

Fourier Analysis and Wavelet Analysis

James S. Walker

In this article we will compare the classical methods of Fourier analysis with the newer methods of wavelet analysis. Given a signal, say a sound or an image, Fourier analysis easily calculates the frequencies and the amplitudes of those frequencies which make up the signal. This provides a broad overview of the characteristics of the signal, which is important for theoretical considerations. However, although Fourier inversion is possible under certain circumstances, Fourier methods are not always a good tool to recapture the signal, particularly if it is highly nonsmooth: too much Fourier information is needed to reconstruct the signal locally. In these cases, wavelet analysis is often very effective because it provides a simple approach for dealing with local aspects of a signal. Wavelet analysis also provides us with new methods for removing noise from signals that complement the classical methods of Fourier analysis. These two methodologies are major elements in a powerful set of tools for theoretical and applied analysis.

This article contains many graphs of discrete signals. These graphs were created by the computer program FAWAV, A Fourier-Wavelet Analyzer, being developed by the author.

James S. Walker is professor of mathematics at the University of Wisconsin-Eau Claire. His e-mail address is walkerjs@uwec.edu.

The author would like to thank Hugo Rossi, Steven Krantz, his colleague Marc Goulet, and two anonymous reviewers for their helpful comments during the writing of this article.

Frequency Information, Denoising

As an example of the importance of frequency information, we will examine how Fourier analysis can be used for removing noise from signals. Consider a signal $f(x)$ defined over the unit interval (where here x stands for time). The period 1 Fourier series expansion of f is defined by $\sum_{n \in \mathbf{Z}} c_n e^{i2\pi nx}$, with $c_n = \int_0^1 f(x) e^{-i2\pi nx} dx$. Each Fourier coefficient, c_n , is an amplitude associated with the frequency n of the exponential $e^{i2\pi nx}$. Although each of these exponentials has a precise frequency, they all suffer from a complete absence of time localization in that their magnitudes, $|e^{i2\pi nx}|$, equal 1 for all time x .

To see the importance of frequency information, let us examine a problem in noise removal. In Figure 1(a)[top] we show the graph of the signal

$$(1) \quad f(x) = (5 \cos 2\pi \nu x) [e^{-640\pi(x-1/8)^2} + e^{-640\pi(x-3/8)^2} + e^{-640\pi(x-4/8)^2} + e^{-640\pi(x-6/8)^2} + e^{-640\pi(x-7/8)^2}]$$

where the frequency, ν , of the cosine factor is 280. Such a signal might be used by a modem for transmitting the bit sequence 1 0 1 1 0 1 1. The Fourier coefficients for this signal are shown in Figure 1(b)[top]. The highest magnitude coefficients are concentrated around the frequencies ± 280 . Suppose that when this signal is received, it is severely distorted by added noise; see Figure 1(a)[middle]. Using Fourier analysis, we can remove most of this noise. Computing the noisy signal's Fourier coefficients, we obtain the graph shown in Figure 1(b)[middle]. The original signal's largest magnitude Fourier coefficients are clus-

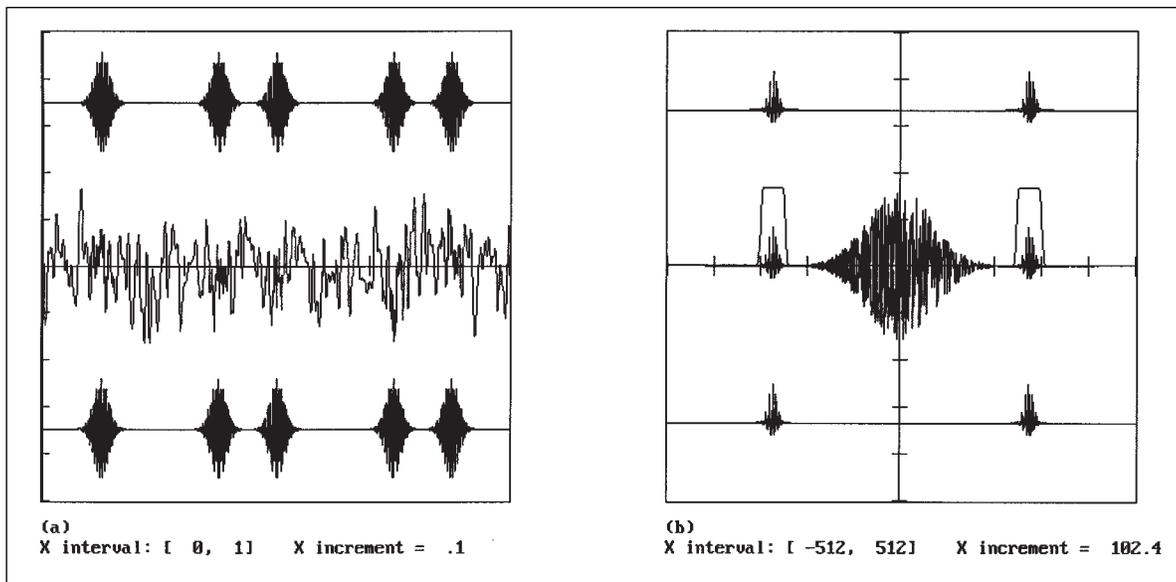


Figure 1. (a)[top] Signal. (b)[top] Fourier coefficients of signal. (a)[middle] Signal after adding noise. (b)[middle] Fourier coefficients of noisy signal and filter function. (b)[bottom] Fourier coefficients after multiplication by filter function. (a)[bottom] Recovered signal.

tered around the frequency positions ± 280 . The Fourier coefficients of the added noise are localized around the origin, and they decrease in magnitude until they are essentially zero near the frequencies ± 280 . Thus, the original signal's coefficients and the noise's coefficients are well separated. To remove the noise from the signal, we multiply the noisy signal's coefficients by a *filter function*, which is 1 where the signal's coefficients are concentrated and 0 where the noise's coefficients are concentrated. We then recover essentially all of the signal's coefficients; see Figure 1(b)[bottom]. Performing a Fourier series partial sum with these recovered coefficients, we obtain the denoised signal, which is shown in Figure 1(a)[bottom]. Clearly, the bit sequence 1 0 1 1 0 1 1 can now be determined from the denoised signal, and the denoised signal is a close match of the original signal. In the section "Signal Denoising" we shall look at another example of this method and also discuss how wavelets can be used for noise removal.

Signal Compression

As the example above shows, Fourier analysis is very effective in problems dealing with frequency location. However, it is often very ineffective at representing functions. In particular, there are severe problems with trying to analyze transient signals using classical Fourier methods. For example, in Figure 2(a)[top] we show a discrete signal obtained from $M = 1024$ values $\{f_j = F(j/M)\}_{j=0}^{M-1}$ of the function $F(x) = e^{-10^5 \pi(x-.6)^2}$. For this example, we compute the discrete Fourier series coefficients \hat{f}_n defined by

$$\hat{f}_n = M^{-1} \sum_{j=0}^{M-1} f_j e^{-i2\pi nj/M}$$

for $n = -\frac{1}{2}M + 1, \dots, 0, \dots, \frac{1}{2}M$. The discrete Fourier coefficients \hat{f}_n can be calculated by a fast Fourier transform (FFT) algorithm and are the discrete analog of the Fourier coefficients c_n for F , when $f_j = F(j/M)$. Moreover, \hat{f}_n is just a Riemann sum approximation of the integral that defines c_n .¹ The magnitudes of the discrete Fourier coefficients for this transient damp down to zero very slowly (their graph is a very wide bell-shaped curve with maximum at the origin). Consequently, to represent the transient well, one must retain most if not all of these Fourier coefficients. In Figure 2(a)[bottom] we show the results obtained from trying to compress the transient by computing a discrete partial sum $\sum_{n=-104}^{104} \hat{f}_n e^{i2\pi nj}$ using only one-fifth of the Fourier coefficients.² Clearly, even a moderate compression ratio of 5:1 is not effective.

Wavelets, however, are often very effective at representing transients. This is because they are designed to capture information over a large range of scales. A wavelet series expansion of a function f is defined by

$$\sum_{n,k \in \mathbb{Z}} \beta_k^n 2^{n/2} \psi(2^n x - k)$$

with

$$\beta_k^n = \int_{-\infty}^{\infty} f(x) 2^{n/2} \psi(2^n x - k) dx.$$

The function $\psi(x)$ is called the *wavelet*, and the coefficients β_k^n are called the *wavelet coefficients*. The function $2^{n/2} \psi(2^n x - k)$ is the

¹For further discussion of discrete Fourier coefficients, see [2] or [11].

