Kai Schneider and Marie Farge, 2006

Wavelets: Mathematical Theory

Encyclopedia of Mathematical Physics
Eds. J.P. Françoise, G. Naber and T.S. Tsun,
Elsevier, 426-438

of wavelets is the Malvar basis which is also a generalization of local Fourier basis, and gives a perfect reconstruction. A new direction of wavelet is the second-generation wavelets which are constructed by lifting scheme and free from the regular dyadic procedure, and thus applicable to compact regions as S^2 and a finite interval.

See also: Fractal Dimensions in Dynamics; Image Processing: Mathematics; Intermittency in Turbulence; Wavelets: Application to Turbulence; Wavelets: Mathematical Theory.

Further Reading

Benedetto JJ and Frazier W (eds.) (1994) *Wavelets: Mathematics and Applications*. Boca Raton, FL: CRC Press.

van den Berg JC (ed.) (1999) *Wavelets in Physics*. Cambridge: Cambridge University Press.

Daubechies I (1992) *Ten Lectures on Wavelets*, SIAM, CBMS61, Philadelphia.

Mallat S (1998) *A Wavelet Tour of Signal Processing*. San Diego: Academic Press.

Strang G and Nguyen T (1997) *Wavelet and Filter Banks*. Wellesley: Wellesley-Cambridge Press.

Wavelets: Mathematical Theory

K Schneider, Université de Provence, Marseille, France
M Farge, Ecole Normale Supérieure, Paris, France

Introduction

The wavelet transform unfolds functions into time (or space) and scale, and possibly directions. The continuous wavelet transform has been discovered by Alex Grossmann and Jean Morlet who published the first paper on wavelets in 1984. This mathematical technique, based on group theory and square-integrable representations, allows us to decompose a signal, or a field, into both space and scale, and possibly directions. The orthogonal wavelet transform has been discovered by Lemarié and Meyer (1986). Then, Daubechies (1988) found orthogonal bases made of compactly supported wavelets, and Mallat (1989) designed the fast wavelet transform (FWT) algorithm. Further developments were done in 1991 by Raffy Coifman, Yves Meyer, and Victor Wickerhauser who introduced wavelet packets and applied them to data compression. The development of wavelets has been interdisciplinary, with contributions coming from very different fields such as engineering (sub-band coding, quadrature mirror filters, time–frequency analysis), theoretical physics (coherent states of affine groups in quantum mechanics), and mathematics (Calderon–Zygmund operators, characterization of function spaces, harmonic analysis). Many reference textbooks are available, some of them we recommend are listed in the "Further reading" section. Meanwhile, a large spectrum of applications has grown and is still developing, ranging from signal analysis and image processing via numerical analysis and turbulence modeling to data compression.

In this article, we will first define the continuous wavelet transform and then the orthogonal wavelet transform based on a multiresolution analysis. Properties of both transforms will be discussed and illustrated by examples. For a general introduction to wavelets, *see* Wavelets: Applications.

Continuous Wavelet Transform

Let us consider the Hilbert space of square-integrable functions $L^2(\mathbb{R}) = \{f : \|f\|_2 < \infty\}$, equipped with the scalar product $\langle f, g\rangle = \int_{\mathbb{R}} f(x) g^*(x)\, \mathrm{d}x$ (* denotes the complex conjugate in the case of complex-valued functions) and where the norm is defined by $\|f\|_2 = \langle f, f\rangle^{1/2}$.

Analyzing Wavelet

The starting point for the wavelet transform is to choose a real- or complex-valued function $\psi \in L^2(\mathbb{R})$, called the "mother wavelet," which fulfills the admissibility condition,

$$C_\psi = \int_0^\infty \left|\widehat{\psi}(k)\right|^2 \frac{\mathrm{d}k}{|k|} < \infty \qquad [1]$$

where

$$\widehat{\psi}(k) = \int_{-\infty}^{\infty} \psi(x)\, \mathrm{e}^{-\iota 2\pi k x}\, \mathrm{d}x \qquad [2]$$

denotes the Fourier transform, with $\iota = \sqrt{-1}$ and k the wave number. If ψ is integrable, that is, $\psi \in L^1(\mathbb{R})$, this implies that ψ has zero mean,

$$\int_{-\infty}^{\infty} \psi(x)\, \mathrm{d}x = 0 \qquad \text{or} \qquad \widehat{\psi}(0) = 0 \qquad [3]$$

In practice, however, one also requires the wavelet ψ to be well localized in both physical and Fourier

$$\int_{-\infty}^{\infty} x^m \psi(x)\,\mathrm{d}x = 0 \qquad \text{for } m = 0,\ M-1 \qquad [4]$$

that is, monomials up to degree $M-1$ are exactly reproduced. In Fourier space, this property is equivalent to

$$\frac{\mathrm{d}^m}{\mathrm{d}k^m}\widehat{\psi}(k)\mid_{k=0} = 0 \quad \text{for } m = 0, M-1 \qquad [5]$$

therefore, the Fourier transform of ψ decays smoothly at $k=0$.

Analysis

From the mother wavelet ψ, we generate a family of continuously translated and dilated wavelets,

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \quad \text{for } a > 0 \text{ and } b \in \mathbb{R} \qquad [6]$$

where a denotes the dilation parameter, corresponding to the width of the wavelet support, and b the translation parameter, corresponding to the position of the wavelet. The wavelets are normalized in energy norm, that is, $\|\psi_{a,b}\|_2 = 1$.

In Fourier space, eqn [6] reads

$$\widehat{\psi}_{a,b(k)} = \sqrt{a}\widehat{\psi}(ak)\,\mathrm{e}^{-\iota 2\pi k b} \qquad [7]$$

where the contraction with $1/a$ in [6] is reflected in a dilation by a [7] and the translation by b implies a rotation in the complex plane.

The continuous wavelet transform of a function f is then defined as the convolution of f with the wavelet family $\psi_{a,b}$:

$$\widetilde{f}(a,b) = \int_{-\infty}^{\infty} f(x)\psi^*_{a,b}(x)\,\mathrm{d}x \qquad [8]$$

where $\psi^*_{a,b}$ denotes, in the case of complex-valued wavelets, the complex conjugate.

Using Parseval's identity, we get

$$\widetilde{f}(a,b) = \int_{-\infty}^{\infty} \widehat{f}(k)\widehat{\psi}^*_{a,b}(k)\,\mathrm{d}k \qquad [9]$$

and the wavelet transform could be interpreted as a frequency decomposition using bandpass filters $\widehat{\psi}_{a,b}$ centered at frequencies $k = k_\psi/a$. The wave number k_ψ denotes the barycenter of the wavelet support in Fourier space

$$k_\psi = \frac{\int_0^\infty k|\widehat{\psi}(k)|\,\mathrm{d}k}{\int_0^\infty |\widehat{\psi}(k)|\,\mathrm{d}k} \qquad [10]$$

Note that these filters have a variable width $\Delta k/k$; therefore, when the wave number increases, the bandwidth becomes wider.

Synthesis

The admissibility condition [1] implies the existence of a finite energy reproducing kernel, which is a necessary condition for being able to reconstruct the function f from its wavelet coefficients $\widetilde{f}$. One then recovers

$$f(x) = \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \widetilde{f}(a,b)\psi_{a,b}(x)\frac{\mathrm{d}a\mathrm{d}b}{a^2} \qquad [11]$$

which is the inverse wavelet transform.

The wavelet transform is an isometry and one has Parseval's identity. Therefore, the wavelet transform conserves the inner product and we obtain

$$\begin{aligned}\langle f, g\rangle &= \int_{-\infty}^{\infty} f(x)g^*(x)\,\mathrm{d}x \\ &= \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \widetilde{f}(a,b)\widetilde{g}^*(a,b)\frac{\mathrm{d}a\mathrm{d}b}{a^2}\end{aligned} \qquad [12]$$

As a consequence, the total energy E of a signal can be calculated either in physical space or in wavelet space, such as

$$\begin{aligned}E &= \int_{-\infty}^{\infty} |f(x)|^2\,\mathrm{d}x \\ &= \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty |\widetilde{f}(a,b)|^2\frac{\mathrm{d}a\mathrm{d}b}{a^2}\end{aligned} \qquad [13]$$

This formula is also the starting point for the definition of wavelet spectra and scalogram (*see* Wavelets: Application to Turbulence).

