end{cases} \qquad (6)$$

This allows direct calculation in [1] and [2] of all the inner product integrals needed in application. Solving (6) by iteration displays f as a fractal function, created from smaller images of itself. Changing the inhomogeneous term (the linear function) would give new possibilities.

Orthogonality: Scaling Functions

In the time domain a filter is represented by an infinite matrix with constant diagonals (a Toeplitz matrix). A lowpass filter has c_k on the kth diagonal. The matrix multiplication is exactly a convolution $\mathbf{c} * \mathbf{x}$. A regular lowpass filter exactly passes the constant functions, $\mathbf{c} * \mathbf{1} = \mathbf{1}$. Exactness for higher-order polynomials means a flatter response at zero frequency, and more accuracy in approximation by V_0. In the scalar case it also means zeros of the response $P(z)$ at $z = e^{i\pi}$.

We mention that the word "regular" is overworked. (Regularity may mean smoothness of the wavelet or it may mean vanishing moments — the two definitions are related but quite different.) Our forthcoming paper [6] extends the vanishing moment condition to multiwavelets — instead of zeros in $P(z)$ it becomes becomes eigenvalues $1, \frac{1}{2}, \ldots$ of the matrix L in equation (7).

When the outputs are decimated, rows of the convolution matrix are thrown away. For a single filter this downsampling leads to aliasing error — different frequencies give the same response. But a bank of filters, each downsampled, can cancel all error and give perfect reconstruction. The convolution matrix — which is now *block Toeplitz* — can be an orthogonal matrix: inverse = transpose. This is our goal.

The lowpass part L — convolution followed by decimation — will be the top half of a doubly infinite orthogonal matrix:

$$L = \frac{1}{\sqrt{2}} \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & & & & & \\ & & c_0 & c_1 & c_2 & c_3 & & & \\ & & & & c_0 & c_1 & c_2 & c_3 & \\ & & & & & & \cdot & \cdot & \cdot & \cdot \end{bmatrix}. \tag{7}$$

Notice the shift by two: half the rows are deleted by downsampling. We do not require all the original rows to be orthogonal; that is too restrictive. There are no interesting FIR allpass filters (= orthogonal Toeplitz band matrices). We need a *bank* of filters, each downsampled. By adjoining a bandpass matrix B we obtain the allpass *block* Toeplitz matrix T below.

Orthogonality of the rows of L means that $LL^T = I$. Explicitly, L times L^T gives ones and zeros , on and off the main diagonal:

$$c_0c_0^T + c_1c_1^T + c_2c_2^T + c_3c_3^T = 2I \tag{8}$$

$$c_0c_2^T + c_1c_3^T = 0 \tag{9}$$

We have written the coefficients c_k as matrices — in our case they are. (For scalars this time-domain description of the filter as a convolution matrix is developed in [9]. The sequence c_k comes there in reverse order — this change of convention has no effect on the computations.) Equations (8) and (9) convert in the scalar case to the familiar condition on the polynomial $P(z)$ — which comes by Fourier transform or z-transform:

$$|\frac{1}{2}\sum c_k e^{ik\xi}|^2 + |\frac{1}{2}\sum c_k e^{ik(\xi+\pi)}|^2 \equiv 1. \tag{10}$$

This is the power complementary condition [8]; it is condition O for wavelets [4]; it gives a quadrature mirror filter (QMF bank); it yields orthogonality. For matrix coefficients c_k the only change is to

$$P(\xi)P^*(\xi) + P(\xi+\pi)P^*(\xi+\pi) \equiv I. \tag{11}$$

Important: The filter matrix acts in discrete time (on vectors). The dilation equation is in continuous time (for functions). The link between filters and wavelets is *iteration with* *rescaling.* In other words, a cascade of filters yields the Φ_i (and the rescaling introduces factors of 2). We may start with the box function $B_0(t)$ — equal to one on $[0,1]$ and elsewhere zero — and iterate:

$$B_{n+1}(t) = \sum c_k B_n(2t-k).$$

The limit as $n \to \infty$ (if it exists!) solves the dilation equation. Rioul and Vetterli show how this cascade can fail; many important filters do not lead to a continuous limit. For filters of Daubechies type (flat at $\xi = 0$, vanishing moments at $\xi = \pi$), we do reach a scaling function. In the frequency domain this cascade is an infinite product of $P(\xi/2^j)$. Please see [9] for a beautiful exposition and [3] for sufficient conditions for convergence. Necessary and sufficient conditions, directly verifiable from the c's, may not be possible.