²The sum $\sum_{n=-511}^{512} \hat{f}_n e^{i2\pi nj}$, which uses all of the discrete Fourier coefficients, equals f_j .

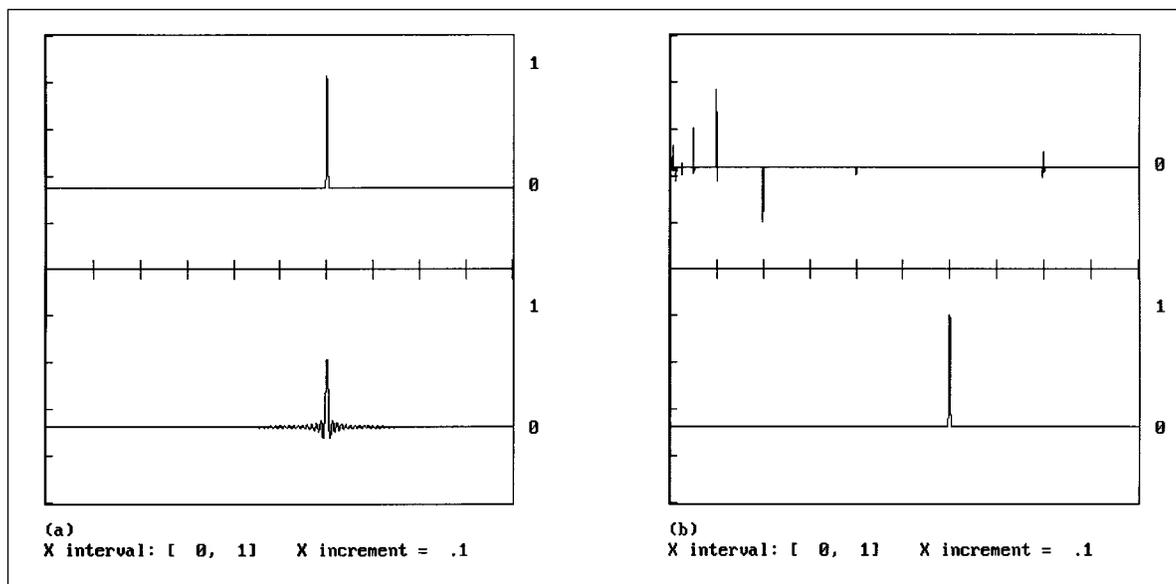


Figure 2. (a)[top] Signal. (a)[bottom] Fourier series using 205 coefficients, 5:1 compression. (b)[top] Wavelet coefficients for signal. (b)[bottom] Wavelet series using only largest 4% in magnitude of wavelet coefficients, 25:1 compression.

wavelet shrunk by a factor of 2^n if n is positive (magnified by a factor of 2^{-n} if n is negative) and shifted by $k2^{-n}$ units. The factor $2^{n/2}$ in the expression $2^{n/2}\psi(2^n x - k)$ preserves the L^2 -norm.

Since the wavelet series depends on two parameters of scale and translation, it can often be very effective in analyzing signals. These parameters make it possible to analyze a signal's behavior at a dense set of time locations and with respect to a vast range of scales, thus providing the ability to *zoom in* on the transient behavior of the signal. For example, let us examine the earlier transient using a discretized version of a wavelet series. We shall use a Daubechies order 4 wavelet (Daub4 for short; see the section "Daubechies Wavelets"). In Figure 2(b)[top] we show all of the 1024 wavelet coefficients of this transient and observe that most of these coefficients are close to 0 in magnitude. Consequently, by retaining only the largest magnitude coefficients for use in a wavelet series, we obtain significant compression. In Figure 2(b)[bottom] we show the reconstruction of the transient using only the top 4% in magnitude of the wavelet coefficients, a 25:1 compression ratio. Notice how accurately the transient is represented. In fact, the maximum error at all computed points is less than 9.95×10^{-14} . There is an important application here to the field of signal transmission. *By transmitting only these 4% of the wavelet coefficients, the information in the signal can be transmitted 25 times faster than if we transmitted all of the original signal.* This provides a considerable boost in efficiency of transmission.

We shall look at more examples of compression in the section "Compression of Signals", but first we shall describe how wavelet analysis works.

The Haar Wavelet

In order to understand how wavelet analysis works, it is best to begin with the simplest wavelet, the Haar wavelet. Let $\mathbf{1}_A(x)$ denote the indicator function of the set A , defined by $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ if $x \notin A$. The Haar wavelet ψ is defined by $\psi(x) = \mathbf{1}_{[0, \frac{1}{2})}(x) - \mathbf{1}_{[\frac{1}{2}, 1)}(x)$. It is 0 outside of $[0, 1)$, so it is well localized in time, and it satisfies

$$\int_{-\infty}^{\infty} \psi(x) dx = 0, \quad \int_{-\infty}^{\infty} |\psi(x)|^2 dx = 1.$$

The Haar wavelet $\psi(x)$ is closely related to the function $\phi(x)$ defined by $\phi(x) = \mathbf{1}_{[0, 1)}(x)$. This function $\phi(x)$ is called the *Haar scaling function*. Clearly, the Haar wavelet and scaling function satisfy the identities

$$(2) \quad \begin{aligned} \psi(x) &= \phi(2x) - \phi(2x - 1), \\ \phi(x) &= \phi(2x) + \phi(2x - 1), \end{aligned}$$

and the scaling function satisfies

$$\int_{-\infty}^{\infty} \phi(x) dx = 1, \quad \int_{-\infty}^{\infty} |\phi(x)|^2 dx = 1.$$

The Haar wavelet $\psi(x)$ generates the system of functions $\{2^{n/2}\psi(2^n x - k)\}$. It is possible to show directly that $\{2^{n/2}\psi(2^n x - k)\}$ is an orthonormal basis for $L^2(\mathbf{R})$, but it is more illuminating to put the discussion on an axiomatic level. This axiomatic approach leads to the Daubechies wavelets and many other wavelets as well. We begin by defining the subspaces $\{V_n\}_{n \in \mathbf{Z}}$ of $L^2(\mathbf{R})$ in the following way:

$V_n = \left\{ \begin{array}{l} \text{step functions in } L^2(\mathbf{R}), \text{ constant} \\ \text{on the intervals } \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right), k \in \mathbf{Z} \end{array} \right\}$.

This set of subspaces $\{V_n\}_{n \in \mathbf{Z}}$ satisfies the following five axioms [6]:

Axioms for a Multi-Resolution Analysis (MRA)

Scaling: $f(x) \in V_n$ if and only if $f(2x) \in V_{n+1}$.

Inclusion: $V_n \subset V_{n+1}$, for each n .

Density: closure $\left\{ \bigcup_{n \in \mathbf{Z}} V_n \right\} = L^2(\mathbf{R})$.

Maximality: $\bigcap_{n \in \mathbf{Z}} V_n = \{0\}$.

Basis: $\exists \phi(x)$ such that $\{\phi(x - k)\}_{k \in \mathbf{Z}}$ is an orthonormal basis for V_0 .

To satisfy the basis axiom, we shall use the Haar scaling function ϕ defined above. Then, by combining the scaling axiom with the basis axiom, we find that $\{2^{n/2}\phi(2^n x - k)\}_{k \in \mathbf{Z}}$ is an orthonormal basis for V_n . But the totality of all these orthonormal bases, consisting of the set $\{2^{n/2}\phi(2^n x - k)\}_{k, n \in \mathbf{Z}}$, is *not* an orthonormal basis for $L^2(\mathbf{R})$ because the spaces V_n are not mutually orthogonal. To remedy this difficulty, we need what are called *wavelet subspaces*. Define the wavelet subspace W_n to be the orthogonal complement of V_n in V_{n+1} . That is, W_n satisfies the equation $V_{n+1} = V_n \oplus W_n$ where \oplus denotes the sum of mutually orthogonal subspaces. From the density axiom and repeated application of the last equation, we obtain $L^2(\mathbf{R}) = V_0 \oplus \bigoplus_{n=0}^{\infty} W_n$. Decomposing V_0 in a similar way, we obtain $L^2(\mathbf{R}) = \bigoplus_{n \in \mathbf{Z}} W_n$. Thus, $L^2(\mathbf{R})$ is an orthogonal sum of the wavelet subspaces W_n .

