Wavelets—A New Orthonormal Basis

Jack K. Cohen

September 1, 1992

1 Introduction

The classical orthogonal bases in applied mathematics are the Fourier and Fourier-like (Legendre polynomials, Hermite polynomials, etc.) bases. The use and success of Fourier methods is so pervasive that many of us have never considered a serious limitation of such bases: they are *non-local*. All the elements of the basis are active for all time t (e.g. for each m, the basis functions $\sin 2\pi mt$ have their support in the entire interval $[0, 1]$). Consequently, Fourier series (usually) converge by dint of enormous cancellations. Given a localized function, a *localized* basis would have the advantage of providing good approximations with only a few terms. Scientists have long sought bases that are localized *both* in time and frequency (or other pairs of conjugate variables). Short time or "windowed" Fourier transforms/series have been extensively studied (Gabor and Slepian are two of the big names) and this work is both fascinating and extremely important. However, it forces one to pick either abutting windows—and get "ringing", that is, poor frequency fall-off (only $1/\omega$ fall-off) or to pick overlapping windows—and lose orthogonality. *Warning*: The last statement is an oversimplification.

An orthonormal *wavelet basis* is a basis generated by a single function—the "mother wavelet", often denoted by $\psi(x)$. The basis is constructed from this single function by the operations of translations and dyadic dilations (other dilations are possible, including continuous dilations, but dyadic dilations are especially nice). Explicitly, the basis is given by the two parameter family

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

Writing $\psi_{j,k}(x)$ as $2^{j/2}\psi(2^j[x - k/2^j])$, we see that the translation steps are $k/2^j$. This means when the scale is doubled, the translation step between wavelets is halved:

$$\left\{\begin{array}{l}\text{large } j\ (j >> 0)\\ \text{large scale}\\ \text{high frequency}\end{array}\right. \quad \Longleftrightarrow \quad \textit{small}\text{ steps}$$

$$\left\{\begin{array}{l}\text{small } j\ (j << 0)\\ \text{small scale}\\ \text{low frequency}\end{array}\right. \quad \Longleftrightarrow \quad \textit{large}\text{ steps}$$

Thus, the time-frequency plane is tiled like this:

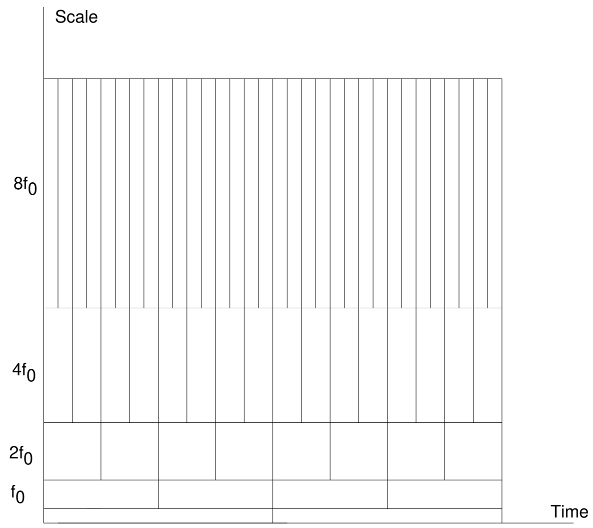

Figure 1: Equal area tiling of time-scale plane.

The equal area boxes indicate the Heisenberg uncertainty trade-off at different scales—contrast this picture with the tiling by the fixed size rectangles afforded by a windowed Fourier representation. *Warning*: The rectangles should not be taken too literally. For example, the support of $\psi_{j,k}(x)$ is not strictly within the j, k-rectangle. However, the diagram rightly indicates the important fact that high frequency events get the fine time resolution they need, while low frequency events get the long time window they require.

There are several methods of constructing orthogonal wavelet families. Chapter 5 of *Ten Lectures on Wavelets* by Daubechies is a good place to look for a full development. The remainder of this article outlines the most direct construction I could devise from the Daubechies chapter. Taking this direct path necessitates omitting some useful insights, especially the multi-resolution analysis of Mallat and Daubechies.