Examples

In the following, we apply the continuous wavelet transform to different academic signals using the Morlet wavelet. The Morlet wavelet is complex valued, and consists of a modulated Gaussian with width k_0/π:

$$\psi(x) = (\mathrm{e}^{2\iota\pi x} - \mathrm{e}^{-k_0^2/2})\,\mathrm{e}^{-2\pi^2x^2/k_0^2} \qquad [14]$$

The envelope factor k_0 controls the number of oscillations in the wave packet; typically, $k_0 = 5$ is used. The correction factor $\mathrm{e}^{-k_0^2/2}$, to ensure its vanishing mean, is very small and often neglected. The Fourier transform is

$$\widehat{\psi}(k) = \frac{k_0}{2\sqrt{\pi}}\,\mathrm{e}^{-(k_0^2/2)(1+k^2)}(\mathrm{e}^{-k_0^2k} - 1) \qquad [15]$$

Figure 1 shows wavelet analyses of a cosine, two sines, a Dirac, and a characteristic function. Below

the four signals we plot the modulus and the phase of the corresponding wavelet coefficients.

Higher Dimensions

The continuous wavelet transform can be extended to higher dimensions in $L^2(\mathbb{R}^n)$ in different ways. Either we define spherically symmetric wavelets by setting $\psi(x) = \psi^{1d}(|x|)$ for $x \in \mathbb{R}^n$ or we introduce in addition to dilations $a \in \mathbb{R}^+$ and translations $b \in \mathbb{R}^n$ also rotations to define wavelets with a directional sensitivity. In the two-dimensional case, we obtain for example,

$$\psi_{a,b,\theta}(x) = \frac{1}{a}\psi\left(R_\theta^{-1}\left(\frac{x-b}{a}\right)\right) \qquad [16]$$

where $a \in \mathbb{R}^+, b \in \mathbb{R}^2$, and where R_θ is the rotation matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad [17]$$

The analysis formula [8] then becomes

$$\widetilde{f}(a,b,\theta) = \int_{\mathbb{R}^2} f(x)\psi^*_{a,b,\theta}(x)\,\mathrm{d}x \qquad [18]$$

and for the corresponding inverse wavelet transform [11] we obtain

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \int_{\mathbb{R}^2}\int_0^{2\pi} \widetilde{f}(a,b,\theta)\psi_{a,b,\theta}(x)\frac{\mathrm{d}a\mathrm{d}b\mathrm{d}\theta}{a^3} \qquad [19]$$

Similar constructions can be made in dimensions larger than 2 using $n-1$ angles of rotation.

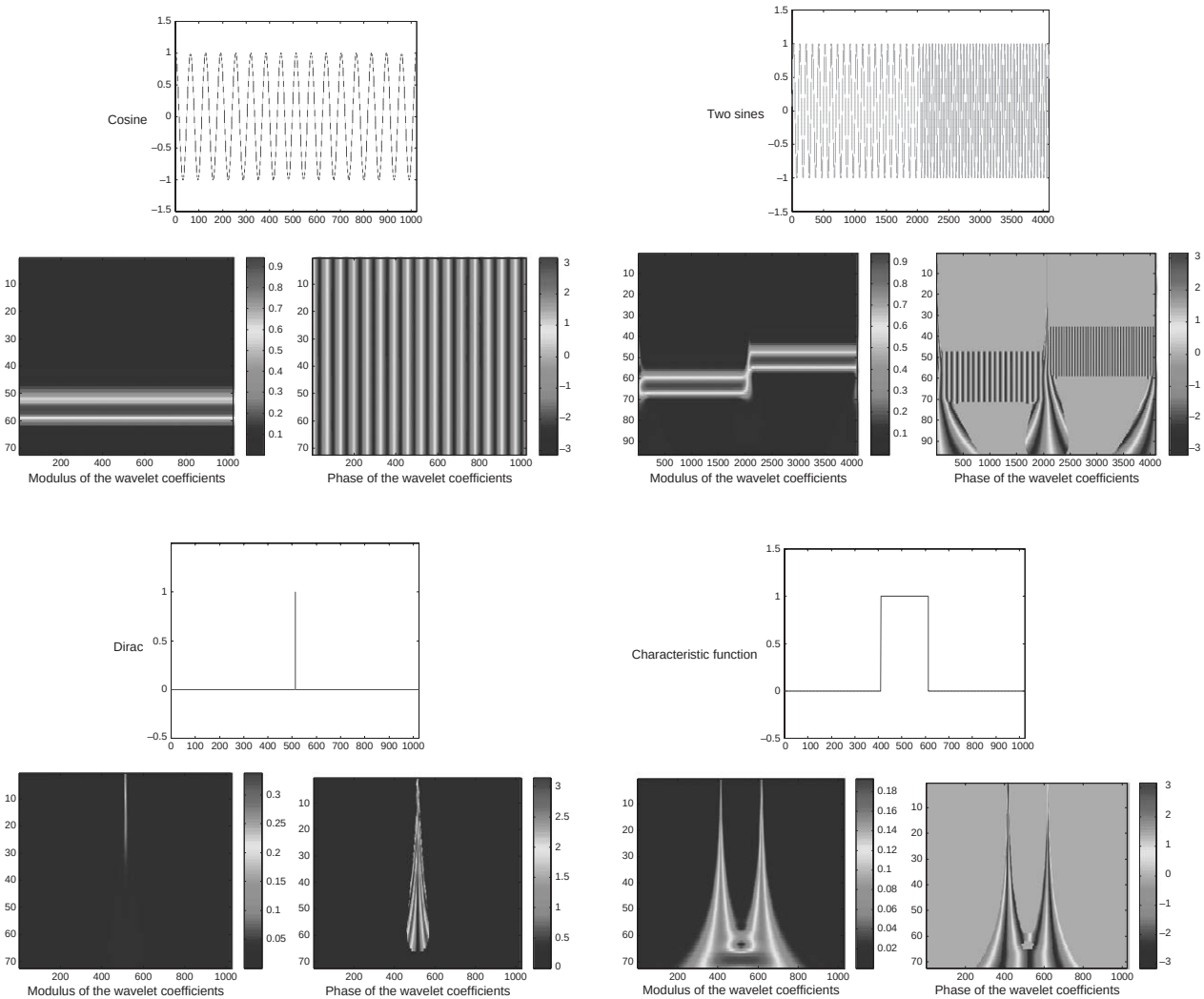

Figure 1 Examples of a one-dimensional continuous wavelet analysis using the complex-valued Morlet wavelet. Each subfigure shows on the top the function to be analyzed and below (left) the modulus of its wavelet coefficients and below (right) the phase of its wavelet coefficients.

Discrete Wavelets

Frames

It is possible to obtain a discrete set of quasiorthogonal wavelets by sampling the scale and position axes a, b. For the scale a we use a logarithmic discretization: a is replaced by $a_j = a_0^{-j}$, where a_0 is the sampling rate of the log a axis ($a_0 = \Delta(\log a)$) and where $j \in \mathbb{Z}$ is the scale index. The position b is discretized linearly: b is replaced by $x_{ji} = ib_0 a_0^{-j}$, where b_0 is the sampling rate of the position axis at the largest scale and where $i \in \mathbb{Z}$ is the position index. Note that the sampling rate of the position varies with scale, that is, for finer scales (increasing j and hence decreasing a_j), the sampling rate increases. Accordingly, we obtain the discrete wavelets (cf. Figure 2)

$$\psi_{ji}(x') = a_j^{-1/2}\psi\left(\frac{x' - x_{ji}}{a_j}\right) \qquad [20]$$

and the corresponding discrete decomposition formula is

$$\widetilde{f}_{ji} = \langle \psi_{ji}, f \rangle = \int_{-\infty}^{\infty} f(x')\psi_{ji}^*(x')\,\mathrm{d}x' \qquad [21]$$

Furthermore, the wavelet coefficients satisfy the following estimate:

$$A\|f\|_2^2 \le \sum_{j,i} |\widetilde{f}_{ji}|^2 \le B\|f\|_2^2 \qquad [22]$$

with frame bounds $B \ge A > 0$. In the case $A = B$ we have a tight frame.