Using (2) and the MRA axioms, it is easy to prove the following lemma.

Lemma 1. The functions $\{\psi(x - k)\}_{k \in \mathbf{Z}}$ are an orthonormal basis for the subspace W_0 .

It follows from the scaling axiom that $\{2^{n/2}\psi(2^n x - k)\}_{k \in \mathbf{Z}}$ is an orthonormal basis for W_n . Therefore, since $L^2(\mathbf{R})$ is the orthogonal sum of all the wavelet subspaces W_n , we have obtained the following result.

Theorem 1. The functions

$$\{2^{n/2}\psi(2^n x - k)\}_{k, n \in \mathbf{Z}}$$

are an orthonormal basis for $L^2(\mathbf{R})$.

This orthonormal basis is the *Haar basis* for $L^2(\mathbf{R})$. There is also a Haar basis for $L^2[0, 1)$. To obtain it, we first define *periodic wavelets* $\tilde{\psi}_{n,k}$ by

$$(3) \quad \tilde{\psi}_{n,k}(x) = \sum_{j \in \mathbf{Z}} 2^{n/2} \psi(2^n(x + j) - k).$$

Note that these wavelets have period 1. Furthermore, $\tilde{\psi}_{n,k} \equiv 0$ for $n < 0$, and $\tilde{\psi}_{n,k+2^n} = \tilde{\psi}_{n,k}$ for all $k \in \mathbf{Z}$ and $n \geq 0$. On the interval $[0, 1)$, the periodic Haar wavelets $\tilde{\psi}_{n,k}$ satisfy $\tilde{\psi}_{n,k}(x) = 2^{n/2}\psi(2^n x - k)$ for $n \geq 0$ and $k = 0, 1, \dots, 2^n - 1$. So we have the following theorem as a consequence of Theorem 1.

Theorem 2. The functions 1 and $\tilde{\psi}_{n,k}$ for $n \geq 0$ and $k = 0, 1, \dots, 2^n - 1$ are an orthonormal basis for $L^2[0, 1)$.

Remark. In the section “Daubechies Wavelets” we will make use of periodized scaling functions, $\tilde{\phi}_{n,k}$, defined by

$$(4) \quad \tilde{\phi}_{n,k}(x) = \sum_{j \in \mathbf{Z}} 2^{n/2} \phi(2^n(x + j) - k)$$

for $n \geq 0$ and $k = 0, 1, \dots, 2^n - 1$.

Fast Haar Transform

The relation between the Haar scaling function ϕ and wavelet ψ leads to a beautiful set of relations between their coefficients as bases. Let $\{\alpha_k^n\}$ and $\{\beta_k^n\}$ be defined by

$$(5) \quad \begin{aligned} \alpha_k^n &= \int_{-\infty}^{\infty} f(x) 2^{n/2} \phi(2^n x - k) dx, \\ \beta_k^n &= \int_{-\infty}^{\infty} f(x) 2^{n/2} \psi(2^n x - k) dx. \end{aligned}$$

Substituting $2^n x$ in place of x in the identities in (2), we obtain

$$\begin{aligned} 2^{n/2} \phi(2^n x) &= \frac{1}{\sqrt{2}} [2^{\frac{n+1}{2}} \phi(2^{n+1} x)] \\ &\quad + \frac{1}{\sqrt{2}} [2^{\frac{n+1}{2}} \phi(2^{n+1} x - 1)] \end{aligned}$$

$$\begin{aligned} 2^{n/2} \psi(2^n x) &= \frac{1}{\sqrt{2}} [2^{\frac{n+1}{2}} \phi(2^{n+1} x)] \\ &\quad - \frac{1}{\sqrt{2}} [2^{\frac{n+1}{2}} \phi(2^{n+1} x - 1)]. \end{aligned}$$

It then follows that

$$(6) \quad \begin{aligned} \alpha_k^n &= \frac{1}{\sqrt{2}} \alpha_{2k}^{n+1} + \frac{1}{\sqrt{2}} \alpha_{2k+1}^{n+1}, \\ \beta_k^n &= \frac{1}{\sqrt{2}} \alpha_{2k}^{n+1} - \frac{1}{\sqrt{2}} \alpha_{2k+1}^{n+1}. \end{aligned}$$

This result shows that the n^{th} level coefficients α_k^n and β_k^n are obtained from the $(n + 1)^{\text{st}}$ level coefficients α_k^{n+1} through multiplication by the following orthogonal matrix:

$$(7) \quad \mathcal{O} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}.$$

Successively applying this orthogonal matrix \mathcal{O} , we obtain $\{\alpha_k^n\}$ and $\{\beta_k^n\}$ starting from some highest level coefficients $\{\alpha_k^N\}$ for some large N . Because of the density axiom, N can be chosen large enough to approximate f by $\sum_{k \in \mathbb{Z}} \alpha_k^N 2^{N/2} \phi(2^N x - k)$ in L^2 -norm as closely as desired.

Let us now discretize these results. Suppose that we are working with data $\{f_j\}_{j=0}^{M-1}$ associated with the time values $\{x_j = j/M\}_{j=0}^{M-1}$ on the unit interval. If we shrink the Haar scaling function $\phi(x) = \mathbf{1}_{[0,1)}(x)$ enough, it covers only the first point, $x_0 = 0$. Consequently, by choosing a large enough N , we may assume that our scaling coefficients $\{\alpha_k^N\}$ satisfy $\alpha_k^N = f_k$, for $k = 0, 1, \dots, M-1$. Assuming that M is a power of 2, say $M = 2^R$, it follows that $N = R$.

The next step involves expressing the coefficient relations in (6) in a matrix form. Let $A \oplus B$ stand for the orthogonal sum of the matrices A and B , that is, $A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$. Now let H_M denote the $M \times M$ orthogonal matrix defined by $H_M = \mathcal{O} \oplus \mathcal{O} \oplus \dots \oplus \mathcal{O}$, where the orthogonal matrix sums are applied $M/2$ times and \mathcal{O} is the matrix defined in (7). Then, by applying the coefficient relations in (6) and using the fact that $f_k = \alpha_k^R$, we obtain

$$\begin{aligned} & H_M [f_0, f_1, \dots, f_{M-1}]^T \\ &= \left[\alpha_0^{R-1}, \beta_0^{R-1}, \alpha_1^{R-1}, \beta_1^{R-1}, \right. \\ & \quad \left. \dots, \alpha_{\frac{1}{2}M-1}^{R-1}, \beta_{\frac{1}{2}M-1}^{R-1} \right]^T. \end{aligned}$$

To sort the coefficients properly into two groups, we apply an $M \times M$ permutation matrix P_M as follows:

$$\begin{aligned} & P_M \left[\alpha_0^{R-1}, \beta_0^{R-1}, \dots, \alpha_{\frac{1}{2}M-1}^{R-1}, \beta_{\frac{1}{2}M-1}^{R-1} \right]^T \\ &= \left[\alpha_0^{R-1}, \dots, \alpha_{\frac{1}{2}M-1}^{R-1}, \beta_0^{R-1}, \dots, \beta_{\frac{1}{2}M-1}^{R-1} \right]^T. \end{aligned}$$

If we go to the next lower level, the transformations just described are repeated, only now the matrices used are $H_{M/2}$ and $P_{M/2}$, and they operate only on the $M/2$ -length vector $\{\alpha_0^{R-1}, \dots, \alpha_{\frac{1}{2}M-1}^{R-1}\}$ to obtain the next level wavelet coefficients $\{\beta_0^{R-2}, \dots, \beta_{\frac{1}{4}M-1}^{R-2}\}$ and scaling coefficients $\{\alpha_0^{R-2}, \dots, \alpha_{\frac{1}{4}M-1}^{R-2}\}$. These operations continue until we can no longer divide the number of components by 2. At the $R = \log_2 M$ step, we obtain a single scale coefficient

α_0^0 and a single wavelet coefficient β_0^0 , and at this last step the permutation P_2 is unnecessary. The complete transformation, denoted by \mathcal{H} , satisfies

$$\begin{aligned} \mathcal{H} &= (H_2 \oplus I_{M-2}) \\ & \quad \dots (P_{M/2} \oplus I_{M/2})(H_{M/2} \oplus I_{M/2})P_M H_M \end{aligned}$$

where I_N is the $N \times N$ identity matrix.

