Wavelets and Other Phase Space Localization Methods

Ingrid Daubechies
Princeton University, Princeton NJ 08544, USA

1. Introduction

Mathematicians have various ways of judging the merits of new theorems and constructions. One very important criterion is esthetic–some developments just "feel" right, fitting, and beautiful. Just as in other venues where beauty or esthetics are discussed, taste plays an important role in this, but I think I am not alone in being especially excited when apparently different fields suddenly meet in a new concept, a new understanding. It is often of the sparks of such encounters that our esthetic enjoyment of mathematics is born.

Another important criterion for according merit to some particular piece of mathematics is the extent to which it can be useful in applications; this is the criterion almost exclusively used by non-mathematicians. Mathematicians themselves do not discount the importance of mathematics for applications (after all, if we were producing only beauty, there wouldn't be as many teaching positions allotted to us), but often beauty is considered the real grail, with applicability second-best. Although we have come some way since Hardy's "A Mathematician's Apology," we often still believe, maybe subliminally, that the two criteria are exclusive–that mathematics, when really close to applications, cannot be beautiful and is often even "dirty."

I believe that this does not have to be so; a wish for beauty and simplicity, and a desire to bring different fields together, can equally well drive developments in "applicable" mathematics.

When mapping out this presentation, I initially thought that I wanted to speak about wavelets, but I soon realized that other developments, aside from or beyond wavelets, should have their place here as well, and the scope was enlarged to add the "other phase space localization methods." Let me start by explaining what I mean by this.

I shall use the term "phase space" when a special type of description is meant, involving several complementary variables. It is really a term that is appropriated here from physics. Imagine that you want to describe the motion of a planet in the solar system. A simple way to do this is to give, as a function of time, its position in space as well as its momentum. This is a phase space description: the two complementary variables are position and momentum, and you are describing the motion by a curve in phase space. "Phase space localization" is no problem here: both position and momentum can be measured, with arbitrarily high precision. ("Phase space" is also used in a more general sense for other dynamical systems, but that is a different story.) The situation is different if we look at a quantum system, say an electron in a solid state crystal, where measuring position and momentum both, simultaneously, with arbitrarily high precision, cannot be done: the uncertainty principle forbids it. Nevertheless, it is still very useful to think in terms of phase space, or momentum and position, when comparing a quantum system with its classical analog, for instance. This poses a problem to the theoretical physicist: how to give a description, localized in phase space, despite the uncertainty principle? The mathematical model for a quantum mechanical particle assigns to a physical state a wave function $\psi(x)$, where x is the position variable; an equivalent description is given by the Fourier transform $\hat{\psi}(p)$, where p is interpreted as the momentum variable. Trying to "localize in phase space" amounts therefore to pinning down, as well as possible, a function's local properties and the local properties of its Fourier transform simultaneously–something analysts have been doing for decades under the name *microlocalization*.

The same problem also crops up in electrical engineering, or in statistics: for instance, when trying to understand signals depending on time, such as a recorded audio signal, it is often useful to gauge its spectrum or frequency content, again modeled naturally by the Fourier transform of the data. But the make-up of such signals, in terms of their different frequency characteristics, seems to change with time–this is immediately clear when you think of a music score which, after all, tells the musician to play different notes (= frequencies) at different times. Once again, the intuitive notion of the mathematical tool needed involves localization in phase space, with the two complementary variables now in the form of time and frequency.

Similarly, the computer scientist or engineer working with images (such as any image on your television screen) finds it helpful to break it up in smaller pieces (localization in space) and to look at the different spatial frequencies present in those pieces: again a phase space localization, now in two dimensions.

Since similar problems occur in different disciplines, it is not surprising that the answers developed, often independently, have some similarity as well. What I want to describe here is how the synthesis of different points of view and different approaches, has led in some cases to new developments, making the whole much more than the sum of its parts.

Before embarking on a more detailed discussion, I would like to point out that this presentation will summarize essential contributions by many people besides myself. At the ICM '90 in Kyoto, both R. Coifman and Y. Meyer gave talks related to this one; at this ICM, related talks include those by W. Dahmen, D. Donoho, and V. Rokhlin. For a more complete list of important contributors, I refer the reader to the references and their references. I would like to take this opportunity to thank especially R. Coifman, A. Cohen, A. Grossmann, S. Mallat, and Y. Meyer, from all of whom I learned a lot.

2. Wavelets

Most of this presentation will concern the development of wavelets, in particular of orthonormal wavelet bases, our growing understanding of their mathematical properties, and the ways in which they can be applied.

What are wavelets? To keep things simple, I shall restrict myself mostly to one dimension; with slight modifications, everything here can be generalized to higher dimensions (the few exceptions will be pointed out explicitly). I shall also almost systematically *not* try to give the most general conditions under which my statements hold, preferring to strip down the technicalities so as to lay bare the essential ideas.

A typical example of a family of wavelets $\psi_{j,k}(x)$ is given by

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k) = 2^{-j/2}\psi(\frac{x - 2^j k}{2^j}) \quad , \quad j,k \in \mathbb{Z} \ , \tag{1}$$

where ψ is a function with reasonable decay (say, $|\psi(x)| < C(1+|x|)^{-(1+\epsilon)}$), with some smoothness (as measured by the decay of the Fourier transform $\hat{\psi}$, say $|\hat{\psi}(\xi)| < C(1+|\xi|)^{-(1+\epsilon)}$), and such that $\int \psi(x) = 0$. For particular choices of ψ, the $\psi_{j,k}$ constitute a(n orthonormal) basis for $L^2(\mathbb{R})$; I shall mainly restrict myself to this case (although there are many interesting applications that use wavelets that are not linearly independent, which fall outside this framework). The first known example of a function ψ for which the $\psi_{j,k}$ give an orthonormal basis is the Haar wavelet, known since 1910,

$$\psi(x) = \begin{array}{rll} 1 & \text{if} & 0 \le x < 1/2 \\ -1 & \text{if} & 1/2 \le x < 1 \\ 0 & \text{otherwise} & ; \end{array} \tag{2}$$

this does not satisfy the smoothness requirement above. Much smoother constructions were found only in the 80's: Stromberg (1982), Meyer (1985), Battle (1987), Lemarié (1988), and Daubechies (1988) are some examples. The first construction, by Stromberg, did not attract a lot of attention at the time, although it later turned out to be very useful, not only for the harmonic analyst, but also computationally. Meyer rediscovered that dilations and translations of a single smooth and decaying function, as in (1) above, could give rise to orthonormal bases for $L^2(\mathbb{R})$; in his example both ψ and $\hat{\psi}$ are C^∞ and $\hat{\psi}$ has compact support. The constructions by Battle and Lemarié use $\psi \in C^m$, where m can be arbitrarily large but finite; moreover ψ has exponential decay. (Stromberg's ψ has similar properties.) These first ad hoc constructions became much more transparent with the development by S. Mallat (1989) and Y. Meyer of multiresolution analysis, a framework which linked wavelets with approximation theory. Interestingly, this construction was triggered by analogies with tools in vision theory, with which S. Mallat was familiar. Multiresolution analysis was then used in Daubechies (1988) to construct a basis of type (1) where ψ is still in C^m but compactly supported.