Orthogonality of the translates $\Phi_i(t-k)$ follows closely from $LL^T = I$. There is a technical "Cohen-Lawton condition", elaborated in [3, p. 182] and not repeated here, which the particular coefficients c_k do satisfy. By displaying a (block) row of L and two columns of L^T, we hope to make it possible for the reader to verify orthogonality. The matrices $c_0,\ c_1,\ c_2,\ c_3$ along the row are taken from $P(z)$ in equation (3):

$$\frac{1}{10}\begin{bmatrix} 6 & 8\sqrt{2} & 6 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{\sqrt{2}} & -3 & \frac{9}{\sqrt{2}} & 10 & \frac{9}{\sqrt{2}} & -3 & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \frac{1}{10}\begin{bmatrix} 6 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 8\sqrt{2} & -3 & 0 & 0 \\ 6 & \frac{9}{\sqrt{2}} & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & \frac{9}{\sqrt{2}} & 6 & -\frac{1}{\sqrt{2}} \\ 0 & -3 & 8\sqrt{2} & -3 \\ 0 & -\frac{1}{\sqrt{2}} & 6 & \frac{9}{\sqrt{2}} \\ 0 & 0 & 0 & 10 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}! \tag{12}$$

Orthogonal Wavelets: General Method

We turn now to our main goal — the construction of wavelets W_1 and W_2. We look for the coefficients d_k in the bandpass (or highpass) filter. Exactly as in the Daubechies case, the two filters will make an orthogonal system with perfect reconstruction. The new point is that we have multifilters with matrix coefficients.

Iterating the lowpass filter L yields Φ_1 and Φ_2. They give half of a "multiresolution analysis" — the continuous analog of Figure 1. The other half comes from the wavelets W_1 and W_2. The translates $\Phi_i(t-k)$ span a space V_0; the translates $W_i(t-k)$ span a space W_0. Those spaces are orthogonal, because those bases are orthogonal. Just as in Figure 1, where the second filter bank orthogonally splits the output from the first lowpass filter, so $V_0 \oplus W_0 = V_1 =$ all combinations of $\Phi_i(2t-k)$.

This means in particular that the $\Phi_i(t)$ are combinations of $\Phi_i(2t-k)$. That is the dilation equation. It also means that the $W_i(t)$ are combinations of $\Phi_i(2t-k)$. This is the *wavelet equation*:

$$\begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix} = \sum d_k \begin{bmatrix} \Phi_1(2t-k) \\ \Phi_2(2t-k) \end{bmatrix}. \tag{13}$$

The filter bank with c's and d's has perfect reconstruction. The infinite matrix that contains both filters (with decimation to remove every second row) is to be an orthogonal matrix:

$$\begin{bmatrix} \\ L \\ \\ B \\ \\ \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & & \\ & & c_0 & c_1 & c_2 & c_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ d_0 & d_1 & d_2 & d_3 & & \\ & & d_0 & d_1 & d_2 & d_3 \end{bmatrix}. \tag{14}$$

We know already that $LL^T = I$. The new coefficients d_k must lead to $LB^T = 0$ and $BB^T = I$. Then the rows of our matrix are orthonormal.

Remember how this is achieved in the scalar case, for wavelets and QMF filters. The c's are flipped to give $d_0 = c_3$, $d_1 = -c_2$, $d_2 = c_1$, and $d_3 = -c_0$. This automatically makes the L and B blocks orthogonal — the rows of c's in (14) are orthogonal to the rows of d's. For example $d_2c_0 + d_3c_1$ equals $c_1c_0 - c_0c_1$. For scalars this is zero. For our matrix coefficients it is not zero, and the flip construction fails.