These matrix multiplications can be performed rapidly on a computer. Multiplying by H_M requires only $O(M)$ operations, since H_M consists mostly of zeroes. Similarly, the permutation P_M requires $O(M)$ operations. Therefore, the whole transformation requires $O(M) + O(M/2) + \dots + O(2) = O(M)$ operations. The transformation \mathcal{H} is called a *fast Haar transform*. It should be noted that FFTs, which have revolutionized scientific practice during the last thirty years, are $O(M \log M)$ algorithms.

Since each H_k is an orthogonal matrix, and so is every permutation P_k , it follows that \mathcal{H} is invertible. Its inverse is

$$\begin{aligned} \mathcal{H}^{-1} &= H_M^T P_M^T (H_{M/2}^T \oplus I_{M/2}) (P_{M/2}^T \oplus I_{M/2}) \\ & \quad \dots (H_2^T \oplus I_{M-2}). \end{aligned}$$

Therefore, the inverse operation is also an $O(M)$ operation.

Discrete Haar Series

The fast Haar transform can be used for computing partial sums of the discretized version of the following Haar wavelet series in $L^2[0, 1)$:

$$(8) \quad \alpha_0^0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \beta_k^n 2^{n/2} \psi(2^n x - k).$$

Let us assume, as in the previous section, that we have a discrete signal $\{f_j\}_{j=0}^{M-1}$ associated with the time values $\{x_j = j/M\}_{j=0}^{M-1}$ on the unit interval. Substituting these time values into (8) and restricting the upper limit of n , we obtain

$$f_j = \alpha_0^0 C_M + \sum_{n=0}^{R-1} \sum_{k=0}^{2^n-1} \beta_k^n C_M 2^{n/2} \psi(2^n x_j - k).$$

The right side of this equation is just the transformation $\mathcal{H}^{-1} \mathcal{H}$ applied to $\{f_j\}$. The first part of this transformation, $\mathcal{H} \{f_j\}$, produces the coefficients $\alpha_0^0, \beta_0^0, \beta_0^1, \beta_1^1, \dots, \beta_{\frac{1}{2}M-1}^{R-1}$, and the second part, the application of \mathcal{H}^{-1} , reproduces the original data $\{f_j\}$. The constant C_M is a scale factor which ensures that the constant vector C_M and the vectors $\{C_M 2^{n/2} \psi(2^n x_j - k)\}$ are unit vectors in \mathbf{R}^M , using the standard inner product. Consequently, $C_M = \sqrt{1/M}$.

There are many ways of forming partial sums of discretized Haar series. The simplest ones con-

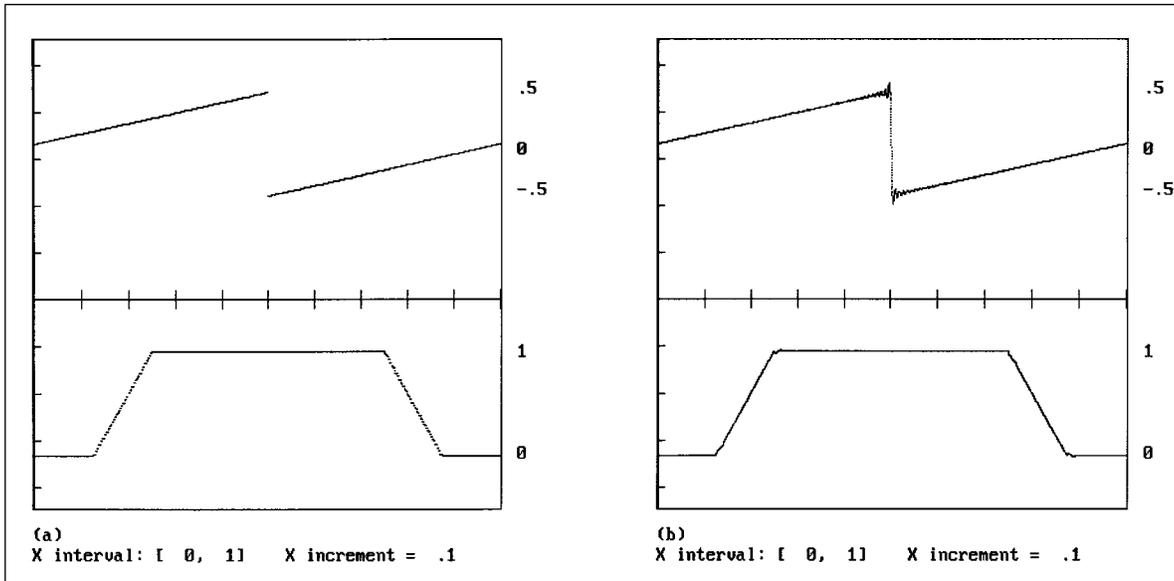


Figure 3. (a)[top] Haar series partial sum, 229 terms. (b)[top] Fourier series partial sum, 229 terms. (a)[bottom] Haar series partial sum, 92 terms. (b)[bottom] Daub4 series partial sum, 22 terms.

sist of multiplying the data by \mathcal{H} , then setting some of the resulting coefficients equal to 0, and then multiplying by \mathcal{H}^{-1} . A widely used method involves specifying a *threshold*. All coefficients whose magnitudes lie below this threshold are set equal to 0. This method is frequently used for noise removal, where coefficients whose magnitudes are significant only because of the added noise will often lie below a well-chosen threshold. We shall give an example of this in the section “Signal Denoising”. A second method keeps only the largest magnitude coefficients, while setting the rest equal to 0. This method is convenient for making comparisons when it is known in advance how many terms are needed. We used it in the compression example in the section “Signal Compression”. A third method, which we shall call the *energy method*, involves specifying a fraction of the signal’s energy, where the energy is the square root of the sum of the squares of the coefficients.³ We then retain the least number of the largest magnitude coefficients whose energy exceeds this fraction of the signal’s energy and set all other coefficients equal to 0. The energy method is useful for theoretical purposes: it is clearly helpful to be able to specify in advance what fraction of the signal’s energy is contained in a partial sum. We shall use the energy method frequently below.

Let us look at an example. Suppose our signal is

$$\{f_j = F(j/8192)\}_{j=0}^{8191}$$

where $F(x) = x\mathbf{1}_{[0,.5)}(x) + (x-1)\mathbf{1}_{[.5,1)}(x)$. In Figure 3(a)[top] we show a Haar series partial sum,

created by the energy method, which contains 99.5% of the energy of this signal. This partial sum, which used 229 coefficients out of a possible 8192, provides an acceptable visual representation of the signal. In fact, the sum of the squares of the errors is 2.0×10^{-3} . By comparison a 229 coefficient Fourier series partial sum suffers from serious drawbacks (see Figure 3(b)[top]). The sum of the squares of the errors is 4.5×10^{-1} , and there is severe oscillation and a Gibbs’ effect near $x = 0.5$. Although these latter two defects could be ameliorated using other summation methods ([11], Ch. 4), there would still be a significant deviation from the original signal (especially near $x = 0.5$).

This example illustrates how wavelet analysis homes right in on regions of high variability of signals and that Fourier methods try to smooth them out. The size of a function’s Fourier coefficients is related to the frequency content of the function, which is measured by integration of the function against completely unlocalized basis functions. For a function having a discontinuity, or some type of transient behavior, this produces Fourier coefficients that decrease in magnitude at a very slow rate. Consequently, a large number of Fourier coefficients are needed to accurately represent such signals.⁴ Wavelet series, however, use compactly supported basis functions which, at increasing levels of resolution, have rapidly decreasing supports and can zoom in on transient behavior. The transient behaviors contribute to the magnitude of only a small portion of the wavelet coefficients.

³The Haar transform is orthogonal, so it makes sense to specify energy in this way.

⁴In recent years, though, significant improvements have been achieved using local cosine bases [3, 7, 9, 5, 1].