3. Multiresolution Analysis

The multiresolution analysis framework views the expansion of f in $L^2(R)$ with respect to an orthonormal wavelet basis,

$$f = \sum_{j,k \in \mathbb{Z}} < f, \psi_{j,k} > \psi_{j,k} \quad , \tag{3}$$

as a decomposition of f into successive layers, each more detailed than the previous one. That is, we write

$$L^2(\mathbb{R}) = \overline{\bigcup_{j \in \mathbb{Z}} V_j} \quad ,$$

where the spaces V_j constitute a nested sequence of approximation spaces,

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} \quad .$$

For fixed j, summing the terms in (3) over k gives exactly the layer to be peeled away from $P_{j-1}f := \mathrm{Proj}_{V_{j-1}} f$ to reach the coarser approximation $P_j f := \mathrm{Proj}_{V_j} f$,

$$P_{j-1}f = P_j f \quad + \quad \sum_k < f, \psi_{j,k} > \psi_{j,k} \quad . \tag{4}$$

For the Haar basis, the corresponding spaces V_j are given by

$$V_j = \{ f \in L^2(\mathbb{R}) \; ; f|_{[2^j k, 2^2(k+1)[} \; = \; \text{constant for each } k \in \mathbb{Z} \}.$$

For the constructions of Stromberg, Battle, and Lemarié, the multiresolution hierarchy consists of spaces of spline functions,

$$\begin{aligned} V_j \quad = \quad & \{ f \in L^2(\mathbb{R}) \; ; f \in C^m \text{ and} \\ & f|_{[2^j k, 2^2(k+1)[} \; = \; \text{polynomial of degree } m+1, \text{ for each } k \in \mathbb{Z} \}. \end{aligned} \tag{5}$$

Additional requirements are that the spaces V_j are all scaled versions of each other,

$$f \in V_j \Leftrightarrow f(2^j \cdot) \in V_0$$

(as is obviously the case in the examples above) and that the central space V_0 is invariant under integer translation. This invariance follows automatically from the final requirement, that there exists a function ϕ in V_0, commonly called the *scaling* function, such that the $\phi(. - k) = 2^{-j/2}\phi(2^{-j}x - k), k \in \mathbb{Z}$, constitute an orthonormal basis for V_j. In the Haar basis case, $\phi(x)$ is taken to be $\chi_{[0,1[}(x)$, the characteristic function of $[0, 1[$; in the spline examples, ϕ is a spline function of the appropriate order and with exponential decay. The work of Lemarié (1993) and Auscher (1992) proves that *any* wavelet basis of type (1) is associated with such a multiresolution analysis, provided that ψ has some smoothness and decay. (Note that this result does not completely translate to higher dimensions.)

As a result of the nesting property of the V_j, we have that $\phi \in V_0 \subset V_{-1}$, so that ϕ can be written as a linear combination of the orthonormal basis functions $\phi_{-1,k}$ in V_{-1}:

$$\phi(x) = \sqrt{2} \sum_n h_n \phi(2x - n) \quad . \tag{6}$$

Similarly $\psi \in V_{-1}$, so that

$$\psi(x) = \sqrt{2} \sum_n g_n \phi(2x - n) \quad . \tag{7}$$

For other scales j, (6) and (7) can be rewritten as

$$\phi_{j,k} = \sum_n h_n \phi_{j-1,2+2k} \quad , \quad \psi_{j,k} = \sum_n g_n \phi_{j-1,2+2k} \quad . \tag{8}$$

4. Fast Algorithm for a Decomposition Into Wavelets

The layered structure of the underlying multiresolution analysis translates into a fast algorithm for the wavelet decomposition of functions. In numerical applications, the function f to decompose will be given with a finite resolution only: for instance, in the form of samples. That is, we really know only the projection of f onto one of the spaces V_{j_0} in the scale; all information pertaining to finer structure (corresponding to $\psi_{j,k}$ with $j \leq j_0$) cannot be recovered (unless we have a priori information on f). Let us rescale our length unit so that $j_0 = 0$. Then we suppose that we know

$$P_0 f = \sum_k s_{0,k} \phi_{0,k} = \sum_k < f, \phi_{0,k} > \phi_{0,k} \qquad ,$$

where the $s_{0,k}$ are either given or computed from the data, depending on the application. Because of (4), we have

$$P_0 f = P_1 f + \sum_k < f, \psi_{1,k} > \psi_{1,k} \quad ,$$

and with the help of (8), the $d_{1,k} \; := \; < f, \psi_{1,k} >$ can be computed by

$$d_{1,k} =< f, \sum_n g_n \phi_{0,n+2k} > \; = \sum_n g^*_{n-2k} s_{0,n} \quad .$$

Similarly, we can also compute $s_{1,k} :=< f, \phi_{1,k} >$ (the coefficients of $P_1 f$ in the orthogonal basis $\{\phi_{1,k}, k \in \mathbb{Z}\}$ of V_1) by using (6) again,

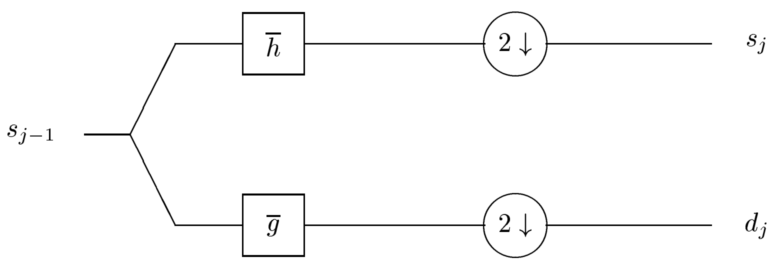
$$s_{1,k} = \sum_n h^*_{n-2k} s_{0,n} \quad .$$