We pose the same problem in the frequency domain. A direct approach is to interlace the rows of L and B (which has no effect on orthogonality). The result is a *block Toeplitz matrix* T:

$$T = \frac{1}{\sqrt{2}} \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & & \\ d_0 & d_1 & d_2 & d_3 & & \\ & & c_0 & c_1 & c_2 & c_3 \\ & & d_0 & d_1 & d_2 & d_3 \\ & & & & \cdot & \cdot \end{bmatrix}.$$

These blocks automatically have "polyphase" form, with even indices in a different column from odd indices:

$$h(0) = \begin{bmatrix} c_0 & c_1 \\ d_0 & d_1 \end{bmatrix} \quad \text{and} \quad h(1) = \begin{bmatrix} c_2 & c_3 \\ d_2 & d_3 \end{bmatrix}. \tag{15}$$

The blocks are 4×4 because each coefficient is 2×2. By direct multiplication, the Toeplitz matrix T is orthogonal when

$$h(0)h(0)^T + h(1)h(1)^T = 2I \text{ and } h(0)h(1)^T = 0. \tag{16}$$

In the z domain, the polyphase (even-odd) matrix $H(z)$ must be *paraunitary*:

$$H(z) = h(0) + z^{-1}h(1) \text{ must satisfy } H(z)\tilde{H}(z) = 2I. \tag{17}$$

Here we adopt the notation of Vaidyanathan: $\tilde{H}(z)$ is the conjugate transpose of $H(1/\overline{z})$. On the unit circle, where $z = 1/\overline{z}$, the polyphase matrix $H(z)/\sqrt{2}$ is to be unitary. $H(z)$ is a polynomial in z^{-1}, because the underlying FIR filter is causal — T is a block triangular band matrix. The aliasing error, which is unavoidably created by decimation in the first filter L, is cancelled by the second filter B. Perfect reconstruction will follow from (16) or (17), and the d_k yield orthogonal wavelets. It is the z-transform that leads from (16) to (17); but one can verify their equivalence directly.

Notice in passing how the flip construction appears when H is only 2×2. We are given its top half (the scalars c_k) and we need its bottom half:

$$H(z) = \begin{bmatrix} c_0 + c_2 z^{-1} & c_1 + c_3 z^{-1} \\ — & — \end{bmatrix}.$$

When the first row of a unitary matrix is $[a \;\; b]$, the second row is $[\overline{b} \;\; -\overline{a}]$ times a factor with $|z| = 1$. Thus $H(z)$ is completed on the unit circle to be unitary:

$$\begin{bmatrix} c_0 + c_2 z^{-1} & c_1 + c_3 z^{-1} \\ c_1 + c_3 z & -c_0 - c_2 z \end{bmatrix} \text{ or better } \begin{bmatrix} c_0 + c_2 z^{-1} & c_1 + c_3 z^{-1} \\ c_3 + c_1 z^{-1} & -c_2 - c_0 z^{-1} \end{bmatrix}.$$

That last row shows the alternating flip to c_3, $-c_2$, c_1, $-c_0$.

We now propose a general method to complete the lower half of $H(z)$. It is inspired by Vaidyanathan's important book [8]. He factors $L(z)$, the top half of the matrix $H(z)$, into $L(1)U(z)$. The paraunitary matrix $U(z)$ is square — in our case 4×4. The constant 2×4 matrix $L(1)$ can be completed by $B(1)$ to a constant unitary matrix (times $\sqrt{2}$). Then we read off the desired lower half $B(z)$ from the product $B(1)U(z)$:

$$H(z) = \begin{bmatrix} L(z) \\ B(z) \end{bmatrix} = \begin{bmatrix} L(1) \\ B(1) \end{bmatrix} \begin{bmatrix} & U(z) & \end{bmatrix}. \tag{18}$$

Vaidyanathan gives a specific form to $U(z)$. It is a product of Householder-type factors $I - vv^T + z^{-1}vv^T$. They can be peeled off one at a time from $L(z)$, each with $||v|| = 1$ and determinant z^{-1}. After finitely many steps, making up $U(z)$, we reach the constant matrix $L(1)$.