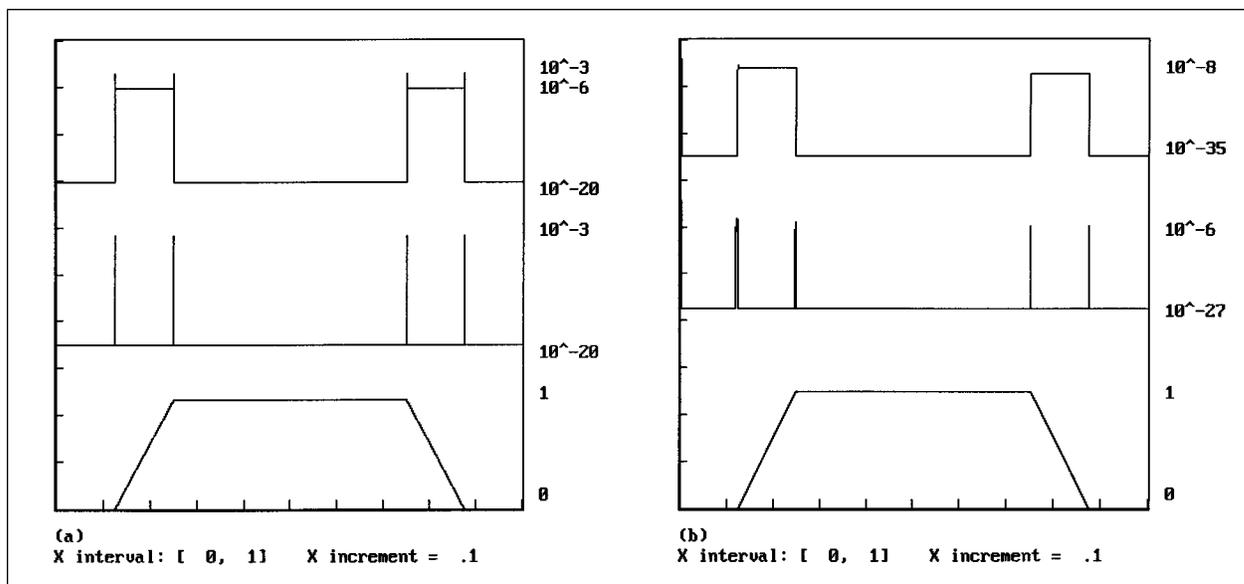


Figure 4. (a) Magnitudes of highest level coefficients for a function F: [top] Haar coefficients, [middle] Daub4 coefficients; [bottom] Graph of F. (b) Sums of squares of all coefficients: [top] Haar coefficients, [middle] Daub4 coefficients; [bottom] Graph of F. Vertical scales for highest level coefficients and sums of squares are logarithmic.

Consequently, a small number of wavelet coefficients are needed to accurately represent such signals.

The Haar system performs well when the signal is constant over long stretches. This is because the Haar wavelet is supported on $[0, 1)$ and satisfies a 0th order moment condition, $\int_{-\infty}^{\infty} \psi(x) dx = 0$. Therefore, if the signal $\{f_j\}$ is constant over an interval $a \leq x_j < b$ such that $[k2^{-n}, (k+1)2^{-n}] \subset [a, b)$, then the wavelet coefficient β_k^n equals 0. For example, suppose $\{f_j = F(j/8192)\}_{j=0}^{8191}$ where

$$F(x) = (8x - 1)\mathbf{1}_{[.125, .25)}(x) + \mathbf{1}_{[.25, .75)}(x) + (7 - 8x)\mathbf{1}_{[.75, .875)}(x).$$

In Figure 3(a)[bottom] we show a Haar series partial sum which contains 99.5% of the energy of this signal and uses only 92 coefficients out of a possible 8192. The fact that the signal is constant over three large subintervals of $[0, 1)$ accounts for the excellent compression in this example. In order to obtain wavelet bases that provide considerably more compression, we need a compactly supported wavelet $\psi(x)$ which has more moments equal to zero. That is, we want

$$(9) \quad \int_{-\infty}^{\infty} x^j \psi(x) dx = 0, \text{ for } j = 0, 1, \dots, L - 1$$

for an integer $L \geq 2$. We say that such a wavelet has its first L moments equal to zero. For example, a Daub4 wavelet has its first 2 moments equal to zero. Using a Daub4 wavelet series for the signal above, it is possible to capture 99.5% of the energy using only 22 coefficients! See Figure 3(b)[bottom]. This improvement in compression is due to the fact that the

Daub4 wavelet is supported on $[0, 3)$ and satisfies $\int_{-\infty}^{\infty} \psi(x) dx = 0$ and $\int_{-\infty}^{\infty} x\psi(x) dx = 0$. Consequently, if the signal is constant or linear over an interval $[a, b)$ which contains $[k2^{-n}, 3(k+1)2^{-n})$, then the wavelet coefficient β_k^n will equal 0. In Figure 4(a) we show graphs of the magnitudes of the highest level Haar coefficients and Daub4 coefficients. Each magnitude $|\beta_k^{12}|$ is plotted at the x -coordinate $k2^{-12}$ for $k = 0, 1, \dots, 2^{12} - 1$. These graphs show that the highest level Haar coefficients are near 0 over the constant parts of F , while the highest level Daub4 coefficients are near 0 over the constant and linear parts of F . In Figure 4(b) we show graphs of the sums of the squares of all the coefficients, which show that almost all the Daub4 coefficients are near 0 over the constant and linear parts of F , while the Haar coefficients are near 0 only over the constant parts of F . Furthermore, the largest magnitude Daub4 coefficients are concentrated around the locations of the points of nondifferentiability of F . This kind of local analysis illustrates one of the powerful features of wavelet analysis.

Looking again at Figure 3(b)[bottom], we see that the most serious defects of the Daub4 compressed signal are near the points where F is nondifferentiable. If, however, we consider the interval $[0.4, 0.6]$ where F is constant, the Daub4 compressed signal values differ from the values of F by no more than 1.2×10^{-15} at all of the 1641 discrete values of x in this interval. In contrast, a Fourier series partial sum using 23 coefficients differs by more than 10^{-3} at 1441 of these 1641 values of x . The Fourier series partial sum exhibits oscillations of amplitude 6.5×10^{-3} around the value 1 over this subin-

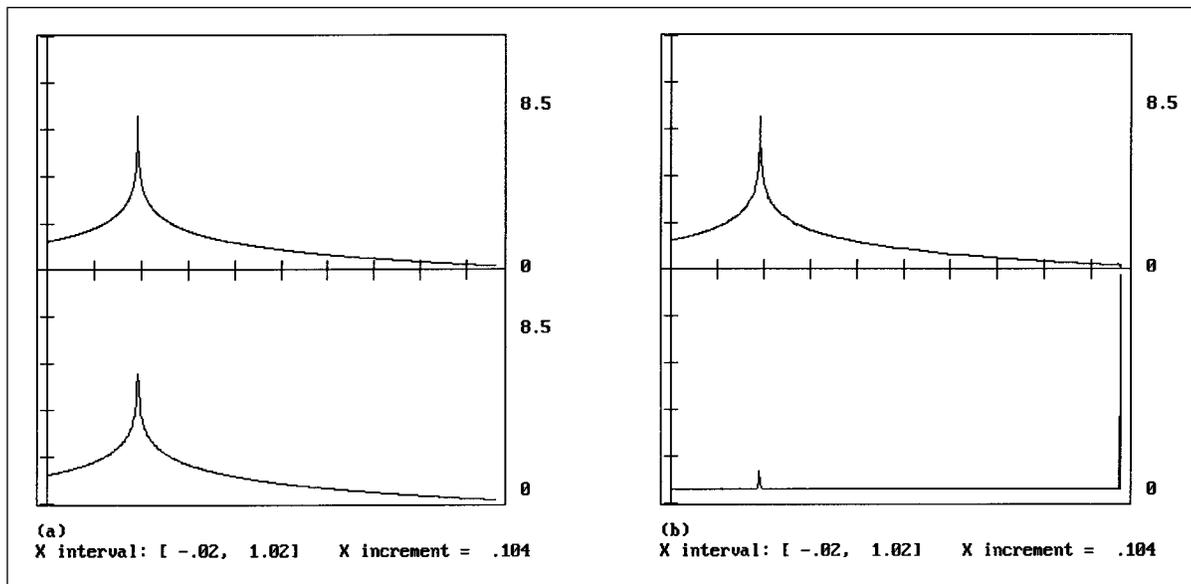


Figure 5. (a)[top] Signal. (b)[top] 37-term Daub4 approximation. (a)[bottom] 257-term Fourier cosine series approximation. (b)[bottom] Highest level Daub4 wavelet coefficients of signal.

terval. Because the Fourier coefficients for F are only $O(n^{-2})$, using just the first 23 coefficients produces an oscillatory approximation to F over all of $[0, 1)$, including the subinterval $[0.4, 0.6]$. The highest magnitude wavelet coefficients, however, are concentrated at the corner points for F , and their terms affect only a small portion of the partial sum (since their basis functions are compactly supported). Consequently, the wavelet series provides an extremely close approximation of F over the subinterval $[0.4, 0.6]$.