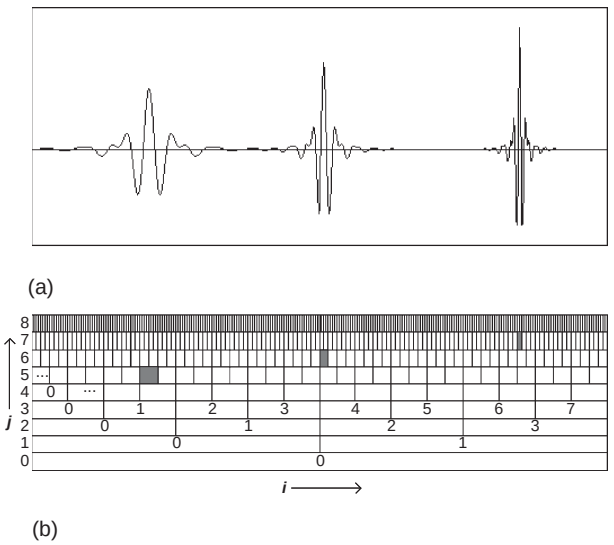

Figure 2 Orthogonal quintic spline wavelets $\psi_{j,i}(x) = 2^{j/2}\psi(2^j x - i)$ at different scales and positions: (a) $\psi_{5,6}(x)$, $\psi_{6,32}(x)$, $\psi_{7,108}(x)$, and (b) corresponding wavelet coefficients.

The discrete reconstruction formula is

$$f(x) = C\sum_{j=-\infty}^{\infty}\sum_{i=-\infty}^{\infty}\widetilde{f}_{ji}\psi_{ji}(x) + R(x) \qquad [23]$$

where C is a constant and $R(x)$ is a residual, both depending on the choice of the wavelet and the sampling of the scale and position axes. For the particular choice $a_0 = 2$ (which corresponds to a scale sampling by octaves) and $b_0 = 1$, we have the dyadic sampling, for which there exist special wavelets ψ_{ji} that form an orthonormal basis of $L^2(\mathbb{R})$, that is, such that

$$\langle \psi_{ji}, \psi_{j'i'} \rangle = \delta_{jj'}\delta_{ii'} \qquad [24]$$

where δ denotes the Kronecker symbol. This means that the wavelets ψ_{ji} are orthogonal with respect to their translates by discrete steps $2^{-j}i$ and their dilates by discrete steps 2^{-j} corresponding to octaves. In this case, the reconstruction formula is exact with $C = 1$ and $R = 0$. Note that the discrete wavelet transform has lost the invariance by translation and dilation of the continuous one.

Orthogonal Wavelets and Multiresolution Analysis

The construction of orthogonal wavelet bases and the associated fast numerical algorithm is based on the mathematical concept of multiresolution analysis (MRA). The underlying idea is to consider approximations f_j of the function f at different scales j. The amount of information needed to go from a coarse approximation f_j to a finer resolution approximation f_{j+1} is then described using orthogonal wavelets. The orthogonal wavelet analysis can thus be interpreted as decomposing the function into approximations of the function at coarser and coarser scales (i.e., for decreasing j), where the differences between the approximations are encoded using wavelets.

The definition of the MRA was introduced by Stéphane Mallat in 1988 (Mallat 1989). This technique constitutes a mathematical framework of orthogonal wavelets and the related FWT.

A one-dimensional orthogonal MRA of $L^2(\mathbb{R})$ is defined as a sequence of successive approximation spaces $V_j, j \in \mathbb{Z}$, which are closed imbedded subspaces of $L^2(\mathbb{R})$. They verify the following conditions:

$$V_j \subset V_{j+1} \quad \forall j \in \mathbb{Z} \qquad [25]$$

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}) \qquad [26]$$

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\} \qquad [27]$$

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1} \qquad [28]$$

A scaling function $\phi(x)$ is required to exist. Its translates generate a basis in each V_j, that is,

$$V_j V_j = \overline{\text{span}}\{\phi_{ji}\}_{i\in\mathbb{Z}} \quad [29]$$

where

$$\phi_{ji}(x) = 2^{j/2}\phi(2^j x - i), \quad j, i \in \mathbb{Z} \quad [30]$$

At a given scale j, this basis is orthonormal with respect to its translates by steps $i/2^j$ but not to its dilates,

$$\langle \phi_{ji}, \phi_{jk} \rangle = \delta_{ik} \quad [31]$$

The nestedness of the approximation spaces [28] generated by the scaling function ϕ implies that it satisfies a refinement equation:

$$\phi_{j-1,i}(x) = \sum_{n=-\infty}^{\infty} h_{n-2i}\phi_{jn}(x) \quad [32]$$

with the filter coefficients $h_n = \langle \phi_{jn}, \phi_{j-1,0} \rangle$, which determine the scaling function completely. In general, only the filter coefficients h_n are known and no analytical expression of ϕ is given. Equation [32] implies that the approximation of a function at coarser scale can be described by linear combinations of the same function at finer scales.

The orthogonal projection of a function $f \in L^2(\mathbb{R})$ on V_J is defined as

$$P_{V_J} : f \longrightarrow P_{V_J} f = f_J \quad [33]$$

with

$$f_J(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_{jk} \rangle \phi_{jk}(x) \quad [34]$$

This coarse graining at a given scale J is done by filtering the function with the scaling function ϕ. As a filter, the scaling function ϕ does not have vanishing mean but is normalized so that $\int_{-\infty}^{\infty} \phi(x)\,dx = 1$.

As V_{J-1} is included in V_J, we can define its orthogonal complement space in V_J:

$$V_J = V_{J-1} \oplus W_{J-1} \quad [35]$$

Correspondingly, the approximation of the function f at scale 2^{-J}, belonging to V_J, can be decomposed as a sum of orthogonal projections on V_{J-1} and W_{J-1}, such that

$$P_{V_J} f = P_{V_{J-1}} f + P_{W_{J-1}} f \quad [36]$$

Based on the scaling function ϕ, one can construct a function ψ, the so-called mother wavelet, given by the relation

$$\psi_{ji}(x) = \sum_{n\in\mathbb{Z}} g_{n-2i}\phi_{j,n}(x) \quad [37]$$

with $g_n = \langle \phi_{jn}, \psi_{j-1,0} \rangle$, and where $\psi_{ji}(x) = 2^{j/2}\psi(2^j x - i)$, $j, i \in \mathbb{Z}$ (cf. **Figure 2**). The filter coefficients g_n can be computed from the filter coefficients h_n using the relation

$$g_n = (-1)^{1-n} h_{1-n} \quad [38]$$

The translates and dilates of the wavelet ψ constitute orthonormal bases of the spaces W_j,

$$W_j = \overline{\text{span}}\{\psi_{ji}\}_{i\in\mathbb{Z}} \quad [39]$$

As in the continuous case, the wavelets have vanishing mean, and also possibly vanishing higher-order moments; therefore,

$$\int_{-\infty}^{\infty} x^m \psi(x)\,dx = 0 \quad \text{for } m = 0, \ldots, M-1 \quad [40]$$

Let us now consider approximations of a function $f \in L^2(\mathbb{R})$ at two different scales j:

- at scale j

$$f_j(x) = \sum_{i=-\infty}^{\infty} \overline{f}_{ji}\phi_{ji}(x) \quad [41]$$

- at scale $j-1$

$$f_{j-1}(x) = \sum_{i=-\infty}^{\infty} \overline{f}_{j-1,i}\phi_{j-1,i}(x) \quad [42]$$

with the scaling coefficients

$$\overline{f}_{ji} = \langle f, \phi_{ji} \rangle \quad [43]$$

which correspond to local averages of the function f at position $i2^{-j}$ and at scale 2^{-j}.

The difference between the two approximations is encoded by the wavelets

$$f_j(x) - f_{j-1}(x) = \sum_{i=-\infty}^{\infty} \widetilde{f}_{j-1,\,i}\psi_{j-1,i}(x) \quad [44]$$

with the wavelet coefficients

$$\widetilde{f}_{ji} = \langle f, \psi_{ji} \rangle \quad [45]$$

which correspond to local differences of the function at position $(2i+1)2^{-(j+1)}$ between approximations at scales 2^{-j} and $2^{-(j+1)}$.