In our particular example the 2×4 $L(z)$ can be completed to the 4×4 $H(z)$ in a single operation — creating $U(z)$ instead of its factors. This produces the wavelet coefficients d_k in $B(z)$. There will be two families of possible $U(z)$ and therefore two families of wavelets.

Orthogonal Wavelets: Particular Construction

We now offer one specific set of wavelet coefficients d_k. They are the four 2×2 matrices in the last rows of the following matrix. The scaling coefficients $c_0,\ c_1,\ c_2,\ c_3$ are still in the first two rows:

$$\frac{1}{10}\begin{bmatrix} 6 & 8\sqrt{2} & 6 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{\sqrt{2}} & -3 & \frac{9}{\sqrt{2}} & 10 & \frac{9}{\sqrt{2}} & -3 & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & -3 & \frac{9}{\sqrt{2}} & -10 & \frac{9}{\sqrt{2}} & -3 & -\frac{1}{\sqrt{2}} & 0 \\ 1 & 3\sqrt{2} & -9 & 0 & 9 & -3\sqrt{2} & -1 & 0 \end{bmatrix} \tag{19}$$

It is easily checked that the four rows are orthogonal (all have length $\sqrt{2}$). Orthogonality must also be verified after the rows are shifted. (Equation (12) above already checked the first two rows.) Imagine that each row is shifted by four places, to produce an 8×12 matrix. Those eight rows are orthogonal!

This is the test that $H(z) = h(0) + z^{-1}h(1)$ is a paraunitary matrix. The filter bank gives perfect reconstruction. Equations (8) and (9) hold for the d_k as well as the c_k, and the two filters give $LB^T = 0$:

$$c_0d_0^T + c_1d_1^T + c_2d_2^T + c_3d_3^T = 0 \quad \text{and} \quad c_0d_2^T + c_1d_3^T = 0. \tag{20}$$

The first wavelet comes from row 3 of our matrix. The eight entries multiply $\Phi_1(2t)$, $\Phi_2(2t)$, $\Phi_1(2t-1)$, $\Phi_2(2t-1), \ldots$ to give $W_1(t)$. Notice that those eight entries agree with row 2, which produces $\Phi_2(t)$, except for the coefficient of $\Phi_2(2t-1)$. We conclude that

$$W_1(t) = \Phi_2(t) - 2\Phi_2(2t-1). \tag{21}$$

This is symmetric around $t = 1$ with support $[0, 2]$.

The second wavelet comes from row 4 of our matrix. The signs are now alternating, to give *anti*symmetry of $W_2(t)$. The combination $\Phi_1(2t) - 9\Phi_1(2t-1) + 9\Phi_1(2t-2) - \Phi_1(2t-3)$ is odd around $t = 1$. So is the other part $3\sqrt{2}\Phi_2(2t) - 3\sqrt{2}\Phi_2(2t-2)$. The sum W_2 has support $[0, 2]$. This is an alternative form of linear phase — odd and even filters in the same filter bank. The wavelets in Figure 4 will be one particular choice in the first of the two families below. They are orthogonal to all translates and dilates $W_i(2^jt - k)$.