A major defect of the Haar wavelet is its discontinuity. For one thing, it is unsatisfying to use discontinuous functions to approximate continuous ones. Even with discrete signals there can be undesirable jumps in Haar series partial sum values (see Figure 3(a)[bottom]). Therefore, we want to have a wavelet that is continuous. In the next section we will describe the Daubechies wavelets, which have their first $L \geq 2$ moments equal to zero and are continuous.

Daubechies Wavelets

It is possible to generalize the construction of the Haar wavelet so as to obtain a continuous scaling function $\phi(x)$ and a continuous wavelet $\psi(x)$. Moreover, Daubechies has shown how to make them compactly supported. We will briefly sketch the main ideas; more details can be found in [4, 5, 8, 10].

Generalizing from the case of the Haar wavelets, we require that $\phi(x)$ and $\psi(x)$ satisfy

$$(10) \quad \int_{-\infty}^{\infty} \phi(x) dx = 1, \quad \int_{-\infty}^{\infty} |\phi(x)|^2 dx = 1, \\ \int_{-\infty}^{\infty} |\psi(x)|^2 dx = 1.$$

The MRA axioms tell us that $\phi(x)$ must generate a subspace V_0 and that $V_0 \subset V_1$. Therefore,

$$(11) \quad \phi(x) = \sum_{k \in \mathbb{Z}} c_k \sqrt{2} \phi(2x - k)$$

for some constants $\{c_k\}$. A wavelet $\psi(x)$, for which $\{\psi(x - k)\}$ spans the wavelet subspace W_0 , can be defined by⁵

$$(12) \quad \psi(x) = \sum_{k \in \mathbb{Z}} (-1)^k c_{1-k} \sqrt{2} \phi(2x - k).$$

Equations (11) and (12) generalize the equations in (2) for the Haar case. The orthogonality of ϕ and ψ leads to the following equation

$$(13) \quad \sum_{k \in \mathbb{Z}} (-1)^k c_{1-k} c_k = 0.$$

This equation, and the second equation in (14) below, imply the orthogonality of the matrices, W_N , used in the fast wavelet transform which we shall discuss later in this section.

Combining (11) with the first two integrals in (10), it follows that

$$(14) \quad \sum_{k \in \mathbb{Z}} c_k = \sqrt{2}, \quad \sum_{k \in \mathbb{Z}} |c_k|^2 = 1.$$

Similarly, assuming that $L = 2$, the equations in (9) combined with (12) imply

$$(15) \quad \sum_{k \in \mathbb{Z}} (-1)^k c_k = 0, \quad \sum_{k \in \mathbb{Z}} k(-1)^k c_k = 0.$$

⁵A simple proof, based on the MRA axioms, that $\{\psi(x - k)\}$ spans W_0 can be found in [10].

And, for $L > 2$, equations (9) and (12) yield additional equations similar to the ones in (15). There is a finite set of coefficients that solves the equations in (14) and (15), namely,

$$(16) \quad \begin{aligned} c_0 &= \frac{1 + \sqrt{3}}{4\sqrt{2}}, & c_1 &= \frac{3 + \sqrt{3}}{4\sqrt{2}}, \\ c_2 &= \frac{3 - \sqrt{3}}{4\sqrt{2}}, & c_3 &= \frac{1 - \sqrt{3}}{4\sqrt{2}} \end{aligned}$$

with all other $c_k = 0$. Using these values of c_k , the following iterative solution of (11)

$$(17) \quad \begin{aligned} \phi_0(x) &= \mathbf{1}_{[0,1)}(x), \\ \phi_n(x) &= \sum_{k \in \mathbb{Z}} c_k \sqrt{2} \phi_{n-1}(2x - k), \end{aligned} \quad \text{for } n \geq 1,$$

converges to a continuous function $\phi(x)$ supported on $[0, 3]$. It then follows from (12) that the wavelet $\psi(x)$ is also continuous and compactly supported on $[0, 3]$. This wavelet ψ we have been referring to as the Daub4 wavelet. The set of coefficients $\{c_k\}$ in (16) is the smallest set of coefficients that produce a continuous compactly supported scaling function. Other sets of coefficients, related to higher values of L , are given in [4] and [12].

Once the scaling function $\phi(x)$ and the wavelet function $\psi(x)$ have been found, then we proceed as we did above in the Haar case. We define the coefficients $\{\alpha_k^n\}$ and $\{\beta_k^n\}$ by the equations in (5), where now ϕ and ψ are the Daubechies scaling function and wavelet, respectively. The scaling identity (11) and the wavelet definition (12) yield the following coefficient relations:

$$(18) \quad \begin{aligned} \alpha_k^n &= \sum_{m \in \mathbb{Z}} c_m \alpha_{m+2k}^{n+1}, \\ \beta_k^n &= \sum_{m \in \mathbb{Z}} (-1)^m c_{1-m} \alpha_{m+2k}^{n+1}. \end{aligned}$$

In order to perform calculations in $L^2[0, 1)$, we define the *periodized wavelet* $\tilde{\psi}_{n,k}$ and the *periodized scaling function* $\tilde{\phi}_{n,k}$ by equations (3) and (4), only now using the Daubechies wavelet ψ and scaling function ϕ in place of the Haar wavelet and scaling function. Theorem 2 remains valid using these periodic wavelets, but the proof is more involved (see section 4.5 of [5] or section 3.11 of [8]). Therefore, for each $f \in L^2[0, 1)$ we can write

$$(19) \quad f(x) = \tilde{\alpha}_0^0 + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \tilde{\beta}_k^n \tilde{\psi}_{n,k}(x),$$

where $\tilde{\alpha}_0^0 = \int_0^1 f(x) dx$ and $\tilde{\beta}_k^n = \int_0^1 f(x) \tilde{\psi}_{n,k}(x) dx$. We also define the coefficients

$\tilde{\alpha}_k^n$ by $\tilde{\alpha}_k^n = \int_0^1 f(x) \tilde{\phi}_{n,k}(x) dx$. And, we periodically extend f with period 1, also denoting this periodic extension by f . Then, for $n \geq 0$ and $k = 0, 1, \dots, 2^n - 1$, we have

$$(20) \quad \begin{aligned} \tilde{\alpha}_k^n &= \int_0^1 f(x) 2^{n/2} \sum_{j \in \mathbb{Z}} \phi(2^n(x+j) - k) dx \\ &= \sum_{j \in \mathbb{Z}} \int_j^{j+1} f(x-j) 2^{n/2} \phi(2^n x - k) dx \\ &= \alpha_k^n. \end{aligned}$$

Similar arguments show that $\tilde{\beta}_k^n = \beta_k^n$ and $\tilde{\alpha}_{k+2^n}^n = \tilde{\alpha}_k^n$ and $\tilde{\beta}_{k+2^n}^n = \tilde{\beta}_k^n$ for $n \geq 0$ and $k = 0, 1, \dots, 2^n - 1$. After periodizing (11) and (12), it follows that

$$(21) \quad \begin{aligned} \tilde{\alpha}_k^n &= \sum_{m \in \mathbb{Z}} c_m \tilde{\alpha}_{m+2k}^{n+1}, \\ \tilde{\beta}_k^n &= \sum_{m \in \mathbb{Z}} (-1)^m c_{1-m} \tilde{\alpha}_{m+2k}^{n+1}. \end{aligned}$$

Remark. In the section “The Haar Wavelet” we saw, for the Haar wavelet ψ , that $\tilde{\psi}_{n,k}(x) = 2^{n/2} \psi(2^n x - k)$ for $n \geq 0$ and $k = 0, 1, \dots, 2^n - 1$. Almost exactly the same result holds for the Daubechies wavelets. For instance, if ψ is the Daub4 wavelet, then ψ is supported on $[0, 3]$. It follows, for $n \geq 2$, that on the unit interval $\tilde{\psi}_{n,0}$ is supported on $[0, 3 \cdot 2^{-n}]$. On the unit interval, we then have $\tilde{\psi}_{n,k}(x) = 2^{n/2} \psi(2^n x - k)$ for $k = 0, 1, \dots, 2^n - 3$. Hence, for $n \geq 2$, the periodized Daub4 wavelets $\tilde{\psi}_{n,k}$ are identical in $L^2[0, 1)$ with the wavelet functions $2^{n/2} \psi(2^n x - k)$, except when $k = 2^n - 2$ and $2^n - 1$. Similar results hold for all the Daubechies wavelets.