Iterating the two-scale decomposition [44], any function $f \in L^2(\mathbb{R})$ can be expressed as a sum of a coarse-scale approximation at a reference scale j_0 that we set to 0 here, and their successive

differences. These details are needed to go from one scale j to the next finer scale $j+1$ for $j=0,\ldots,J-1$,

$$f(x)=\sum_{i=-\infty}^{\infty}\overline{f}_{0,i}\phi_{0,i}(x)+\sum_{j=0}^{\infty}\sum_{i=-\infty}^{\infty}\widetilde{f}_{ji}\psi_{ji}(x) \qquad [46]$$

For numerical applications, the sums in eqn [46] have to be truncated in both scale j and position i. The truncation in scale corresponds to a limitation of f to a given finest scale J, which is in practice imposed by the available sampling rate. Due to the finite length of the available data, the sum over i also becomes finite. The decomposition [46] is orthogonal, as, by construction,

$$\langle\psi_{ji},\psi_{j'i'}\rangle=\delta_{jj'}\delta_{ii'} \qquad [47]$$

$$\langle\psi_{ji},\phi_{j'i'}\rangle=0 \quad \text{for } j\geq j' \qquad [48]$$

in addition to [31].

Fast Wavelet Transform

Starting with a function $f\in L^2(\mathbb{R})$ given at the finest resolution 2^{-J} (i.e., we know $f_J\in V_J$ and hence the coefficients $\overline{f}_{Ji}$ for $i\in\mathbb{Z}$), the FWT computes its wavelet coefficients $\widetilde{f}_{ji}$ by decomposing successively each approximation f_j into a coarser scale approximation f_{j-1}, plus the corresponding details which are encoded by the wavelet coefficients. The algorithm uses a cascade of discrete convolutions with the low pass filter h_n and the bandpass filter g_n, followed by downsampling, in which only one coefficient out of two is retained. The direct wavelet transform algorithm is

- initialization

$$\text{given } f\in L^2(\mathbb{R}) \text{ and } \overline{f}_{Ji}=f\left(\frac{i}{2^J}\right) \text{ for } i\in\mathbb{Z}$$

- decomposition
  for $j=J$ to 1, step -1, do

$$\overline{f}_{j-1,i}=\sum_{n\in\mathbb{Z}}h_{n-2i}\overline{f}_{jn} \qquad [49]$$

$$\widetilde{f}_{j-1,i}=\sum_{n\in\mathbb{Z}}g_{n-2i}\overline{f}_{jn} \qquad [50]$$

The inverse wavelet transform is based on successive reconstructions of fine-scale approximations f_j from coarser scale approximations f_{j-1}, plus the differences between approximations at scale $j-1$ and the finer scale j which are encoded by $\widetilde{f}_{j-1,i}$. The algorithm uses a cascade of discrete convolutions with the filters h_n and g_n, preceded by upsampling which adds zeros in between two successive coefficients.

- reconstruction
  for $j=1$ to J, step 1, do

$$\overline{f}_{ji}=\sum_{n=-\infty}^{\infty}h_{i-2n}\overline{f}_{j-1,n}+\sum_{n=-\infty}^{\infty}g_{i-2n}\widetilde{f}_{j,n} \qquad [51]$$

The FWT has been introduced by Stéphane Mallat in 1989. If the scaling functions (and wavelets) are compactly supported, the filters h_n and g_n have only a finite number of nonvanishing coefficients. In this case, the numerical complexity of the FWT is $\mathcal{O}(N)$ where N denotes the number of samples.

Choice of Wavelets

Orthogonal wavelets are typically defined by their filter coefficients h_n, since in general no analytic expression for ψ is available. In the following, we give the filter coefficients of h_n for some typical orthogonal wavelets. The filter coefficients of g_n can be obtained using the quadrature relation between the two filters [38].

- Haar D1 (one vanishing moment):

$$h_0=1/\sqrt{2}$$
$$h_1=1/\sqrt{2}$$

- Daubechies D2 (two vanishing moments):

$$h_0=0.482\,962\,913\,145$$
$$h_1=0.836\,516\,303\,736$$
$$h_2=0.224\,143\,868\,042$$
$$h_3=-0.129\,409\,522\,551$$

- Daubechies D3 (three vanishing moments):

$$h_0=0.332\,670\,552\,950$$
$$h_1=0.806\,891\,509\,311$$
$$h_2=0.459\,877\,502\,118$$
$$h_3=-0.135\,011\,020\,010$$
$$h_4=-0.085\,441\,273\,882$$
$$h_5=0.035\,226\,291\,882$$

- Coiflets C12 (four vanishing moments): the wavelets and the corresponding scaling function are shown in Figure 3.

Remarks The construction of orthogonal wavelets in $L^2(\mathbb{R})$ can be modified to obtain wavelets on the interval, that is, in $L^2([0,1])$. Therewith, boundary wavelets are introduced, while in the interior of the interval the wavelets are not modified.

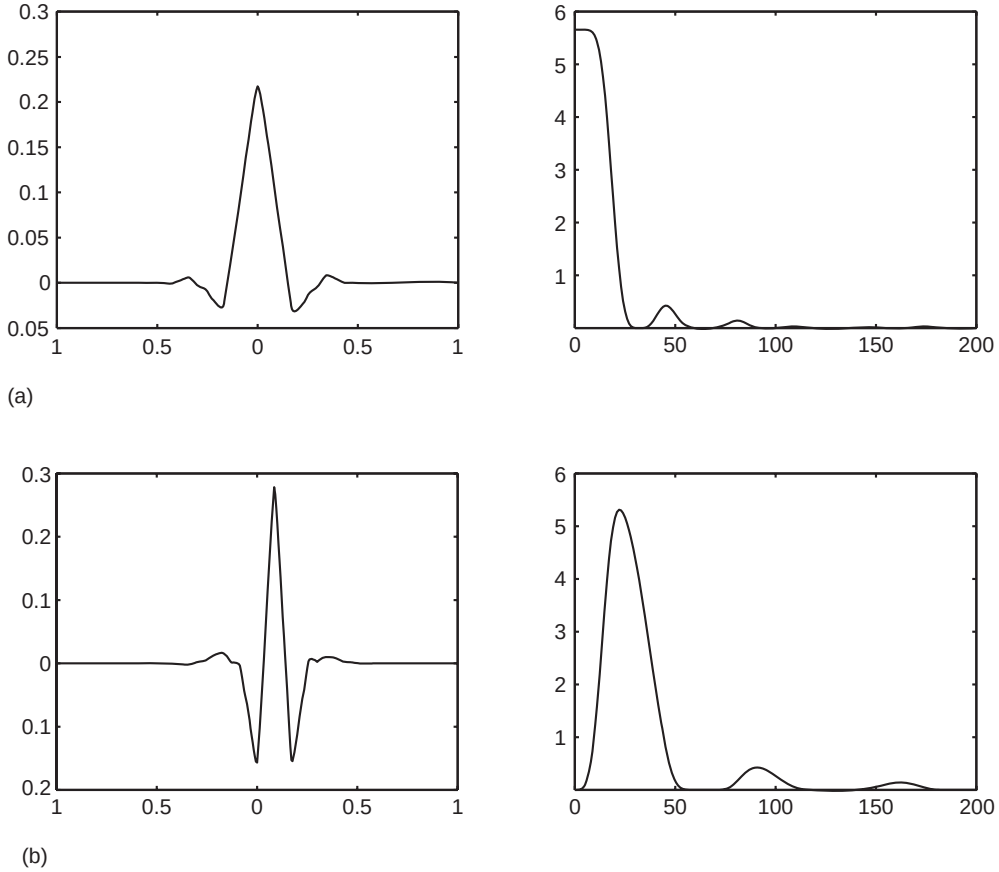

Figure 3 Orthogonal wavelets Coiflet C12. (a) Scaling function $\phi(x)$ (left) and $|\hat{\phi}(\omega)|$. (b) Wavelet $\psi(x)$ (left) and $|\hat{\psi}(\omega)|$.

A periodic MRA of $L^2(\mathbb{T})$, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denotes the torus, can also be constructed by periodizing the wavelets in $L^2(\mathbb{R})$, using

$$\psi^{\text{per}}(x) = \sum_{k \in \mathbb{Z}} \psi(x + k)$$

Relaxing the condition of orthogonality allows greater flexibility in the choice of the basis functions. For example, biorthogonal wavelets can be designed using different basis functions for analysis ($^{\text{a}}$) and synthesis ($^{\text{s}}$) which are related but no longer orthogonal. A couple of refinable scaling functions $(\phi^{\text{a}}, \phi^{\text{s}})$ with related wavelets $(\psi^{\text{a}}, \psi^{\text{s}})$ which are by construction biorthogonal generate a biorthogonal MRA $V_j^{\text{a}}, V_j^{\text{s}}$. From an algorithmic point of view, only two different filter couples $(g^{\text{a}}, h^{\text{a}})$ for the forward and $(g^{\text{s}}, h^{\text{s}})$ for the backward FWT are used, without changing the algorithm.