$P_1 f$ can be decomposed further into $P_2 f$ and a combination of the $\psi_{2,k}$; this can be repeated for successively higher values of j, resulting in

$$d_{j,k} := < f, \psi_{j,k} > = \sum_n g^*_{n-2k} s_{j-1,n} \tag{9}$$

$$s_{j,k} := < f, \phi_{j,k} > = \sum_n h_{n-2k} s_{j-1,n} \quad . \tag{10}$$

These formulas consist of a convolution of the sequence s_{j-1} with $\overline{h} = (h^*_{-n})_{n\in\mathbb{Z}}$ or $\overline{g} = (g^*_{-n})_{n\in\mathbb{Z}}$, followed by retaining only the even entries of the result. Schematically, this is represented by



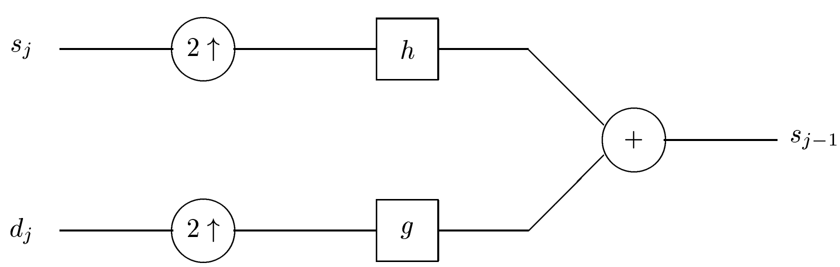
, (11)

where the symbol $\boxed{a}$ stands for convolution with the sequence a, and $2\downarrow$ is "decimation by a factor 2."

The transition from the sequence $(s_{j-1,n})_{n\in\mathbb{Z}}$ to the two sequences $(s_{j,k})_{k\in\mathbb{Z}}$, $(d_{j,k})_{k\in\mathbb{Z}}$ corresponds to a change of basis in V_{j-1}, from $\{\phi_{j-1}, n \in \mathbb{Z}\}$ to $\{\phi_{j,k}, \psi_{j,k}; k \in \mathbb{Z}\}$. The inverse operation corresponds to the adjoint unitary operator, and we have

$$s_{j-1,n} = \sum_k [h_{n-2k} s_{j,k} + g_{n-2k} d_{j,k}] \quad . \tag{12}$$

Each of the two terms in the right hand side of (12) can be viewed as the result of first "upsampling by 2," i.e. taking the given sequence as the even entries of a new sequence in which all the odd entries are zero, followed by a convolution. Schematically, this becomes



. (13)

For the Haar basis one finds $h_0 = \frac{1}{\sqrt{2}} = h_1, g_0 = \frac{1}{\sqrt{2}} = -g_1$, with all other $h_n, g_n = 0$. The decomposition steps (9) and (10) then correspond to breaking up the sequence s_{j-1} into pairs, and replacing every pair of numbers by its average (a coarser level approximation, giving s_j) and the difference between the two numbers (the detail d_j). The reconstruction (12) then adds the sum and difference to recover the first number, whereas a subtraction gives the second number in every pair. The resulting algorithm is fast: starting from a sequence s_0 with N entries, we compute sums and differences for $\frac{N}{2}$ pairs to obtain s_1 and d_1. The $\frac{N}{2}$ entries in s_1 give $\frac{N}{4}$ pairs, for each of which we have another sum and difference to compute, and so on. The total number of computations is therefore $2\frac{N}{2} + 2\frac{N}{4} + \cdots \simeq 2N$, (where we have swept edge effect terms under the rug if N is not a power of 2, but they don't matter here: their contribution to the complexity is $O(\log N)$). If we have K non-vanishing h_n, g_n instead of only 2, then the total number of computations is KN, still linear in N. This type of wavelet transform has therefore a lower complexity than the FFT, which uses $O(N \log N)$ computations.

5. What do Wavelets Buy You?

So we have a fast algorithm for a neat kind of basis, in which all basis functions are shifted and dilated versions of just one template (or a few templates, in some generalizations or in higher dimensions). Why should anyone care? In fact, surprisingly many people do care, and many fields have something to tell us about these wavelet bases.

To harmonic analysts, wavelet bases are a convenient way to carry out a Littlewood-Paley (LP) decomposition. In a traditional LP decomposition of a function f, one writes

$$f = \sum_{j=-\infty}^{\infty} \Delta_j f = f_0 + \sum_{j=0}^{\infty} \Delta_j f \quad ,$$

where the Fourier transform $(\Delta_j f)\hat{}(\xi)$ of each $\Delta_j f$ is non-vanishing only for, say, $2^{j-1} \leq |\xi| \leq 2^{j+1}$. One way of obtaining such $\Delta_j f$ is to construct a smooth function w, supported on $\frac{1}{2} \leq |\xi| \leq 2$, such that, for $1 \leq |\xi| \leq 2$, $w(\xi)+w(\xi/2) = 1$, and to define $(\Delta_j f)\hat{}(\xi) = \hat{f}(\xi) w(2^{-j}\xi)$. The different $\Delta_j f$ decouple different frequency ranges of f; yet, unlike the Fourier transform itself, they retain some spatial information. This information is sufficient, for instance, to characterize the Hölder spaces C^s: even though it is impossible to characterize (i.e. give an "if and only if" condition) the Hölder exponent of f by the decay of its Fourier transform, nevertheless decay conditions on the $\Delta_j f$, as a function of their frequency range label j, permit such a characterization. More precisely, for any $f \in L^\infty$, we have

$$f \in C^s \Leftrightarrow \sup_{j \in \mathbb{N}} 2^{js} \parallel \Delta_j f \parallel_{L^\infty} < \infty \quad . \tag{14}$$

Similarly, LP decompositions can be used for much more sophisticated estimates (Stein (1993), Frazier *et al.* (1990)). A wavelet decomposition carves up f likewise in dyadic frequency blocks, with $Q_{-j} f := \sum_k < f, \psi_{-j,k} > \psi_{-j,k}$ corresponding to

$\Delta_j f$. This means that many achievements of LP decompositions have their mirror image in wavelet terms. For instance, if the wavelet ψ and the scaling function ϕ are in C^r and have sufficiently rapid decay, then we have, for all $s < r$, a characterization of the Hölder spaces similar to (14). Specifically, for $f \in L^\infty$,

$$f \in C^s \Leftrightarrow \sup_{j \in \mathbb{N}} 2^{j(s+\frac{1}{2})} \sup_{k \in \mathbb{Z}} | < f, \psi_{-j,k} > | < \infty \quad . \tag{15}$$

The similarity with (14) is obvious (the extra $\frac{1}{2}$ in the exponent is due to the normalization we chose for the $\psi_{j,k}$ in (1)); more sophisticated estimates using LP-type decompositions translate into wavelet estimates analogously. Wavelets then provide a way to write powerful techniques in harmonic analysis in a language that can also be read as an algorithm. On the other hand, their very convenient orthogonality properties in L^2 also lead to shortcuts in proofs in harmonic analysis (see e.g. Meyer (1990)).