We can discretize the series in (19) by analogy with the Haar series. The coefficient relations in (21) yield a *fast wavelet transform*, \mathcal{W} , an orthogonal matrix defined by

$$\mathcal{W} = (W_2 \oplus I_{M-2}) \cdots (P_{M/2} \oplus I_{M/2})(W_{M/2} \oplus I_{M/2}) P_M W_M$$

where each matrix W_N is an $N \times N$ orthogonal matrix (as follows from (13) and the second equation in (14)). The matrix W_N is used to produce the $(N-1)$ st level coefficients $\{\tilde{\alpha}_k^{N-1}\}$ and $\{\tilde{\beta}_k^{N-1}\}$ from the N th level scaling coefficients $\{\tilde{\alpha}_k^N\}$ as follows:

$$\begin{aligned} W_N \left[\tilde{\alpha}_0^N, \dots, \tilde{\alpha}_{2^N-1}^N \right]^T \\ = \left[\tilde{\alpha}_0^{N-1}, \tilde{\beta}_0^{N-1}, \dots, \tilde{\alpha}_{2^{N-1}-1}^{N-1}, \tilde{\beta}_{2^{N-1}-1}^{N-1} \right]^T. \end{aligned}$$

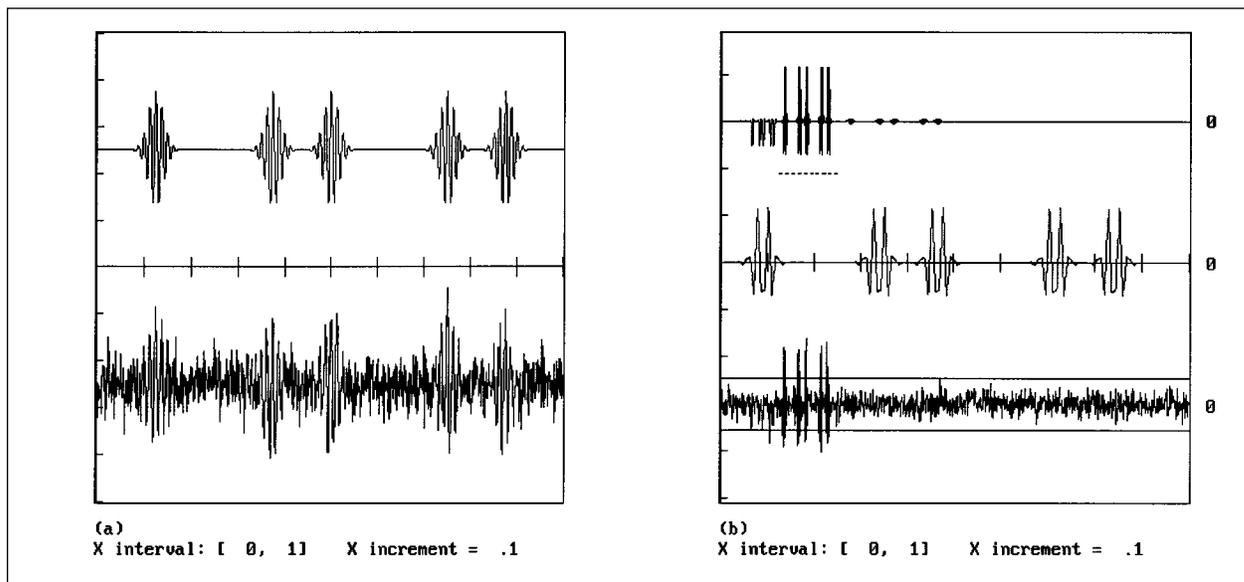


Figure 6. (a)[top] Signal. (b)[top] Signal's coiflet30 transform, 7th level coefficients lie above the dotted line. (b)[middle] 7th level coefficients. (a)[bottom] Signal with added noise. (b)[bottom] Noisy signal's coiflet30 transform. The horizontal lines are thresholds equal to ± 0.15 .

the highest level scaling coefficients, beyond the argument we gave above for Daubechies wavelets. The construction of coiflets was first carried out by Daubechies and named after Coifman (who first suggested them).

Compression of Signals

One of the most important applications of wavelet analysis is to the compression of signals. As an example, let us use a Daub4 series to compress the signal $\{f_j = F(j/1024)\}_{j=0}^{1023}$ where $F(x) = -\log|x - 0.2|$. See Figure 5(a)[top]. For this signal, a partial sum containing 99% of the energy required only 37 coefficients (see Figure 5(b)[top]). It certainly provides a visually acceptable approximation of $\{f_j\}$. In particular, the sharp maximum in the signal near $x = 0.2$ seems to be reproduced quite well. The compression ratio is $1024:37 \approx 27:1$, which is an excellent result considering that we also have 99% accuracy. In addition, wavelet analysis has identified the singularities of F . Notice in Figure 5(b)[bottom] the peak in the wavelet coefficients is near $x = 0.2$, where F has a singularity, and the largest wavelet coefficient is near $x = 1$, where the periodic extension of F has a jump discontinuity.

Turning to Fourier series, since the even periodic extension is continuous, we used a discrete Fourier cosine series to compress this signal. In Figure 5(a)[bottom] we show a 257-term discrete Fourier cosine series partial sum for $\{f_j\}$. Even using seven times as many coefficients as the wavelet series, the cosine series cannot reproduce the sharp peak in the signal. Better results could be obtained in this case by either seg-

menting the interval and performing a cosine expansion on each segment, or by using a smoother version of the same idea involving *local cosine bases* [3, 9, 12, 1, 5].

One way to quantify the accuracy of these approximations is to use relative R.M.S. differences. Given two sets of data $\{f_j\}_{j=0}^{M-1}$ and $\{g_j\}_{j=0}^{M-1}$, their *relative R.M.S. difference*, relative to $\{f_j\}$, is defined by

$$\mathcal{D}(f, g) = \sqrt{\frac{\sum_{j=0}^{M-1} |f_j - g_j|^2}{\sum_{j=0}^{M-1} |f_j|^2}}$$

For the example above, if we denote the wavelet approximation by f^w , then $\mathcal{D}(f, f^w) = 9.8 \times 10^{-3}$. For the Fourier cosine series approximation, call it f^c , we have $\mathcal{D}(f, f^c) = 2.7 \times 10^{-2}$. A *rule of thumb* for a visually acceptable approximation is to have a relative R.M.S. difference of less than 10^{-2} . The approximations in this example are consistent with this rule of thumb.

We can also do more localized analysis with R.M.S. differences. For example, over the subinterval $[.075, .325]$ centered on the singularity of F , we find that $\mathcal{D}(f, f^w) = 9.7 \times 10^{-3}$ and $\mathcal{D}(f, f^c) = 3.2 \times 10^{-2}$. These numbers confirm our visual impression that the wavelet series does a better job reproducing the sharp peak in the signal. Or, using the subinterval $[.25, .75]$, we get $\mathcal{D}(f, f^w) = 1.0 \times 10^{-2}$ and $\mathcal{D}(f, f^c) = 3.3 \times 10^{-3}$, confirming our impression that both series do an adequate job approximating $\{f_j\}$ over this subinterval.