The multiresolution approach can be further generalized, for samplings on nonequidistant grids leading to the so-called second-generation wavelets.

Higher Dimensions

The previously presented one-dimensional construction can be extended to higher dimensions. For simplicity, we will consider only the two-dimensional case, since higher dimensions can be treated analogously.

Tensor product construction Having developed a one-dimensional orthonormal basis ψ_{ji} of $L^2(\mathbb{R})$, one could use these functions as building blocks in higher dimensions. One way of doing so is to take the tensor product of two one-dimensional bases and to define

$$\psi_{j_x, j_y, i_x, i_y}(x, y) = \psi_{j_x, i_x}(x)\psi_{j_y, i_y}(y) \qquad [52]$$

The resulting functions constitue an orthonormal wavelet basis for $L^2(\mathbb{R}^2)$. Each function $f \in L^2(\mathbb{R}^2)$ can then be developed into

$$f(x, y) = \sum_{j_x, i_x} \sum_{j_y, i_y} \widetilde{f}_{j_x, j_y, i_x, i_y} \psi_{j_x, j_y, i_x, i_y}(x, y) \qquad [53]$$

with $\widetilde{f}_{j_x, j_y, i_x, i_y} = \langle f, \psi_{j_x, j_y, i_x, i_y} \rangle$. However, in this basis the two variables x and y are dilatated separately

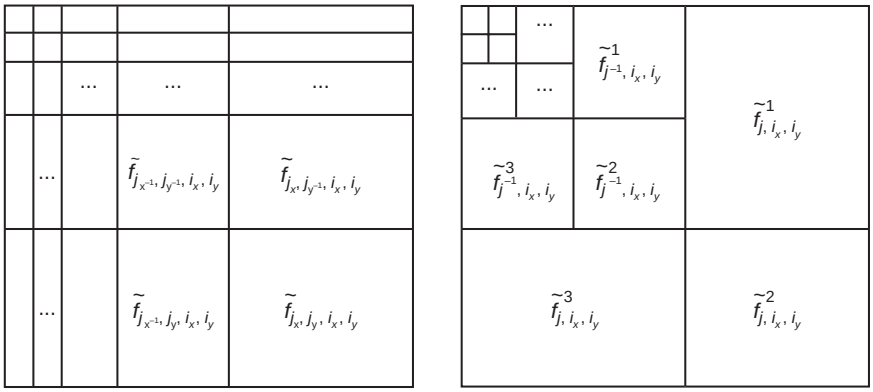

Figure 4a Schematic representation of the 2D (b) wavelet transforms: (a) Tensor product construction and (b) 2D MRA.

and therefore no longer form an MRA. This means that the functions ψ_{j_x,j_y} involve two scales, 2^{j_x} and 2^{j_y}, and each of the functions is essentially supported on a rectangle with these side-lengths. Hence, the decomposition is often called rectangular wavelet decomposition (cf. Figure 4a). From the algorithmic viewpoint, this is equivalent to applying the one-dimensional wavelet transform to the rows and the columns of a matrix or a function. For some applications, such a basis is advantageous, for others not. Often the notion of a scale has a certain meaning. For an application, one would like to have a unique scale assigned to each basis function.

Multiresolution construction Another much more interesting construction is the construction of a truly two-dimensional MRA of $L^2(\mathbb{R}^2)$. It can be obtained through the tensor product of two one-dimensional MRAs of $L^2(\mathbb{R})$. More precisely, one defines the spaces $V_j, j \in \mathbb{Z}$ by

$$V_j = V_j \otimes V_j \qquad [54]$$

and $V_j = \overline{\text{span}}\{\phi_{j,i_x,i_y}(x,y) = \phi_{j,i_x}(x)\phi_{j,i_y}(y), i_x, i_y \in \mathbb{Z}\}$ fulfilling analogous properties as in the one-dimensional case.

Likewise, we define the complement space W_j to be the orthogonal complement of V_j in V_{j+1}, that is,

$$\begin{aligned} V_{j+1} &= V_{j+1} \otimes V_{j+1} \\ &= (V_j \oplus W_j) \otimes (V_j \oplus W_j) \qquad [55] \\ &= V_j \otimes V_j \oplus ((W_j \otimes V_j) \\ &\quad \oplus (V_j \otimes W_j) \oplus (W_j \otimes W_j)) \qquad [56] \\ &= V_j \oplus W_j \qquad [57] \end{aligned}$$

It follows that the orthogonal complement $W_j = V_{j+1} \ominus V_j$ consists of three different types of functions and is generated by three different wavelets

$$\psi^{\varepsilon}_{j,i_x,i_y}(x,y) = \begin{cases} \psi_{j,i_x}(x)\phi_{j,i_y}(y), & \varepsilon = 1 \\ \phi_{j,i_x}(x)\psi_{j,i_y}(y), & \varepsilon = 2 \\ \psi_{j,i_x}(x)\psi_{j,i_y}(y), & \varepsilon = 3 \end{cases} \qquad [58]$$

Observe that here the scale parameter j simultaneously controls the dilatation in x and y. We recall that in d dimensions this construction yields $2^d - 1$ types of wavelets spanning W_j.

Using [58], each function $f \in L^2(\mathbb{R}^2)$ can be developed into a multiresolution basis as

$$f(x,y) = \sum_j \sum_{i_x,i_y} \sum_{\varepsilon=1,2,3} \tilde{f}^{\varepsilon}_{j,i_x,i_y} \psi^{\varepsilon}_{j,i_x,i_y}(x,y) \qquad [59]$$

with $\tilde{f}^{\varepsilon}_{j,i_x,i_y} = <f, \psi^{\varepsilon}_{j,i_x,i_y}>$. A schematic representation of the wavelet coefficients is shown in Figure 4b. The algorithmic structure of the one-dimensional transforms carries over to the two-dimensional case by simple tensorization, that is, applying the filters at each decomposition step to rows and columns.

Remark The described two-dimensional wavelets and scaling functions are separable. This advantage is the ease of generation starting from one-dimensional MRAs. However, the main drawback of this construction is that three wavelets are needed to span the orthogonal complement space W_j. Another property should be mentioned. By construction, the wavelets are anisotropic, that is, horizontal, diagonal, and vertical directions are preferred.

Approximation Properties

Reproduction of Polynomials

A fundamental property of the MRA is the exact reproduction of polynomials. The vanishing moments of the wavelet ψ, that is, $\int_{\mathbb{R}} x^m \psi(x)\mathrm{d}x = 0$

for $m = 0, M-1$, is equivalent to the fact that polynomials up to degree $M-1$, can be expressed exactly as a linear combination of scaling functions, $p_m(x) = \sum_{n\in\mathbb{Z}} n^m \phi(x-n)$ for $m = 0, M-1$. This so-called Strang–Fix condition proves that ψ has M vanishing moments if and only if any polynomial of degree $M-1$ can be written as a linear combination of scaling functions ϕ. Note that, as $p_m \notin L^2(\mathbb{R})$, the coefficients n^m are not in $l^2(\mathbb{Z})$.

Regularity and Local Decay of Wavelet Coefficients

The local or global regularity of a function is closely related to the decay of its wavelet coefficients. If a function is locally in $C^s(\mathbb{R})$ (the space of s-times continuously differentiable functions), it can be well approximated locally by a Taylor series of degree s. Consequently, its wavelet coefficients are small at fine scales, as long as the wavelet ψ has enough vanishing moments. The decay of the coefficients hence determines directly the error being made when truncating a wavelet sum at some scale.

Depending on the type of norm used and whether global or local characterization is concerned, various relations of this kind have been developed. Let us take as example the case of an α-Lipschitz function.

Suppose $f \in L^2(\mathbb{R})$, then for $[a,b] \subset \mathbb{R}$ the function f is α-Lipschitz with $0 < \alpha < 1$ for any $x_0 \in [a,b]$, that is, $|f(x_0+h) - f(x_0)| \le C|h|^\alpha$, if and only if there exists a constant A such that $|\widetilde{f}_{ji}| \le A2^{-j\alpha-1/2}$ for any (j,i) with $i/2^j \in [a,b]$.