To electrical engineers, wavelet bases are a mathematical framework that links up with a filtering technique developed earlier, called subband filtering. Diagram (11) is in fact the electrical engineering notation for a filter bank with two channels, one *low-pass* (the transition $s_{j-1} \to s_j$) and one *high-pass* ($s_{j-1} \to d_j$); the downsampling makes this a *critically downsampled* filtering procedure, meaning that after the operation we end up with exactly as many entries as before. Diagram (13) then means that we have in fact a *perfect reconstruction, critically downsampled 2-channel* filter bank. The standard reference in electrical engineering for such filter banks is Smith & Barnwell (1986); similar constructions also appear in Mintzer (1985) and Vetterli (1986). Before these perfect reconstruction filter banks, electrical engineers had constructed similar filter banks that were almost perfect, in the sense that the reconstructed sequence is very close to the original. Such near-perfect filter banks are still designed and used for many applications; giving up perfect reconstruction leads to more degrees of freedom in the design and, if things are done right, to perceptually equally good results. All this was developed without any input from mathematicians, with the result that electrical engineers sometimes and understandably feel that the present popularity of wavelets gives a lot of "undeserved" credit to mathematicians for re-inventing the wheel while engineers were already driving cars. This view would be correct if there were nothing more to wavelets than the algorithm. The realization that the perfect reconstruction banks are linked to a rich underlying mathematical structure, associated with powerful and deep mathematical theorems is a different matter, however. Even for applications of interest to electrical engineering, this link has led to new applications that use the mathematical insights, and that would not have been developed from only the subband filtering concept (examples are Mallat & Hwang (1992), Wornell & Oppenheim (1992)).

To the computer scientist or engineer interested in studying vision, the multiresolution analysis framework, with its different levels of detail, is very reminiscent of multi-scale models in vision analysis, such as Witkin (1983), or in a more algorithmic version, the pyramids of Burt & Adelson (1983). (As mentioned above, it was S. Mallat's background in vision theory that inspired him to re-interpret wavelet bases via the mathematical concept of multiresolution analysis.)

Independently of and in parallel with the wavelet development, E. Adelson had in fact already switched from the (redundant) pyramid schemes to cascaded subband filtering for image analysis (see Adelson *et al.* (1987)).

Approximation theorists also recognized familiar concepts in wavelet theory: the space V_j, with their varying degrees of resolution, are basic standard fare in approximation theory. The example in (5) of spline spaces V_j really stems from approximation theory (de Boor (1978)). Similarly, formulas (10) and (12) are reminiscent of subdivision schemes, a technique developed to generate smooth curves and surfaces (Cavaretta *et al.* (1991)), Dyn *et al.* (1987)). There is a "philosophical" difference between many theorems in approximation theory and, for instance, the way function spaces are characterized via wavelets. Wavelet coefficients $< f, \psi_{j,k} >$ capture the *difference* between the successive approximations $P_{j-1}f$ and $P_j f$, rather than studying f via the $P_j f$. Similarly, subdivision schemes typically contain in their coarse-to-fine-formulas, only the first term in the right hand side of (12). Nevertheless, approximation theorists immediately recognized the kinship of wavelets, and are very active in the field now.

Wavelet techniques have some similarity as well with the multipole algorithms developed by Rokhlin (1985) for fast numerical computations. (See also Rokhlin's presentation in this ICM.) In a multipole expansion, the quantity to be computed (such as, e.g., the total potential energy of a completely arbitrary distribution of a large number of particles) is taken apart into many contributions living on different scales; for many of them, a coarse scale description suffices. Moreover, the taking apart is done in a hierarchical way. All this results in a fast algorithm. The wavelet algorithms in Beylkin *et al.* (1991) and subsequent work (e.g. Beylkin (1993)) work on the same principle as these fast multipole expansion techniques.

Finally, and as promised, wavelets buy you a time-frequency decomposition: once a function f is decomposed as in (3), it is written as a superposition of building blocks, the wavelets $\psi_{j,k}$, each of which is well localized in frequency (in a frequency band of width proportional to 2^{-j}, i.e. $2^{-j}\alpha \leq |\xi| \leq 2^{-j}\beta$) and in time (around the position $2^j k$, with a resolution proportional to k). Note that this means that high-frequency wavelets have very sharp time resolution, whereas low-frequency wavelets are much more spread out in time but have sharp frequency resolution. A decomposition of this type is well suited to signals f that consist of short-lived high-frequency *transients* superposed on more placid longer-lived low-frequency components. Many signals are of this type. But many more are really more complicated, and require a battery of tools of which wavelets are only one; we shall come back to this later.

6. Back to the Algorithm

For many applications, the powerful mathematical properties of wavelets can be exploited only if the associated algorithm is truly efficient. We saw earlier that the total complexity of a decomposition into orthonormal wavelets is KN, if there are K non-vanishing h_n, g_n in the associated filters. However, many "natural" orthonormal wavelet bases correspond to filters with infinitely many non-vanishing

h_n, ruining the complexity estimate. This is the case, for instance, if the V_j are taken to be spline spaces of higher order than 1. In this case, the most natural choice for the function ϕ, the translates $\phi(x-n)$ of which should span all of V_0, would seem to be the B-spline function, obtained by convolving $\chi_{[0,1]}$ with itself $k-1$ times (for splines of order k). For this choice, the $\phi(x-n)$ are not orthogonal, so that ϕ needs to be replaced by an "orthogonalized" version, which is however now supported on all of $I\!R$ (with exponential decay), leading to infinitely many $h_n \neq 0$.

So how does one get MRA with finitely many non-vanishing h_n? The answer lies in the filtering approach from electrical engineering. If one takes (9), (10), and (12) as the point of departure, rather than as a corollary of the MRA structure, then one finds easily that the h_n should satisfy

$$\sum_n h_n h^*_{n+2k} = \delta k, 0 \quad , \tag{16}$$

while the g_n can be chosen as

$$g_n = (-1)^n h^*_{-n+1}$$

Smith & Barnwell (1986) had found ways to construct finite sequences h that satisfy (16). Such sequences do not necessarily correspond to an L^2-function ϕ such that (6) holds, however; necessary and sufficient conditions for this correspondence were found by Cohen (1990). If these conditions are satisfied, then there always exists an MRA associated with h. To obtain high approximation order for the MRA-ladder and smoothness, one needs to impose additional conditions on the sequence h, of the form

$$\sum_n (-1)^n h_n n^l = 0 \qquad l = 0, \ldots, L-1 \quad . \tag{17}$$

Daubechies (1988) constructs such finite sequences h and proves that by this method one can obtain compactly supported ϕ, ψ which are C^k, where k is arbitrarily large (but finite). These functions $\phi(x)$, $\psi(x)$ are not given by an explicit analytic expression, although the Fourier transform of ϕ can be written as an infinite product,

$$\hat{\phi}(\xi) = \hat{\phi}(0) \prod_{j=1}^{\infty} m_0(2^{-j}\xi) \quad , \tag{18}$$

with $m_0(\xi) = 2^{-1/2} \sum_n h_n e^{-in\xi}$.