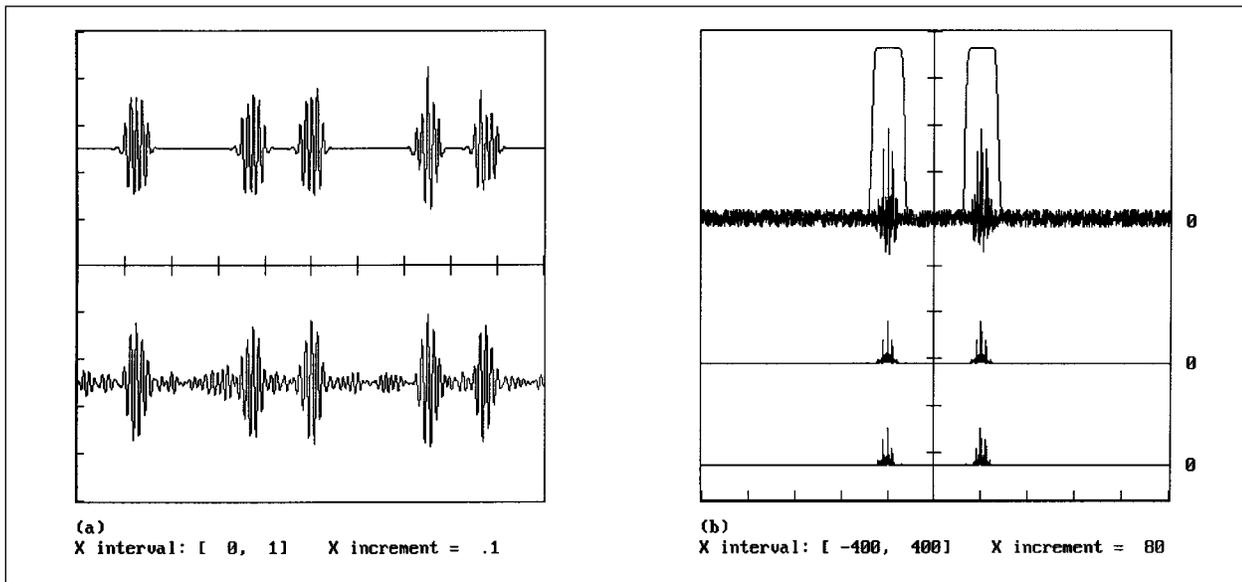


Figure 7. (a)[top] Denoised signal using wavelet analysis. (a)[bottom] Denoised signal using Fourier analysis. (b)[top] Fourier coefficients of noisy signal and filter function. (b)[middle] Moduli-squared of Fourier coefficients of original signal. (b)[bottom] Moduli-squared of Fourier coefficients of wavelet denoised signal.

Although in the examples we have discussed so far Fourier analysis did not compress the signals very well, we do not wish to create the impression that this will always be true. In fact, if a signal is composed of relatively few sinusoids, then Fourier analysis will provide very good compression. For example, consider the signal $\{f_j = f(j/1024)\}_{j=0}^{1023}$ where $f(x)$ is defined in (1) with $\nu = 280$. The Fourier coefficients for f are graphed in Figure 1(b)[top]. They tend rapidly to 0 away from the frequencies ± 280 ; hence the signal is composed of relatively few sinusoids. By computing a Fourier series partial sum that uses only the 122 Fourier coefficients whose frequencies are within ± 30 of ± 280 , we obtained a signal g that was visually indistinguishable from the original signal. In fact, $\mathcal{D}(f, g) = 5.1 \times 10^{-3}$. However, by compressing $\{f_j\}$ with the largest 122 Daub4 wavelet coefficients, we obtained $\mathcal{D}(f, f^w) = 2.7 \times 10^{-1}$ and the compressed signal f^w was only a crude approximation of the original signal. The reason that compactly supported wavelets perform poorly in this case is that the large number of rapid oscillations in the signal produce a correspondingly large number of high magnitude wavelet coefficients at the highest levels. Consequently, a significant fraction of all the wavelet coefficients are of high magnitude, so it is not possible to significantly compress the signal using compactly supported wavelets. This example illustrates that wavelet analysis is not a panacea for the problem of signal compression. In fact, much work has been done in creating large collections of wavelet bases and Fourier bases and choosing for each signal a basis which best compresses it [12, 9, 3, 5].

Signal Denoising

Wavelet analysis can also be used for removing noise from signals. As an example, we show in Figure 6(a)[top] a discrete signal $\{f(j/1024)\}_{j=0}^{1023}$ where $f(x)$ is defined by formula (1) with $\nu = 80$. Each term of the form

$$(5 \cos 2\pi\nu x)e^{-640\pi(x-k/8)^2}$$

we shall refer to as a *blip*. Notice that each blip is concentrated around $x = k/8$, since $e^{-640\pi(x-k/8)^2}$ rapidly decreases to 0 away from $x = k/8$. This signal can be interpreted as representing the bit sequence 1011011. In Figure 6(a)[bottom] we show this signal after it has been corrupted by adding noise. In Figure 6(b)[top] we show the *coiflet30* wavelet coefficients for the original signal. The rationale for using wavelets to remove the noise is that the original signal's wavelet coefficients are closely correlated with the points near $x = k/8$ where the blips are concentrated. To demonstrate this, we show in Figure 6(b)[middle] a graph of the 7th level wavelet coefficients $\{\tilde{\beta}_k^7\}$ corresponding to the points $\{k2^{-7}\}_{k=0}^{2^7-1}$ on the unit interval. Comparing this to Figure 6(a)[top], we can see that the positions of this level's largest magnitude wavelet coefficients are closely correlated with the positions of the blips. Similar graphs could also be drawn for other levels, but the 7th level coefficients have the largest magnitude. In Figure 6(b)[bottom] we show the *coiflet30* transform of the noisy signal. In spite of the noise, the 7th level coefficients clearly stand out, although in a distorted form. By introducing a threshold, in this case 0.15, we can retain these 7th level coefficients and remove

most of the noise. In Figure 7(a)[top] we show the reconstructed signal obtained by computing a partial sum using only those coefficients whose magnitudes do not fall below 0.15. This reconstruction is not a flawless reproduction of the original signal, but nevertheless the amount of noise has been greatly reduced, and the bit sequence 1011011 can be determined.

In Figure 7(a)[bottom] we show the denoised signal obtained by filtering the Fourier coefficients of the noisy signal (see Figure 7(b)[top]) using the method of denoising described in the section “Frequency Information, Denoising”. In contrast to the wavelet denoising, the Fourier denoising has retained a significant amount of noise in the spaces between the blips. The source of this retained noise is that most of the original noise’s Fourier coefficients are of uniform magnitude distributed across all frequencies. Consequently, the filter preserves noise coefficients corresponding to frequencies that were not present in the original signal. These coefficients generate sinusoids that oscillate across the entire interval $[0, 1]$. The noise’s wavelet coefficients also have almost uniform magnitude, but the thresholding process eliminates them all, except the ones modifying the 7th level coefficients of the original signal. Since these coefficients’ wavelet basis functions are compactly supported, this causes distortions in the recovered signal that are limited to neighborhoods of the positions of the 7th level coefficients. Consequently, there is still noise distorting the blips, but very little noise in between them.

It is also interesting to observe that the wavelet reconstructed signal and the original signal have similar frequency content. In Figure 7(b)[middle] and Figure 7(b)[bottom], we have graphed the moduli-squared of the Fourier coefficients of the original signal and of the wavelet denoised signal, respectively. These graphs show that the frequencies of the wavelet reconstruction are, like the frequencies of the original signal, concentrated around ± 80 with the highest magnitude frequencies located precisely at ± 80 . This shows that the *coiflet30* wavelet has the ability to extract frequency information. Much work has been done in refining this ability, including the development of another class of bases called wavelet packets [12, 9, 3, 5].

Conclusion

In this paper we have tried to show how the two methodologies of Fourier analysis and wavelet analysis are used for various kinds of work. Of course, we have only scratched the surface of both fields. Much more information can be found in the references and their bibliographies.

References

- [1] P. AUSCHER, G. WEISS, and M. V. WICKERHAUSER, *Local sine and cosine basis of Coifman and Meyer and the construction of smooth wavelets*, Wavelets: A Tutorial in Theory and Applications (C. K. Chui, ed.), Academic Press, 1992.
- [2] W. BRIGGS and V. E. HENSON, *The DFT. An owner’s manual for the Discrete Fourier Transform*, SIAM, 1995.
- [3] R. COIFMAN and M. V. WICKERHAUSER, *Wavelets and adapted waveform analysis*, Wavelets: Mathematics and Applications (J. Benedetto and M. Frazier, eds.), CRC Press, 1994.
- [4] I. DAUBECHIES, *Ten lectures on wavelets*, SIAM, 1992.
- [5] E. HERNANDEZ and G. WEISS, *A first course on wavelets*, CRC Press, 1996.
- [6] S. MALLAT, *Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbf{R})$* , Trans. Amer. Math. Soc. **315** (1989), 69–87.
- [7] H. MALVAR, *Signal processing with lapped transforms*, Artech House, 1992.
- [8] Y. MEYER, *Wavelets and operators*, Cambridge Univ. Press, 1992.
- [9] ———, *Wavelets: Algorithms and applications*, SIAM, 1993.
- [10] R. STRICHARTZ, *How to make wavelets*, MAA Monthly **100**, no. 6 (June–July 1993).
- [11] J. WALKER, *Fast Fourier transforms*, second edition, CRC Press, 1996.
- [12] M. V. WICKERHAUSER, *Adapted wavelet analysis from theory to software*, IEEE and A. K. Peters, 1994.