This shows the relation between the local regularity of a function and the decay of its wavelet coefficients in scale.

Example To illustrate the local decay of the wavelet coefficients, we consider in Figure 5 the function $f(x) = \sin(2\pi x)$ for $x \le 1/4$ and $x \ge 3/4$ and $f(x) = -\sin(2\pi x)$ for $1/4 < x < 3/4$. The corresponding wavelet coefficients for quintic spline wavelets are plotted in logarithmic scale. The wavelet coefficients show that only in a local region around singularities the fine-scale coefficients are significant.

Linear Approximation

The exact reproduction of polynomials can be used to derive error estimates for the approximation of a function f at a given scale, which corresponds to linear approximation. We consider f belonging to the Sobolev space $W^{s,p}(\mathbb{R}^d)$, that is, the weak derivatives of f up to order s belong to $L^p(\mathbb{R}^d)$. The linear approximation of f at scale J, corresponding to the projection of f onto V_J, is then given by

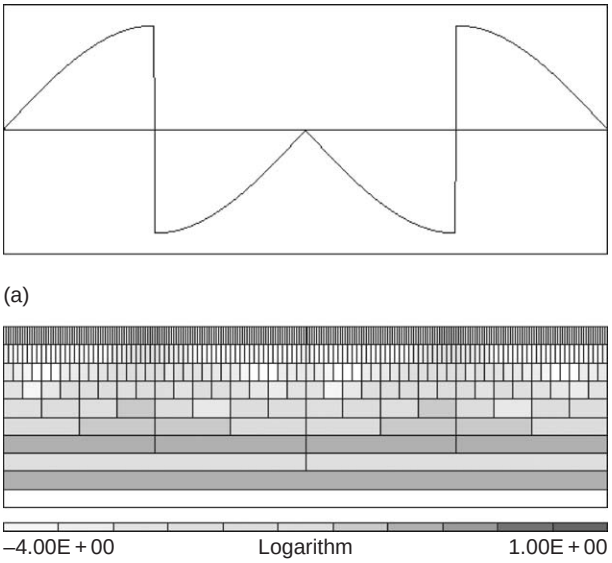

Figure 5 Orthogonal wavelet decomposition using quintic spline wavelets: (a) function $f(x) = \sin(2\pi x)$ for $x \le 1/4$ and $x \ge 3/4$ and $f(x) = -\sin(2\pi x)$ for $1/4 < x < 3/4$ sampled on a grid $x_i = i/2^J, i = 0, \ldots, 2^J - 1$ with $J = 9$ and (b) corresponding wavelet coefficients $\log_{10}|\widetilde{f}_{j,i}|$ for $i = 0, \ldots, 2^j - 1$ and $j = 0, \ldots, J-1$.

$$f_J(x) = \sum_{j=0}^{J-1} \sum_{i\in\mathbb{Z}} \widetilde{f}_{j,i}\psi_{j,i}(x) \qquad [60]$$

The approximation error can be estimated by

$$\|f - f_J\|_{L^p} < C2^{-J\min(s,m)/d} \qquad [61]$$

where s denotes the smoothness of the function in L^p, d the space dimension, and m the number of vanishing moments of the wavelet ψ. In the case of poor global regularity of f, that is, for small s, a large number of scales J is needed to get a good approximation of f.

In Figure 6, we plot the linear approximation of the function f shown in Figure 5. The function f_6 is reconstructed using wavelet coefficients up to scale $J-1 = 5$, so that in total only 64 out of 512 coefficients are retained. We observe an oscillating behavior of f_J near the discontinuities of f which dominates the approximation error.

Nonlinear Approximation

Retaining the N largest wavelet coefficients in the wavelet expansion of f in [46], without imposing any *a priori* cutoff scale, yields the best N-term approximation f^N. In contrast to the linear approximation [60], it is called nonlinear approximation, since the choice of the retained coefficients depends

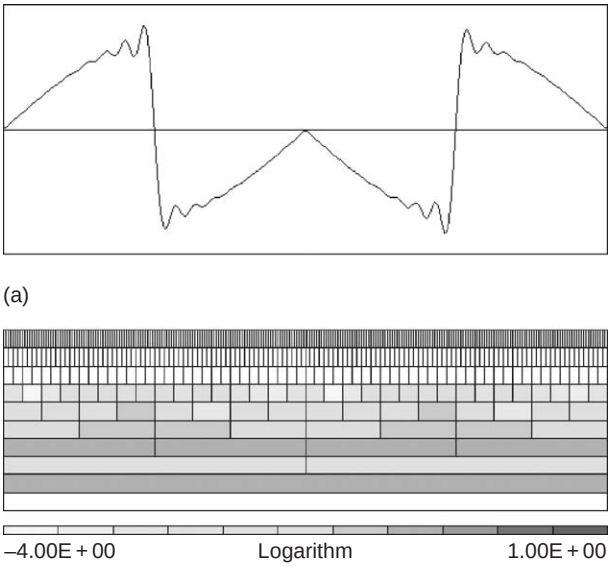

Figure 6 (a) Linear approximation f_J of the function f in Figure 5 for $J=6$, reconstructed from 64 wavelet coefficients using quintic splines wavelets and (b) corresponding wavelet coefficients $\log_{10}|\tilde{f}_{j,i}|$ for $i=0,\ldots,2^j-1$ and $j=0,\ldots,J-1$. Note that the coefficients for $J>5$ have been set to zero.

on the function f. The mathematical theory has been formalized by Cohen, Dahmen, and De Vore.

The nonlinear approximation of the function f can then be written as

$$f^N(x)=\sum_{(j,i)\in\Lambda_N}\tilde{f}_{j,i}\,\psi_{j,i}(x) \qquad [62]$$

where Λ_N denotes the ensemble of all multi-indices $\lambda=(j,i)$, indexing the N largest coefficients (measured in the l^p norm),

$$\Lambda_N=\{\lambda_k,k=1,N|\ \ \|\tilde{f}_{\lambda_k}\|_{l^p}>\|\tilde{f}_{\mu}\|_{l^p}\ \ \forall\mu\in\Lambda\} \qquad [63]$$

with $\Lambda=\{\mu=(j,i),j\geq 0,i\in\mathbb{Z}\}$. The nonlinear approximation leads to the following error estimate:

$$\|f-f^N\|_{L^p}<CN^{-s/d} \qquad [64]$$

where s denotes the smoothness of f in the larger space $L^q(\mathbb{R}^d)$ with

$$\frac{1}{q}=\frac{1}{p}+\frac{s}{d}$$

which corresponds to the Sobolev embedding line (Figure 7). This estimate shows that the nonlinear approximation converges faster than the linear one, if f has a larger regularity in L^q, that is, $f\in W^{s,q}(\mathbb{R}^d)$, which is for example the case for functions with isolated singularities and for small q.

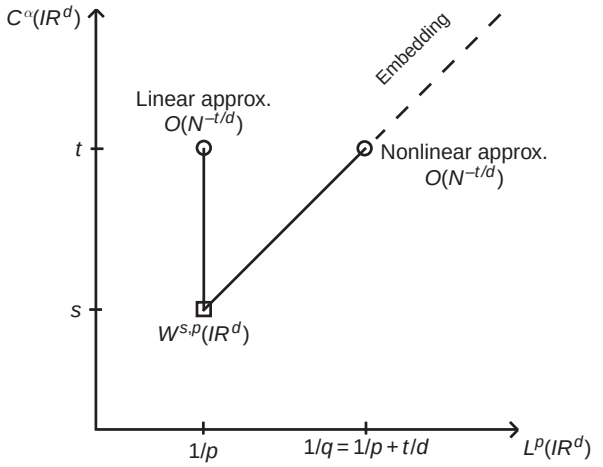

Figure 7 Schematic representation of linear and nonlinear approximation.

In Figure 8, we plot the nonlinear approximation of the function f shown in Figure 5. The function f^N is reconstructed using the strongest 64 wavelet coefficients out of 512 coefficients. Compared to the linear approximation (cf. Figure 6), the oscillations around the discontinuities disappear and the approximation error is reduced while using the same number of coefficients.