2 A Wavelet Construction

The construction presented here uses an auxiliary function called the *scaling function*, often denoted by $\phi(x)$.

Assumption 1 *There exist a set of constants $\{c_n\}$ such that*

$$\phi(x) = \sum_n c_n \phi(2x - n). \tag{1}$$

Equation (1) is called the "scaling equation". If the factor of 2 in the argument of ϕ on the right side were not there, the scaling equation would be an ordinary difference equation that could be solved by Fourier transform. The 2 is what makes the study of wavelets non-trivial. Without some additional assumptions, equation (1) could lead to some strange functions, ordinarily found only in the pathology section of Real Analysis textbooks.

Exercise 1 *Solve the specific 2-scale equation: $\phi(x) = 2\phi(2x)$.*

Exercise 2 *Show that equation (1) is equivalent to the Fourier domain equation*

$$\hat{\phi}(\xi) = m_0(\xi/2)\hat{\phi}(\xi/2), \tag{2}$$

where $m_0(\xi)$ is a 1-periodic function that is related to the $\{c_n\}$ by

$$m_0(\xi) = \frac{1}{2}\sum_n c_n e^{-2\pi i n\xi} \text{ or } c_n = 2\int_0^1 d\xi\, e^{2\pi i n\xi} m_0(\xi). \tag{3}$$

Here, the Fourier transform of $f(x)$ is taken to be $\hat{f}(\xi) = \int_{-\infty}^{+\infty} dx\, e^{-2\pi i x\xi} f(x)$ with inversion $f(x) = \int_{-\infty}^{+\infty} d\xi\, e^{2\pi i x\xi}\hat{f}(\xi)$.

Assumption 2 *The set $\{\phi_{0,n}(x)\}$ is an orthonormal sequence in $L^2(\mathcal{R})$, that is,*

$$\int_{-\infty}^{+\infty} dx\, \bar{\phi}(x-m)\phi(x-n) = \delta_{m,n}. \tag{4}$$

Exercise 3 *Show that Assumption (2) is equivalent to the Fourier domain equation*

$$\sum_l |\hat{\phi}(\xi + l)|^2 = 1. \tag{5}$$

Finally, to guarantee *smooth* solutions of the scaling equation, we make two more assumptions:

Assumption 3 *$\hat{\phi}(\xi)$ is bounded for all ξ.*

Assumption 4 *$\hat{\phi}(\xi)$ is continuous and non-zero at $\xi = 0$.*

Exercise 4 *Show that Assumption 4 is equivalent to the temporal domain result: $\int dx\, \phi(x) \neq 0$.*

As a practical matter, we also assume that both ϕ and $\hat{\phi}$ have reasonable decay (better than $1/x$ fall-off), so that the corresponding wavelet ψ affords good time-frequency localization.

Theorem 1 *If Assumptions 1-4 hold and we define*

$$\hat{\psi}(\xi) = e^{-\pi i \xi}\, \bar{m}_0(\xi/2 + 1/2)\hat{\phi}(\xi/2), \tag{6}$$

then the collection $\{\psi_{j,k}(x)\}$ is an orthonormal basis of $L^2(\mathcal{R})$.

In particular, for an arbitrary function $f \epsilon L^2(\mathcal{R})$, we have

$$f(x) = \sum \gamma_{j,k}\psi_{j,k}, \quad \gamma_{j,k} =< \psi_{j,k}, f >,$$

where $< \psi_{j,k}, f >$ represents the inner product integral $\int dx\, \bar{\psi}_{j,k}(x)\, f(x)$.

The proof of Theorem 1 takes most of Daubechies' Chapter 5, so is omitted here.