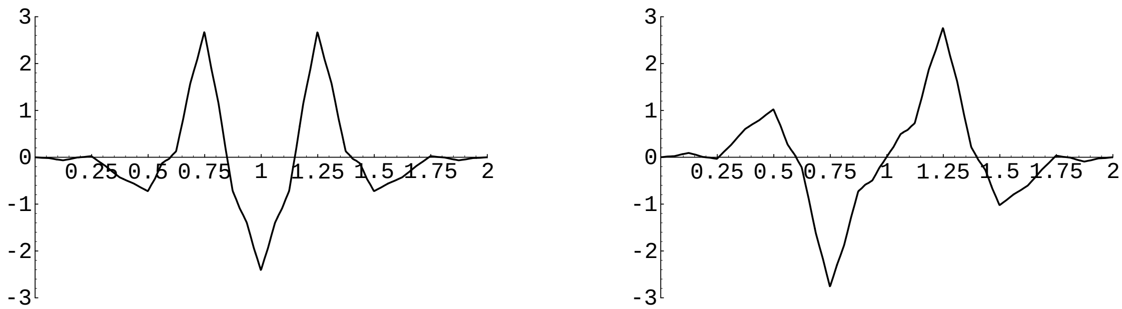

Figure 4. Symmetric-antisymmetric pair of wavelets.

Orthogonal Wavelets: Two Families

The procedure described earlier was based on degree-one paraunitary factors $I - vv^T + z^{-1}vv^T$. We propose to combine two or three factors into a single matrix:

$$U(z) = I - P + z^{-1}P.$$

Here P is a symmetric projection matrix: $P^2 = P$. (The rank-one case has $P^2 = vv^Tvv^T = vv^T$ from the requirement $v^Tv = 1$.) We choose P so that

$$L(z) = L(1)U(z). \tag{22}$$

Recall that $L(z) = h(0) + z^{-1}h(1)$ is the polyphase matrix for the low-pass filter. Thus

$$L(z) = \begin{bmatrix} c_0 + c_2z^{-1} & c_1 + c_3z^{-1} \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 6 & 8\sqrt{2} & 6 & 0 \\ \frac{-1+9z^{-1}}{\sqrt{2}} & -3 - 3z^{-1} & \frac{9-z^{-1}}{\sqrt{2}} & 10 \end{bmatrix}. \tag{23}$$

In one family the matrix P will have rank one. The pair of wavelets computed by Donovan, Hardin, Geronimo, and Massopust lies in this family. So does the symmetric-antisymmetric pair described above — which those authors also found. In the second family P has rank 2 and the filter bank has one extra delay. The option of two families exists because the 2×4 matrix $h(1) = [c_2 \quad c_3]$ only has rank one. Substituting for $L(z)$ and $U(z)$ in (22) yields

$$[c_0 + c_2z^{-1} \quad c_1 + c_3z^{-1}] = [c_0 + c_2 \quad c_1 + c_3]\,[I - P + z^{-1}P]. \tag{24}$$

The possibilities for the 4×4 projection matrix are P_1 with rank 1 and P_2 with rank 2:

$$P_1 = \frac{1}{100}\begin{bmatrix} 81 & -27\sqrt{2} & -9 & 0 \\ -27\sqrt{2} & 18 & 3\sqrt{2} & 0 \\ -9 & 3\sqrt{2} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad P_2 = \begin{bmatrix} \frac{49}{60} & \frac{-\sqrt{2}}{4} & \frac{-3}{20} & \frac{\sqrt{2}}{30} \\ \frac{-\sqrt{2}}{4} & \frac{3}{10} & \frac{-3\sqrt{2}}{20} & \frac{1}{5} \\ \frac{-3}{20} & \frac{-3\sqrt{2}}{20} & \frac{11}{20} & \frac{-3\sqrt{2}}{10} \\ \frac{\sqrt{2}}{30} & \frac{1}{5} & \frac{-3\sqrt{2}}{20} & \frac{1}{3} \end{bmatrix}. \quad (25)$$

The other step in the construction is to extend the 2×4 matrix $L(1)$ to a 4×4 matrix $H(1)$ with $HH^T = 2I$. One extension is

$$H(1) = \begin{bmatrix} L(1) \\ B(1) \end{bmatrix} = \begin{bmatrix} c_0 + c_2 & c_1 + c_3 \\ d_0 + d_2 & d_1 + d_3 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 6 & 8\sqrt{2} & 6 & 0 \\ 4\sqrt{2} & -6 & 4\sqrt{2} & 10 \\ -10 & 0 & 10 & 0 \\ -4\sqrt{2} & 6 & -4\sqrt{2} & 10 \end{bmatrix}. \quad (26)$$

There is freedom to rotate the plane of those last two rows. They can be premultiplied by the matrix $\begin{bmatrix} c & s \\ -s & c \end{bmatrix}$, with $c = \cos\phi$ and $s = \sin\phi$ and no change in orthogonality.