Compression and Preconditioning of Operators

The nonlinear approximation of functions can be extended to certain operators leading to an efficient

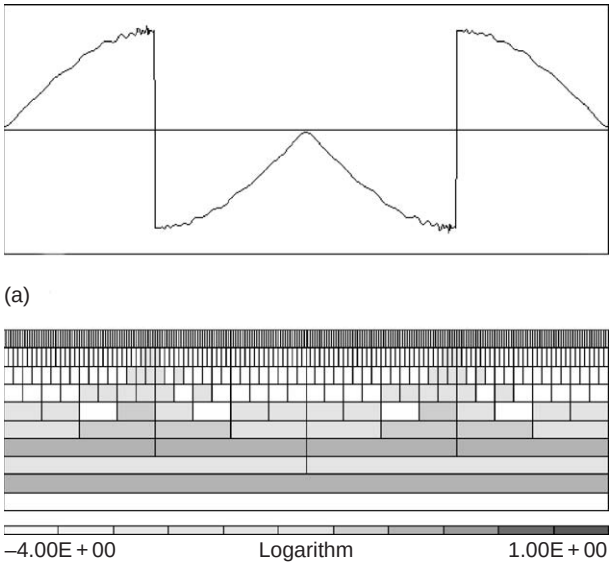

Figure 8 (a) Nonlinear approximation f^N of the function f in Figure 5 reconstructed from the 64 largest wavelet coefficients using quintic splines wavelets, (b) retained wavelet coefficients $\log_{10}|\tilde{f}_{j,i}|$ for $i=0,\ldots,2^j-1$ and $j=0,\ldots,J-1$.

representation in wavelet space, that is, to sparse matrices. For integral operators, for example, Calderon–Zygmund operators T on $\mathbb{R}$ defined by

$$Tf(x) = \int_{\mathbb{R}} K(x,y)f(y)\,dy \qquad [65]$$

where the kernel k satisfies

$$|k(x,y,)| \leq \frac{C}{|x-y|}$$

and

$$\left|\frac{\partial}{\partial x}k(x,y)\right| + \left|\frac{\partial}{\partial y}k(x,y,)\right| \leq \frac{C}{|x-y|^2}$$

their wavelet representation $\langle T\psi_{j,i}, \psi_{j',i'}\rangle$ is sparse and a large number of weak coefficients can be suppressed by simple thresholding of the matrix entries while controlling the precision. The resulting numerical scheme is called BCR algorithm and is due to Beylkin *et al.* (1991).

The characterization of function spaces by the decay of the wavelet coefficients and the corresponding norm equivalences can be used for diagonal preconditioning of integral or differential operators which leads to matrices with uniformly bounded condition numbers. For elliptic differential operators, for example, the Laplace operator ∇^2 the norm equivalence $\|\nabla^2 f\| \simeq \|2^{2j}\widetilde{f}_{ji}\|$ can be used for preconditioning the matrix $\langle \nabla^2\psi_{j,i}, \psi_{j',i'}\rangle$ by a simple diagonal scaling with 2^{-2j} to obtain a uniformly bounded condition number. For further details, we refer to the book of Cohen (2000).

Wavelet Denoising

We consider a function f which is corrupted by a Gaussian white noise $n \in \mathcal{N}(0,\sigma^2)$. The noise is spread over all wavelet coefficients $\widetilde{s}_\lambda$, while, typically, the original function f is determined by only few significant wavelet coefficients. The aim is then to reconstruct the function f from the observed noisy signal $s = f + n$.

The principle of the wavelet denoising can be summarized in the following procedure:

- *Decomposition*. Compute the wavelet coefficients $\widetilde{s}_\lambda$ using the FWT.
- *Thresholding*. Apply the thresholding function ρ_ε to the wavelet coefficients $\widetilde{s}_\lambda$, thus reducing the relative importance of the coefficients with small absolute value.
- *Reconstruction*. Reconstruct a denoised version s_C from the thresholded wavelet coefficients using the fast inverse wavelet transform.

The thresholding parameter ε depends on the variance of the noise and on the sample size N. The thresholding function ρ we consider corresponds to hard thresholding:

$$\rho_\varepsilon(a) = \begin{cases} a & \text{if } |a| > \varepsilon \\ 0 & \text{if } |a| \leq \varepsilon \end{cases} \qquad [66]$$

Donoho and Johnstone (1994) have shown that there exists an optimal ε for which the relative quadratic error between the signal s and its estimator s_C is close to the minimax error for all signals $s \in \mathcal{H}$, where $\mathcal{H}$ belongs to a wide class of function spaces, including Hölder and Besov spaces. They showed using the threshold

$$\varepsilon_D = \sigma_n\sqrt{2\ln N} \qquad [67]$$

yields an error which is close to the minimum error. The threshold ε_D depends only on the sampling N and on the variance of the noise σ_n; hence, it is called universal threshold. However, in many applications, σ_n is unknown and has to be estimated from the available noisy data s. For this, the present authors have developed an iterative algorithm (see Azzolini *et al.* (2005)), which is sketched in the following:

1. *Initialization*
   (a) given $s_k, k = 0,\ldots,N-1$. Set $i = 0$ and compute the FWT of s to obtain $\widetilde{s}_\lambda$;
   (b) compute the variance σ_0^2 of s as a rough estimate of the variance of n and compute the corresponding threshold $\varepsilon_0 = (2\ln N\sigma_0^2)^{1/2}$;
   (c) set the number of coefficients considered as noise $N_{\text{noise}} = N$.
2. *Main loop repeat*
   (a) set $N'_{\text{noise}} = N_{\text{noise}}$ and count the wavelet coefficients N_{noise} with modulus smaller than ε_i;
   (b) compute the new variance σ_{i+1}^2 from the wavelet coefficients whose modulus is smaller than ε_i and the new threshold $\varepsilon_{i+1} = (2(\ln N)\sigma_{i+1}^2)^{1/2}$;
   (c) set i = i + 1 until $(N'_{\text{noise}} = = N_{\text{noise}})$.
3. *Final step*
   (a) compute s_C from the coefficients with modulus larger than ε_i using the inverse FWT.

Example To illustrate the properties of the denoising algorithm, we apply it to a one-dimensional test signal. We construct a noisy signal s by superposing a Gaussian white noise, with zero mean and variance $\sigma_W^2 = 1$, to a function f, normalized such that $((1/N)\sum_k |f_k|^2)^{1/2} = 10$. The number of samples is