One can use (6) to make a detailed study of their different, sometimes intriguing, properties. For instance, it turns out that the Hölder exponent of ϕ in a point x in its support depends on the frequency of the digits 1 and 0 in the binary expansion of x, as shown in Daubechies & Lagarias (1992); this means that these ϕ have *multifractal* properties (Daubechies & Lagarias (1994); see also Jaffard (1994)).

Although the wavelet bases constructed in Daubechies (1988) have been used in various applications, they are by no means ideal for all circumstances, and many other constructions have been carried out which improve on them in some respects, while giving up on other properties. For instance, one can give up some of the orthogonality in the constructions above, and construct a Riesz basis rather than an orthonormal basis of wavelets (together with the dual Riesz basis), as in Chui & Wang (1991), Chui & Wang (1992), Auscher (1989), or Cohen *et al.* (1992); this relaxing of orthonormality buys more smoothness and/or symmetry for the wavelets. Another useful construction restricts these wavelet bases to an interval while retaining their powerful mathematical properties (see e.g. Cohen *et al.* (1993), and Andersson *et al.* (1994)). Not all applications require absolutely that the filter h be finite; if m_0, defined as in (18), can be written as the quotient of two trigonometric polynomials, then there still exist fast algorithms to implement convolution with h, and such filters and wavelet bases have been proposed as well (Lemarié & Malgouyres (1989), Evangelista (1992), Herley & Vetterli (1993))–in fact, the original construction by Stromberg (1982) falls into this class.

7. Higher Dimensions

So far, we have been working in one dimension only. There exist several possible generalizations to higher dimensions. Usually they involve several wavelets $\psi^1, \ldots, \psi^K$, and the wavelet basis is then given by the collection $\psi^n_{j,k}(x)2^{-jd/2}\psi^n(2^{-j}x-k), j \in \mathbb{Z}, k \in \mathbb{Z}^d, n = 1, \ldots, K$. The easiest construction starts from a one-dimensional multiresolution analysis, with scaling function ϕ and wavelet ψ, and uses these to build one scaling function Φ and 2^d-1 wavelets Ψ^k in d-dimensions, by taking products of $\phi(x_k)$ and $\psi(x_m)$. For $d=2$, for instance, one takes $\Phi(x_1,x_2) = \phi(x_1)\phi(x_2)$, $\Psi^1(x_1,x_2) = \psi(x_1)\phi(x_2)$, $\Psi^2(x_1,x_2) = \phi(x_1)\psi(x_2)$, $\Psi^3(x_1,x_2) = \psi(x_1)\psi(x_2)$. This corresponds to a two-dimensional multiresolution where the spaces $\mathbf{V}_j$ are tensor products $V_j \otimes V_j$, and the $\Psi^n_{j,k}, k \in \mathbb{Z}^2, n = 1,2,3$ then exactly span $\mathbf{W}_j$, the orthogonal complement of $\mathbf{V}_j$ on $\mathbf{V}_{j-1}$. The higher-dimensional Ψ^k and Φ inherit, of course, recursion relations similar to (6) and (7) from their one-dimensional progenitors, so that the algorithms remain basically as simple as in one dimension. There exist other, fancier constructions as well, with "non-separable" higher-dimensional wavelets, possibly with a dilation matrix A replacing the simple scaling by 2, but the simple tensor product multiresolution analysis above is the most used. One can also introduce special bases of multi-dimensional wavelets, such as the divergence-free wavelet bases of Battle & Federbush (1993) or Lemarié-Rieusset (1992), useful for decomposing divergence-free vector fields.

In most of what follows, I will stick to the one-dimensional notation, but all statements (unless qualified) will be true for these d-dimensional wavelets as well.

8. Mathematical Properties

A first important property of wavelet bases is that they provide unconditional bases for many classical function spaces. A family of functions $\{g_\alpha; \alpha \in A\}$ is an unconditional basis for a Banach space $B \subset S'$ if it is a Schauder basis and there exists a criterion to decide whether $f \in B$ by using only the *absolute values* $|< f, g_\alpha >|, \alpha \in A$. Equivalently, the g_α constitute an unconditional basis if, whenever $\sum_{\alpha \in A} c_\alpha g_\alpha \in B$, multiplying the coefficients c_α with arbitrarily chosen $\epsilon_\alpha = \pm 1$ always leads to another element of B, i.e. $\sum_{\alpha \in A} \epsilon_\alpha c_\alpha g_\alpha \in B$. It turns out that the orthonormal wavelet bases (or more generally, Riesz bases of wavelets) give such unconditional bases for L^p $(1 < p < \infty)$, the Sobolev spaces W^s, the Besov spaces $B_q^{p,s}$, the Hölder spaces C^s, as well as for the Hardy space H^1 and its dual BMO (see Meyer (1990)). For instance, (15) gives a characterization of $f \in C^s$, using only the $|< f, \psi_{j,k} >|$, if we know a priori that $f \in L^\infty$. This last requirement can be dropped if we also impose that $\sup_{k \in \mathbb{Z}} |< f, \phi_{0,k} >| < \infty$; this then means that $\{\phi_{0,k}; k \in \mathbb{Z}\} \cup \{\psi_{-j,k}; k \in \mathbb{Z}, j \in \mathbb{N}\}$ is an unconditional basis for the (inhomogeneous) Hölder space C^s (provided $\phi \in C^r$ with $r > s$).

Intuitively speaking, wavelet expansions do so well, in such a variety of frameworks, because their smoothness allows them to adjust well to smooth functions (or to smooth portions of functions), their scaling properties allow them to "zoom in" on singularities, and their good spatial concentration allows them to handle decay well.