Exercise 5 *Show that Equation (6) is equivalent to the temporal domain result: $\psi(x) = \sum_n d_n \phi(2x-n), \quad d_n = (-1)^{n-1} c_{1-n}$.*

Exercise 6 *Show that* $\widehat{\psi_{j,k}}(x) = 2^{-j/2} e^{-2\pi i k\xi/2^j} \hat{\psi}(\xi/2^j)$ *(so that as* j *increases,* $\psi_{j,k}$ *becomes more localized, while* $\hat{\psi}_{j,k}$ *becomes less localized).*

Theorem 2

$$|m_0(\xi)|^2 + |m_0(\xi + 1/2)|^2 = 1. \qquad (7)$$

Proof: Use equation (2) in equation (5) to obtain

$$\sum_l |m_0(\xi/2 + l/2)|^2 |\hat{\phi}(\xi/2 + l/2)|^2 = 1.$$

Break this sum into even and odd terms:

$$\sum_n |m_0(\xi/2+n)|^2 |\hat{\phi}(\xi/2+n)|^2 + \sum_n |m_0(\xi/2+n+1/2)|^2 |\hat{\phi}(\xi/2+n+1/2)|^2 = 1.$$

Since $m_0(\xi)$ has period 1, we can factor out the m_0's to obtain:

$$|m_0(\xi/2)|^2 \sum_n |\hat{\phi}(\xi/2 + n)|^2 + |m_0(\xi/2 + 1/2)|^2 \sum_n |\hat{\phi}((\xi+1)/2 + n)|^2 = 1.$$

But by Assumption 2, both remaining sums are equal to 1. □

Exercise 7 *Show that* $m_0(0) = 1$ *and* $m_0(1/2) = 0$.

Exercise 8 *Show that* $\hat{\phi}(k) = 0$ *for all non-zero integers* k.

Exercise 9 *Show that* $|\int dx\, \phi(x)| = 1$.

Because of the result of Exercise 7, $|m_0|$ can be considered as a *low pass* filter (recall that m_0 is 1-periodic). Figure 2 shows an example that will be discussed in detail in the next section.

Defining $m_1(\xi) = e^{-2\pi i\xi}\, \bar{m}_0(\xi + 1/2)$, equation (6) can be recast as $\hat{\psi}(\xi) = m_1(\xi/2)\hat{\phi}(\xi/2)$ and equation (7) as $|m_0(\xi)|^2 + |m_1(\xi)|^2 = 1$.

Exercise 10 *Show that* $m_1(0) = 0$ *and* $|m_1(1/2)| = 1$.

Because of this result, m_1 can be considered as a *high pass* filter, see Figure 3.

Notice that neither the "low pass" nor the "high pass" filters are *strict*—they each have a tail in the wrong half of the spectrum. The choice of scaling

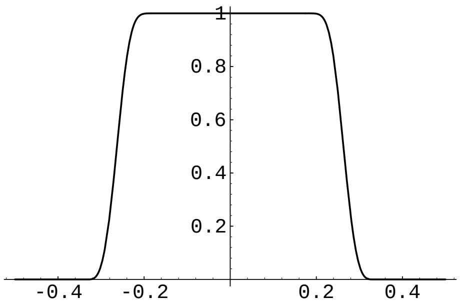

Figure 2: A Meyer $|m_0|$ filter on the period $[-1/2, 1/2]$.

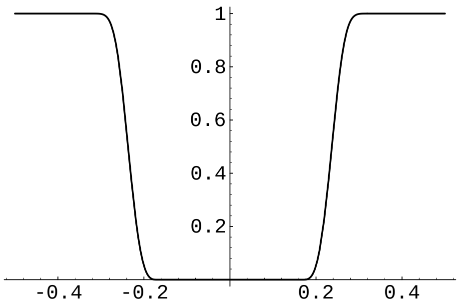

Figure 3: A Meyer $|m_1|$ filter on the period $[-1/2, 1/2]$.

function (or mother wavelet) can decrease this effect, but this is a matter of trading off one desirable feature against another: good localization.

It is remarkable that both the scaling function and the mother wavelet come from convolution with essentially the same filter (except for a phase factor, m_1 is just a translate of m_0). Because of the symmetry about $\pm 1/4$ shown in Figure 4, such filters are known as "Quadrature *Mirror* Filters" or sometimes "Conjugate *Mirror* Filters".

3 Example—The Meyer Wavelet

3.1 The Taper Function

The Meyer wavelet is defined in terms of a taper function $\nu(x)$ with the following properties:

1. $\nu(x)$ is continuous and monotonically increasing.