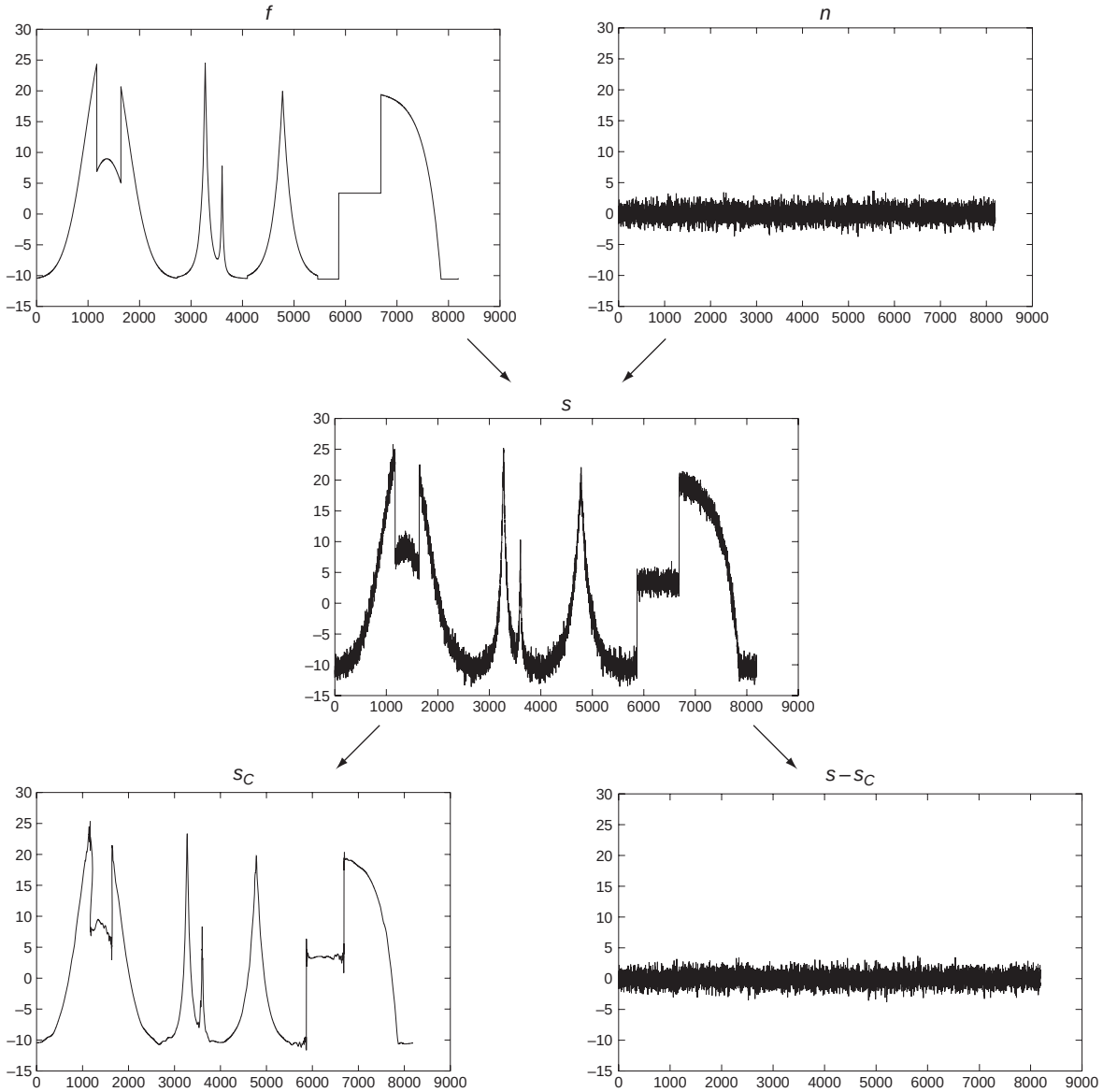

Figure 9 Construction (top) of a 1D noisy signal $s = f + n$ (middle), and results obtained by the recursive denoising algorithm (bottom).

$N = 8192$. Figure 9a shows the function f together with the noise n; Figure 9b shows the constructed noisy signal s and Figure 9c shows the wavelet denoised signal s_C together with the extracted noise.

Acknowledgments

Marie Farge thankfully acknowledges Trinity College, Cambridge, UK, and CIRM, Marseille, France, for support while writing this paper. The authors also thank Barbara Burke for kindly revising their English.

See also: Coherent States; Fractal Dimensions in Dynamics; Homeomorphisms and Diffeomorphisms of the Circle; Image Processing: Mathematics; Wavelets: Application to Turbulence; Wavelets: Applications.

Further Reading

Azzolini A, Farge M, and Schneider K (2005) Nonlinear wavelet thresholding: A recursive method to determine the optimal denoising threshold. *Applied and Computational Harmonic Analysis* 18(2): 177.

Beylkin, Coifman, and Rohklin (1991) Fast wavelet transforms and numerical algorithms. *Communications in Pure and Applied Mathematics* 44: 141.

Cohen A (2000) Wavelet methods in numerical analysis. In: Ciarlet PG and Lions JL (eds.) *Handbook of Numerical Analysis*, vol. 7. Amsterdam: Elsevier.

Dahmen W (1997) Wavelets and multiscale methods for operator equations. *Acta Numerica* 6: 55–228.

Daubechies I (1988) Orthonormal bases of compactly supported wavelets. *Communications in Pure and Applied Mathematics* 41: 909.

Daubechies I (1992) *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM.

Donoho and Johnstone (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81: 425.

Grossmann A and Morlet J (1984) Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal of Mathematical Analysis*. 15(4): 723–736.

Lemarié P-G and Meyer Yves (1986) Ondolettes et bases hilbertiennes. *Revista Matematica Iberoamericano* 2: 1.

Mallat S (1989) Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Transactions of the American Mathematical Society* 315: 69.

Mallat S (1998) *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic Press.

WDVV Equations and Frobenius Manifolds

B Dubrovin, SISSA-ISAS, Trieste, Italy

Main Definition

WDVV equations of associativity (after E Witten, R Dijkgraaf, E Verlinde, and H Verlinde) is tantamount to the following problem: find a function $F(v)$ of n variables $v=(v^1,v^2,\dots,v^n)$ satisfying the conditions [1], [3], and [4] given below. First,

$$\frac{\partial^3 F(v)}{\partial v^1\partial v^\alpha\partial v^\beta}\equiv\eta_{\alpha\beta} \qquad [1]$$

must be a constant symmetric nondegenerate matrix. Denote $(\eta^{\alpha\beta})=(\eta_{\alpha\beta})^{-1}$ the inverse matrix and introduce the functions

$$c^\gamma_{\alpha\beta}(v)=\eta^{\gamma\epsilon}\frac{\partial^3 F(v)}{\partial v^\epsilon\partial v^\alpha\partial v^\beta},\quad \alpha,\beta,\gamma=1,\dots,n \qquad [2]$$

The main condition says that, for arbitrary $v^1,\dots,v^n$ these functions must be structure constants of an associative algebra, that is, introducing a v-dependent multiplication law in the n-dimensional space by

$$a\cdot b:=\left(c^1_{\alpha\beta}(v)a^\alpha b^\beta,\dots,c^n_{\alpha\beta}(v)a^\alpha b^\beta\right)$$

one obtains an n-parameter family of n-dimensional associative algebras (these algebras will automatically be also commutative). Spelling out this condition one obtains an overdetermined system of nonlinear PDEs for the function $F(v)$ often also called WDVV associativity equations

$$\frac{\partial^3 F(v)}{\partial v^\alpha\partial v^\beta\partial v^\lambda}\eta^{\lambda\mu}\frac{\partial^3 F(v)}{\partial v^\mu\partial v^\gamma\partial v^\delta}=\frac{\partial^3 F(v)}{\partial v^\delta\partial v^\beta\partial v^\lambda}\eta^{\lambda\mu}\frac{\partial^3 F(v)}{\partial v^\mu\partial v^\gamma\partial v^\alpha} \qquad [3]$$

for arbitrary $1\le\alpha,\beta,\gamma,\delta\le n$. (Summation over repeated indices will always be assumed.) The last one is the so-called quasihomogeneity condition

$$EF=(3-d)F+\tfrac{1}{2}A_{\alpha\beta}v^\alpha v^\beta+B_\alpha v^\alpha+C \qquad [4]$$

where

$$E=\left(a^\alpha_\beta v^\beta+b^\alpha\right)\frac{\partial}{\partial v^\alpha}$$

for some constants a^α_β, b^α satisfying

$$a^\alpha_1=\delta^\alpha_1,\qquad b^1=0$$

$A_{\alpha\beta},B_\alpha,C,d$ are some constants. E is called Euler vector field and d is the charge of the Frobenius manifold.

For $n=1$ one has $F(v)=(1/6)v^3$. For $n=2$ one can choose

$$F(u,v)=\tfrac{1}{2}uv^2+f(u)$$

only the quasihomogeneity [4] makes a constraint for $f(v)$. The first nontrivial case is for $n=3$. The solution to WDVV is expressed in terms of a function $f=f(x,y)$ in one of the two forms (in the examples all indices are written as lower):

$$\begin{aligned} d\neq 0:\quad & F=\tfrac{1}{2}v_1^2v_3+\tfrac{1}{2}v_1v_2^2+f(v_2,v_3)\\ & f^2_{xxy}=f_{yyy}+f_{xxx}f_{xyy}\\ d=0:\quad & F=\tfrac{1}{6}v_1^3+v_1v_2v_3+f(v_2,v_3)\\ & f_{xxx}f_{yyy}-f_{xxy}f_{xyy}=1\end{aligned} \qquad [5]$$

The function $f(x,y)$ satisfies additional constraint imposed by [4]. Because of this the above PDEs [5] can be reduced (Dubrovin 1992, 1996) to a particular case of the Painlevé-VI equation (*see* Painlevé Equations).

The problem [1], [3], [4] is invariant with respect to linear changes of coordinates preserving the direction of the vector $\partial/\partial v^1$:

$$v^\alpha\mapsto\tilde v^\alpha=P^\alpha_\beta v^\beta+Q^\alpha,\ \det(P^\alpha_\beta)\neq 0,\quad P^\alpha_1=\delta^\alpha_1$$