Finally we multiply those two rows by $U(z)$ to find $B(z) = [d_0 + d_2 z^{-1} \quad d_1 + d_3 z^{-1}]$. Using the projection P_1 in $U = I - P_1 + z^{-1}P_1$, we read off the wavelet coefficients

$$[d_0 \;\; d_1 \;\; d_2 \;\; d_3] = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \frac{1}{10}\begin{bmatrix} -1 & -3\sqrt{2} & 9 & 0 & -9 & 3\sqrt{2} & 1 & 0 \\ \frac{1}{\sqrt{2}} & 3 & -\frac{9}{\sqrt{2}} & 10 & -\frac{9}{\sqrt{2}} & 3 & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}. \quad (27)$$

This is the first family, with rotation angle ϕ. The choice $\phi = \pi/2$ yields our symmetric-antisymmetric pair.

Using the projection P_2 in the same way yields the second family

$$[d_0 \;\; d_1 \;\; d_2 \;\; d_3] = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}\begin{bmatrix} -\frac{1}{30} & -\frac{\sqrt{2}}{10} & -\frac{3}{10} & \frac{\sqrt{2}}{3} & -\frac{29}{30} & \frac{\sqrt{2}}{10} & \frac{7}{10} & -\frac{\sqrt{2}}{3} \\ -\frac{\sqrt{2}}{60} & -\frac{1}{10} & \frac{3\sqrt{2}}{20} & \frac{1}{3} & -\frac{23\sqrt{2}}{60} & \frac{7}{10} & -\frac{11\sqrt{2}}{20} & \frac{2}{3} \end{bmatrix}. \quad (28)$$

Choosing $s = \sin\phi = \sqrt{\frac{2}{3}}$, each family contains a wavelet with support length reduced from 2 to $\frac{3}{2}$. The other wavelet in each pair still has support $[0, 2]$. We have gone through both families to search for shorter support; it is not possible. Nor can both wavelets have support length $\frac{3}{2}$. This is proved in [2].

The simple expression for the odd-even pair on $[0, 2]$ may make that the preferred choice.

Conclusion

This paper takes a further step in the construction of "double wavelets". The extra freedom allows orthogonality (perfect reconstruction) with short support and symmetry-antisymmetry. We anticipate that this new direction in the design of wavelets and filters can be extended toward higher accuracy (more vanishing moments) and improved performance .

References

[1] J. S. Geronimo, D. P. Hardin, and P. R. Massopust, *Fractal functions and wavelet expansions based on several scaling functions* , J. Approximation Theory, to appear.

[2] G. Donovan, J. S. Geronimo, D. P. Hardin, and P. R. Massopust, *Construction of orthogonal wavelets using fractal interpolation functions*, preprint.

[3] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia (1992).

[4] G. Strang, *Wavelet transforms versus Fourier transforms*, Bull. Amer. Math. Soc. **28** (1993) 288–305: *Wavelets and dilation equations: A brief introduction*, SIAM Review **31** (1989) 614–627.

[5] G. Strang and G. Fix, *An Analysis of the Finite Element Method*, Wellesley-Cambridge Press, Wellesley MA (1973).

[6] G. Strang and V. Strela, *Orthogonal multiwavelets with vanishing moments*, SPIE Proceedings, Orlando (April 1994); Optical Engineering, to appear.

[7] T. N. T. Goodman and S. L. Lee, *Wavelets of multiplicity r*, preprint.

[8] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs (1993).

[9] O. Rioul and M. Vetterli, *Wavelets and signal processing*, IEEE Signal Processing Mag. **8** (1991) 14–38.

[10] M. Vetterli and D. J. Legall, *Perfect reconstruction FIR filter banks: Some properties and factorizations*, IEEE Trans. ASSP **37** (1989) 267–278.