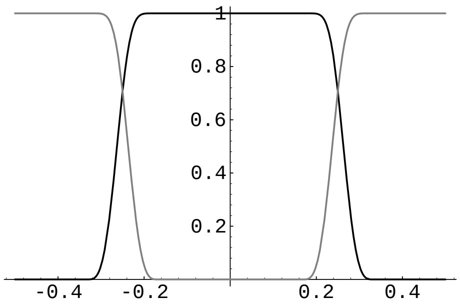

Figure 4: The Meyer $|m_0|$ and $|m_1|$ filters on the period $[-1/2, 1/2]$.

2. $\nu(x) \equiv 0$ for $x \leq 0$.
3. $\nu(x) \equiv 1$ for $x \geq 1$.
4. $\nu(x) + \nu(1-x) = 1$ for $0 < x < 1$.

The smoothness of ν dictates the smoothness of the Meyer $\hat{\phi}$, moreover, the smoother ν, the better the frequency localization. It is easy to find functions that rise smoothly from 0 to 1 on the unit interval, but smoothness at $x = 0$ depends on the order of the zero of $\nu(x)$ as $x \to 0^+$ and the analogous order of matching at $x = 1$. The fourth condition on ν listed above guarantees that each zero at $x = 0$ is matched by a factor of $1 - x$ at $x = 1$. For example, condition 4 shows that $\nu(0) = 0$ implies $\nu(1) = 1$. Similarly, if also $\nu'(0) = 0$, then condition 4 shows that $\nu'(1) = 0$ and so forth.

It is not difficult to think up a few ν functions:

Exercise 11 *Show that defining $\nu(x) = x$ for $0 < x < 1$ yields a continuous taper function.*

Exercise 12 *Show that defining $\nu(x) = \sin^2(\pi x/2)$ for $0 < x < 1$ yields a taper function with a continuous derivative.*

It turns out that there exist *polynomial* taper functions with an arbitrary degree of match-up at $x = 0$ and $x = 1$. From the above considerations, we need only define

$$\nu(x) = \begin{cases} 0 & x \leq 0 \\ x^n \sum_{k=0}^{n-1} a_k x^k & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

and then pick the a_n to satisfy condition 4, that is, solve n linear equations in n unknowns. Here is a *Mathematica* module to carry out this calculation:

```
NuTaper[n_Integer?Positive, x_] :=
Module[
        {v, xi, a, eqs, solns},

        v[xi_] = xi^n * Sum[a[k] xi^k, {k, 0, n-1}];

        eqs =
            CoefficientList[v[xi] + v[1 - xi] - 1, xi] == 0 //Thread;

        solns = Solve[eqs, Array[a, n, 0]];

        v[x] /. solns //First //Factor
]
```

And here is a transcript of the calculation for the case of continuous third derivative, $n = 4$:

```
v[x] = NuTaper[4, x]
     4                  2        3
    x  (35 - 84 x + 70 x  - 20 x )
```

Remark: Daubechies cites this example repeatedly, unfortunately misprinting the quadratic term as $80x^2$. One can see that this is wrong by evaluating at $x = 1$ and this error motivated writing the above module. Let's verify that the above polynomial has the right behavior at $x = 1$:

```
Factor@Normal@Series[1 - v[x], {x, 1, 7}]
            4                  2        3
    (-1 + x)  (1 + 4 x + 10 x  + 20 x )
```

so, indeed, we have 4^{th} order matching at $x = 1$ too.