It follows that wavelet expansions for a function f can converge in many different topologies (depending on which spaces f belongs to). They can converge in even other ways as well. For instance, if we restrict ourselves to an interval, and order the wavelet basis properly (exhausting every scale first before moving on to the next finer scale), then the truncated sums of the correspondingly ordered wavelet expansions will converge in L^1 on the interval. (L^1 does not have unconditional bases, so some ordering is necessary.) When looking at pointwise convergence, one finds easily (provided ϕ and ψ satisfy a minimum of decay and smoothness conditions, as always) that the wavelet expansion of f converges in all points of continuity of f. It is also true that for L^2-functions f, the wavelet expansion of f converges pointwise almost everywhere (more precisely: in every Lebesgue point of f). For compactly supported ϕ, this last point follows from standard harmonic analysis arguments once one realizes that $\sup_j \sum_k |< f, \phi_{j,k} >| |\phi_{j,k}(x)|$ is essentially a maximal function for f, bounded above (up to a constant factor) by the standard Hardy-Littlewood maximal function. The result is also true for less constrained ϕ (Auscher (1989), Kelly *et al.* (1994), and it carries over (as usual) to other L^p-spaces as well.

9. Applications

Among the many successful applications of wavelets, only a few can be presented here. Particularly attractive are those (at least to me) where the mathematical properties of wavelets play an essential role in their effectiveness. A first example was the matrix or operator compression in Beylkin *et al.* (1991). The matrices $A_{i,j}$ they consider are finely sampled versions, $A_{i,j} = K(i\alpha, j\alpha)$ of an integral kernel $K(x, y)$ corresponding to a Calderón-Zygmund operator, i.e. K satisfies bounds of the type $|K(x, y)| \leq C|x - y|^{-1}, |\partial_x K(x, y)| + |\partial_y K(x, y)| \leq C|x - y|^{-2}$ (with often similar bounds for higher order derivatives). For the matrix $A_{i,j}$, this means that the matrix elements vary smoothly with i, j as long as (i, j) stays away from a region around the diagonal; near the diagonal wilder behavior is allowed. Replacing the sequence $A_{i,j}$ by its wavelet coefficients (obtained by "filtering" in both horizontal and vertical directions, with the fast algorithms explained above) results in a new matrix in which the majority of entries are exceedingly small. Thresholding them by ϵ (i.e. the entries smaller than ϵ are replaced by 0) gives a sparse matrix, so that computing (a truncated version of) the action of A on a vector can be done much faster. The beauty is that one can actually control the damage done by thresholding–not a trivial matter, since a large number of small errors can still add up to a sizeable total error. If the thresholding is done a little bit more carefully than by simple truncation (some sum rules need to be respected), then Beylkin, Coifman, and Rokhlin proved that the truncated matrix $A_\epsilon^{\text{trunc}}$ obtained by thresholding and then returning, via the inverse algorithm, to the "real world" from the "wavelet coefficient world," satisfies $\| A - A_\epsilon^{\text{trunc}} \| \leq C\epsilon^\gamma$ in L^2*-operator* norm, with C, γ independent of ϵ and also independent of the size of the matrix. The proof essentially repeats the argument of the "T(1) theorem" by David & Journé (1984).

The orthonormality of wavelet bases, as well as their different scales, are exploited by **?** in an application closer to physics. They use wavelets as a tool to generate random velocity fields that accurately model fractal, self-similar fields, important in turbulent diffusion.

Applications to a very different field (although, strictly speaking, not of wavelet bases but of another type of wavelet representation) can be found in the work by Mallat & Hwang (1992). In one application, they seek to remove noise from very noisy images. This noise has particularly large effects on the fine scale wavelet coefficients, which also contain the information necessary to keep "sharp" edges in the image–discarding these corrupted fine scale coefficients altogether would result in a less noisy image, but it would also look blurred. Mallat and Hwang exploit the characterization of singularities given by the rate at which local wavelet coefficients decay as a function of scale, to sort out the chaff from the grain in the fine scale coefficients, leading to a restored denoised image with sharp edges.

Yet a different set of applications is in the work of Donoho (1993). He also discusses denoising. The starting point is a function f, supposed to belong to a Banach space B (which describes the class of problems of interest in a particular application); f is known only through noisy samples or estimations. Suppose that $(g_\alpha)_{\alpha \in A}$ is an unconditional basis for B. Then the data for f can be translated into noisy estimates for the coefficients of the expansion of f into the g_α. The denoising consists in a thresholded shrinking of these coefficients (all the ones below a threshold are set to zero, the ones above the threshold are multiplied with a nonzero coefficient < 1 depending on their size) and reconstruction. Donoho proves that if the g_α constitute an unconditional basis for B, then the worst-case error for this method cannot be significantly larger than the worst-case error for any other method, however fancy. Because wavelet bases are unconditional bases for many function spaces, they provide therefore a near-optimal method for a large variety of frameworks.

Wavelet bases are also, because of their adaptivity, a good tool to use in nonlinear approximation of e.g. piecewise smooth functions; see e.g. DeVore *et al.* (1992), Donoho (1993). Linear approximation theory discusses how well successive truncations of an expansion approach the desired function. For instance, if $(g_n)_{n\in\mathbb{N}}$ is a basis for B, then linear approximation is concerned with the behavior, as a function of N, of $\text{dist}_B(f, \Sigma_N)$, where Σ_N is the linear subspace $\Sigma_N = \{f = \sum_{n=1}^{N} c_n g_n; c_n \in \mathbb{C}\}$. In nonlinear approximation, the N-th approximation of f still involves N terms, but they need not correspond to the first N basis functions. That is, one studies $\text{dist}_B(f, S_N)$, where $S_N = \{f = \sum_{n\in I_{f,N}} c_n g_n; c_n \in \mathbb{C}, \#I_{f,N} = N\}$; S_N is no longer a linear subspace of B. An example of how this affects things: if f is a piecewise C^s function with good decay, and possibly discontinuities between the pieces, and if we choose a wavelet basis (with $\phi, \psi \in C^r$ with $r > s$), then $\text{dist}_{L^2}(f, \Sigma_N) \sim CN^{-1/2}$, but $\text{dist}_{L^2}(f, S_N) \sim CN^{-s}$: the nonlinear approximation does not suffer from the presence of the discontinuities. In contrast, if one chooses a Fourier basis, one finds that both $\text{dist}_{L^2}(f, \Sigma_N)$ and $\text{dist}_{L^2}(f, S_N)$ decay like $N^{-1/2}$.

10. Shortcomings of Wavelets

In spite of all their good qualities, wavelets are, of course, not the universal panacea. They are markedly inefficient for coherently oscillating components. Wavelet bases also suffer from being very translationally non-invariant, and no entirely satisfactory solution has been found, so far, to deal with boundary problems in higher dimensions for non-rectangular domains. Other recently developed harmonic analysis tools are much better at dealing with oscillations: wavelet packets and localized trigonometric bases.