3.2 The Meyer Wavelet

The Meyer scaling function is defined in Fourier domain by the explicit formula:

$$\hat{\phi}(\xi) = \begin{cases} 1 & |\xi| \leq 1/3 \\ \cos[\frac{\pi}{2}\nu(3|\xi| - 1)] & 1/3 < |\xi| < 2/3 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where ν is an arbitrary taper function with the properties listed in the previous sub-section. See Figure 5 for an example. Remark: Even allowing for the

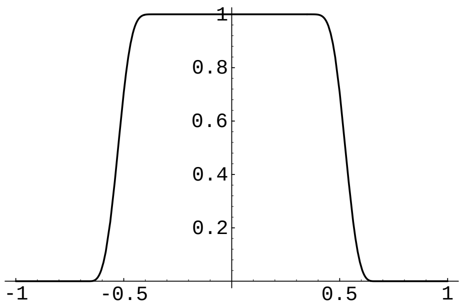

Figure 5: Meyer's $\hat{\phi}(\xi)$ with $\nu(x) = x^4(35 - 84x + 70x^2 - 20x^3)$.

various 2π's scattered around by Daubechies's Fourier transform convention, there is a misprint in her analog of equation (8): in her expression on page 137, the factor $3/4\pi$ should be replaced by $3/2\pi$.

The Meyer scaling function has compact support in frequency domain, so is certainly localized in frequency. Figure 6 shows that it is localized in time as well. As we will see, this dual time-frequency localization will carry over to the wavelet basis functions as well.

We now set about satisfying Assumption 1 in its Fourier domain form given by equation (2). Examining that equation, it seems natural to *define* $m_0(\xi) = \hat{\phi}(2\xi)/\hat{\phi}(\xi)$. But watch out! We must be careful to avoid dividing by 0 and moreover, $m_0(\xi)$ is supposed to be 1-periodic. The solution is to define m_0 as indicated *only* in the fundamental period interval $-1/2 < \xi < 1/2$ and then extend it periodically off this interval. But notice that the denominator $\hat{\phi}(\xi)$ is identically 1 for $|\xi| < 1/3$, while the support of the numerator $\hat{\phi}(2\xi)$ is itself this *same* interval! Hence we can omit the denominator on the fundamental period, and the definition becomes

$$m_0(\xi) = \hat{\phi}(2\xi), \quad |\xi| < 1/2. \tag{9}$$

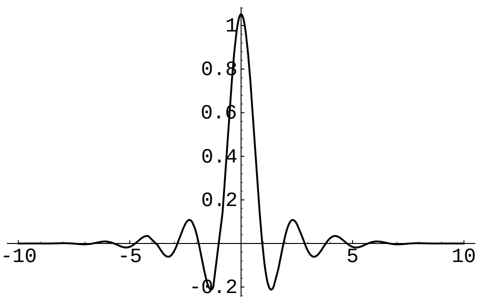

Figure 6: Meyer's $\phi(x)$ with $\nu(x) = x^4(35 - 84x + 70x^2 - 20x^3)$.

From the considerations leading to this equation, it is obvious that equation (2) and thus Assumption 1 is satisfied. This triumph is a first indication of Meyer's cleverness. See Figure 7 and also Figure 2 for graphs of m_0.

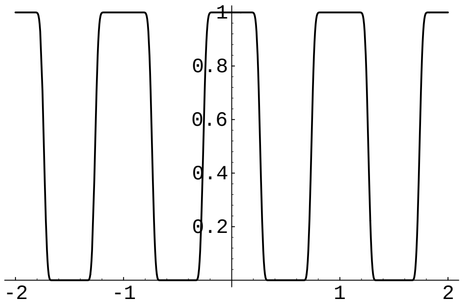

Figure 7: Meyer's $m_0(\xi)$ with $\nu(x) = x^4(35 - 84x + 70x^2 - 20x^3)$.

We next show that Assumption 2 is satisfied, again using the Fourier domain form given by equation (5). Figure 8 shows 5 successive summands of equation (5).

On the interval $[-1/3, 1/3]$ and its periodic replicas, only one summand in equation (5) is non-zero and that summand is indeed 1. Thus, we need only check the sum on the "overlap" region $[1/3, 2/3]$ and its periodic replicas. In these regions, exactly *two* terms are non-zero. For example, on $[1/3, 2/3]$, the sum reduces to $\hat{\phi}^2(\xi) + \hat{\phi}^2(\xi - 1)$. In turn, using the definition of $\hat{\phi}$, this quantity is given by $\cos^2[\frac{\pi}{2}\nu(3\xi - 1)] + \cos^2[\frac{\pi}{2}\nu(-3(\xi - 1) - 1)]$. Let $3\xi - 1 = x$, then $-3(\xi - 1) - 1 = 2 - 3\xi = 1 - x$ and the sum in question reduces to $\cos^2[\frac{\pi}{2}\nu(x)] + \cos^2[\frac{\pi}{2}\nu(1 - x)]$. Finally use the taper equation $\nu(1 - x) = 1 - \nu(x)$ to write the second summand as $\cos^2[\frac{\pi}{2}(1 - \nu(x)] = \sin^2[\frac{\pi}{2}\nu(x)]$. Thus