11. Wavelet Packets

The algorithm that we sketched above for a decomposition into wavelets consists of concatenating diagram (11) several times, starting a new stage from the preceding "s_j" output. The "d_j"-branches are left untouched. We could also choose to attach another splitting diagram (11) to the "d_j"-branches; this still results in fast algorithms, corresponding to a decomposition into different functions, called wavelet packets. The wavelet bases we saw before are just one (extreme) example of wavelet packet bases. As explained before, the wavelet bases correspond to a Littlewood-Paley decomposition: in the frequency domain, $\widehat{\psi_{j,k}}(\xi)$ is essentially concentrated in and near the region $2^j\pi \leq |\xi| \leq 2^{j+1}\pi$. When the extra splittings are introduced that lead to wavelet packets, they correspond to further splits of these frequency blocks. One can, for instance, choose to keep splitting the branch of the wavelet algorithm diagram that would normally have ended in the "d_j"; if we split j times, at every intermediate step splitting all the subbranches that have been sprouted from the d_j-branch, then we will have subdivided the region $2^j\pi \leq |\xi| \leq 2^{j+1}\pi$ into 2^j subregions. If we do this for all $j \geq 0$, we end up with wavelet packets that all have the same "width," for their Fourier transforms as well as in "physical" space; these are therefore much closer to a standard windowed Fourier type basis than to the dyadic frequency decomposition given by wavelets. By choosing to split fewer times, one can generate a wide variety of wavelet packet bases that are intermediary between the "pure" wavelet bases and these Fourier-type wavelet packet bases.

Among all these bases, one can adaptively choose the one that is most "efficient" for a given function f (meaning, coarsely speaking, that the decomposition into this basis is achieved by a few large coefficients that represent most of the L^2-norm of f, with a small "tail" in the other coefficients) by basing the decision whether or not to split, at every step in the algorithm, on the results obtained for f. Detailed descriptions of these wavelet packet bases, first constructed by Coifman and Meyer, and of their mathematical properties and the associated algorithms can be found in Coifman *et al.* (1992), Coifman & Wickerhauser (1993), Wickerhauser (1994), and references therein. Note that when many splittings are carried out, the carving up of the frequency domain is not really as "clean" as the description above indicates; see Coifman *et al.* (1992).

12. Localized Trigonometric Bases

Wavelet packet bases are already much better at dealing with oscillations than wavelets. Even better are the localized cosine or sine bases constructed by Coifman & Meyer (1991), related to the independently constructed overlapped cosine transforms of Malvar (1990). These are orthogonal bases of the type

$$f_{k,l}(t) = w_k(t)\sin(\Omega_{k,l}t)$$

where the functions w_k are window functions, well-localized in space (e.g., with compact support) but with possibly varying widths W_k, and the $\Omega_{k,l}$ are a corresponding discrete sequence of frequencies; in first approximation, the $\Omega_{k,l}$ behave like $\pi l/W_k$. More precisely, every $w_k(t)$ is supported on an interval $[a_k - \epsilon_k, a_{k+1} + \epsilon_{k+1}]$, and is $\equiv 1$ on the smaller interval $[a_k + \epsilon_k, a_{k+1} - \epsilon_{k+1}]$; here we assume $\cdots < a_{k-1} < a_k < a_{k+1} < \cdots$, with the ϵ_j chosen so that, for all k, $a_k + \epsilon_k \leq a_{k+1} - \epsilon_{k+1}$. In the transition regions $[a_k - \epsilon_k, a_k + \epsilon_k]$, the window functions w_k and w_{k-1} must satisfy the complementarity requirement $w_{k-1}^2(x) + w_k^2(x) = 1$ as well as the symmetry condition $w_{k-1}(a_k - t) = w_k(a_k + t)$ (for $|t| \leq \epsilon_k$). The width W_k is then defined as $W_k = a_{k+1} - a_k$, and the $f_{k,l}$ are given by

$$f_{k,l}(t) = (2/W_k)^{1/2} w_k(t) \sin[\frac{\pi}{W_k}(l + \frac{1}{2})(t - a_k)] \quad .$$

It is quite surprising that the functions w_k and the frequencies $\Omega_{k,l}$ can be chosen in such a way that the $f_{k,l}$ are all smooth (even C^∞) and nevertheless provide an orthonormal basis for $L^2(\mathbb{R})$. The construction is ingenious, but it doesn't use any modern techniques–this construction could have been carried out in the 18th century, and maybe the biggest surprise is that it wasn't. A remarkable feature of the construction is that neighboring window functions can be "merged," leading to the replacement of the $f_{k,l}$ and $f_{k+1,l'}$ by different functions $\tilde{f}_{k,l''}$; together with the remaining (and untouched!) $f_{n,l}(n < k \text{ or } n > k+1)$ these then provide a different orthonormal basis. As in the case of wavelet packets, this choice between two options (to merge or not) can be exploited to construct a whole family of different bases, all "living" within one fast algorithm, so that the "best basis" can be chosen adaptively. See Coifman & Wickerhauser (1993), or Wickerhauser (1994).

13. Libraries of Bases

In practice, functions are usually quite complicated, and even these "best basis" algorithms do not necessarily give the most efficient decomposition. A simple example is a nicely oscillating function with just one superposed spike–the oscillations are best represented with a localized trigonometric basis or a wavelet packet basis, whereas the spike is "asking for" a wavelet representation. To address this, Mallat proposed a "pursuit" algorithm (Mallat & Zhang (1993)), adapted by Coifman and Meyer into an algorithm using libraries of bases. Based on the function to be decomposed, one first selects the best basis from a library which can contain wavelet packets, various localized trigonometric bases, and other possible bases as well (as long as they are associated with fast algorithms). One monitors the coefficients computed for a decomposition in this basis, ranked by decreasing size. Beyond a certain threshold (which can depend on the total norm of the remaining tail, or the slowing down of the decay rate of the coefficients in this tail), one calls it quits–the selected basis was good for the first components but may not be optimal now. Reconstructing the first components and subtracting from the original leads to a remainder, for which one starts anew: again a best basis is selected, and one sticks to this basis until it becomes less satisfactory, etc .... This process can be repeated several times (see Coifman & Wickerhauser (1993)). This type of approach leads to very flexible and efficient time-frequency, or phase space decompositions.

14. Conclusion

In the last ten years, mathematical tools have emerged that combine insights from harmonic analysis with fast algorithms. They turn out to be very powerful for many applications, especially when used in conjunction with each other, and in combination with many existing tools. Not suprisingly, they can be linked with many other earlier insights in a variety of fields; one way of viewing them is as the synthesis of these varied strands. The result of this synthesis is more than just the sum of its parts, and as these new tools are becoming a familiar part of many a researcher's toolbox, they will turn up in many applications.