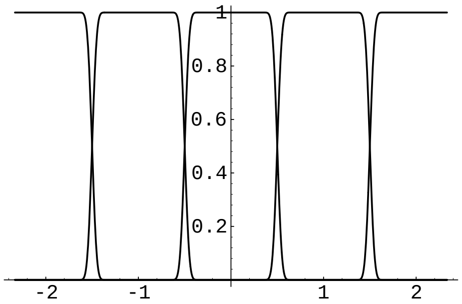

Figure 8: Five successive summands of equation (5).

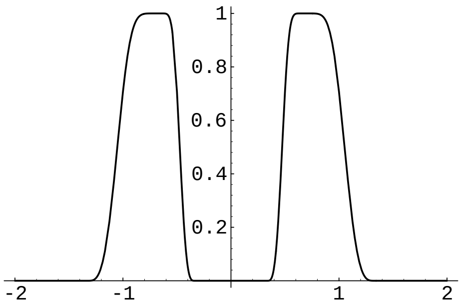

Figure 9: Meyer's $\hat{\psi}(\xi)$ with $\nu(x) = x^4(35 - 84x + 70x^2 - 20x^3)$.

by the Pythagorean identity, $\cos^2\theta + \sin^2\theta = 1$, the sum in the overlap region is also 1 and Assumption 2 is established. Since Assumptions 3 and 4 are patently true, we can apply Theorem 1 to obtain the Meyer mother wavelet of an orthonormal family $\{\psi_{j,k}\}$ in $L^2(\mathcal{R})$. Figures 9 and 10 show the Meyer mother wavelet in respectively Fourier and temporal domain. Notice that in temporal domain, the Meyer wavelet is not strictly of compact support—but it is clearly well localized in time nonetheless.

Exercise 13 *Show that*

$$\hat{\psi}(\xi) = e^{-\pi i\xi}(\hat{\phi}(\xi+1) + \hat{\phi}(\xi-1))\hat{\phi}(\xi/2) \tag{10}$$

Exercise 14 *Show that*

$$\hat{\psi}(\xi) = \begin{cases} e^{-\pi i\xi}\sin[\frac{\pi}{2}\nu(3|\xi|-1)] & 1/3 \le |\xi| \le 2/3 \\ e^{-\pi i\xi}\cos[\frac{\pi}{2}\nu(\frac{3}{2}|\xi|-1)] & 2/3 < |\xi| < 4/3 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

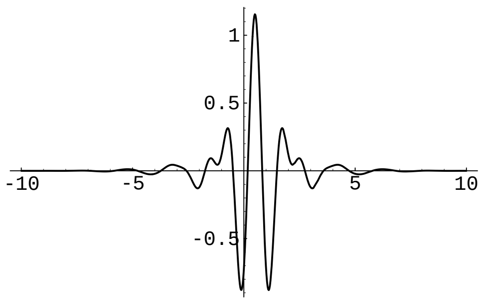

Figure 10: Meyer's $\psi(x)$ with $\nu(x) = x^4(35 - 84x + 70x^2 - 20x^3)$.

Meyer did not have Theorem 1 when he constructed his example, indeed, the original construction is deemed "miraculous" by Daubechies. Thus

Theorem 3 *Meyer is a genius.*

There are other wavelets besides the ones constructed by Meyer and their tribe increases daily. The most exciting development in recent years was the construction of wavelets with compact support in time. This construction was done by Daubechies. Hence our final theorem:

Theorem 4 *Daubechies is a genius too.*