References

Adelson, E.H., Simoncelli, E.P., & Hingorani, R. 1987. Orthogonal pyramid transforms for image coding. *Proc. SPIE, Cambridge*, **845**(Oct.), 50–58.

Andersson, L., Hall, N., Jawerth, B., & Peters, G. 1994. Wavelets on closed subsets of the real line. *In:* Schumaker, Larry L., & Webb, G. (eds), *Topics in the Theory and Applications of Wavelets*. Boston: Academic Press.

Auscher, P. 1989. *Ondelettes fractales et applications*. Ph.D. thesis, Université de Paris IX – Dauphine, Paris, France.

Auscher, P. 1992. Toute base d'ondelettes régulières de $L^2(R)$ est issue d'une analyse multirésolution régulière. *C. R. Acad. Sci. Paris, Série 1*, **315**, 1227–1230.

Battle, G. 1987. A block spin construction of ondelettes, Part I: Lemarié functions. *Comm. Math. Phys.*, **110**, 601–615.

Battle, G., & Federbush, P. 1993. Divergence-free wavelets. *Michigan Math. J.*, **40**, 181.

Beylkin, G. 1993. Wavelets and fast numerical algorithms. *Pages 89–117 of:* Daubechies, I. (ed), *Different perspectives on wavelets*. AMS Proceedings of Symposia in Applied Mathematics, no. 47.

Beylkin, G., Coifman, R., & Rokhlin, V. 1991. Fast wavelet transforms and numerical algorithms. *Comm. Pure and Appl. Math.*, **44**, 141–183.

Burt, P., & Adelson, E.H. 1983. The Laplacian pyramid as a compact image code. *IEEE Trans. Communications*, **31**, 482–540.

Cavaretta, A.S., Dahmen, W., & Micchelli, C.A. 1991. Stationary subdivision. *Memoirs of the Am. Math. Soc.*, **93**, 1–186.

Chui, C.K., & Wang, J.Z. 1991. A cardinal spline approach to wavelets. *Proc. Amer. Math. Soc.*, **113**, 785–793.

Chui, C.K., & Wang, J.Z. 1992. On compactly supported spline wavelets and a duality principle. *Trans. Amer. Math. Soc.*, **330**, 903–915.

Cohen, A. 1990. *Ondelettes, analyses multirésolutions et traitement numérique du signal*. Ph.D. thesis, Université Paris – Dauphine. Also *Ondelettes et traitement numérique du signal*, (1992), Masson, Paris; English translation to appear with Chapman and Hall, London.

Cohen, A., Daubechies, I., & Feauveau, J.C. 1992. Biorthogonal bases of compactly supported wavelets. *Comm. Pure & Appl. Math.*, **45**, 485–560.

Cohen, A., Daubechies, I., & Vial, P. 1993. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, **1**, 54–81.

Coifman, R., & Meyer, Y. 1991. Remarques sur l'analyse de Fourier à fenêtre. *C. R. Acad. Sc. Paris. I*, **312**, 259–261.

Coifman, R., & Wickerhauser, M.V. 1993. Wavelets and adapted waveform analysis: a toolkit for signal processing and numerical analysis. *Pages 119–153 of:* Daubechies, I. (ed), *Different perspectives on wavelets.* AMS Short Course Series.

Coifman, R., Meyer, Y., & Wickerhauser, M.V. 1992. Size properties of wavelet packets. *Pages 453–470 of:* M.B. Ruskai, et al. (ed), *Wavelets and their applications.* Boston: Jones and Bartlett.

Daubechies, I. 1988. Orthonormal bases of compactly supported wavelets. *Comm. Pure & Appl. Math.*, **41**, 909–996.

Daubechies, I., & Lagarias, J.C. 1992. Two-scale difference equations, II. Local regularity, infinite products of matrices and fractals. *SIAM J. Math. Anal.*, **23**(4), 1031–1079.

Daubechies, I., & Lagarias, J.C. 1994. On the thermodynamic formalism for multifractal functions. *Reviews of Mathematical Physics*, **6**, 1033–1070.

David, G., & Journé, J.L. 1984. A boundedness criterion for generalized Calderón-Zygmund operators. *Ann. of Math.*, **120**, 371–397.

de Boor, C. 1978. A practical guide to splines. *In: Applied Math. Science Series nr. 27.* New York: Springer.

DeVore, R.A., Jawerth, B., & Popov, V. 1992. Compression of wavelet decompositions. *Amer. J. Math.*, **114**, 737–785.

Donoho, D. 1993. Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, **1**, 100–115.

Dyn, N., Gregory, A., & Levin, D. 1987. A 4-point interpolatory subdivision scheme for curve design. *Computer Aided Geometric Design*, **4**, 257–268.

Evangelista, G. 1992. Wavelet transforms and digital filters. *Pages 396–412 of:* Meyer, Y. (ed), *Wavelets and Applications.* Paris: Masson.

Frazier, M., Jawerth, B., & Weiss, G. 1990. Littlewood-Paley theory and the study of function spaces. CBMS – Conference held at Auburn University in 1989, published by AMS.

Herley, C., & Vetterli, M. 1993. Wavelets and recursive filter banks. *IEEE Trans. Sign. Proc.*, **41**, 2536–2556.

Jaffard, S. 1994. *Multifractal formalism for functions. I & II.* Preprints, CERMA, Ecole Nationale des Ponts et Chaussées, Noisy-le-Grand, France.

Kelly, S.E., Kon, M.A., & Raphael, L.A. 1994. Local convergence for wavelet expansions. *Journal of Functional Analysis*, **126**(1), 102–138.

Lemarié, P.G. 1988. Une nouvelle base d'ondelettes de $L^2(\mathbb{R}^n)$. *J. de Math. Pures et Appl.*, **67**, 227–236.

Lemarié, P.G. 1993. Sur l'existence des analyses multirésolutions en théorie des ondelettes. *Revista Matematica Iberoamericana*, **8**, 457–474.

Lemarié, P.G., & Malgouyres, G. 1989. Ondelettes sas peine. *Technical Memorandum.*

Lemarié-Rieusset, P.G. 1992. Analyses multi-résolutions non orthogonales, commutation entre projecteurs et dérivation, et ondelettes vecteurs à divergence nulle. *Rev. Mat. Iberoamericana*, **8**, 221–237.

Mallat, S. 1989. Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, **315**, 69–88.

Mallat, S., & Hwang, W.L. 1992. Singularity detection and processing with wavelets. *IEEE Trans. Information Theory*, **38**, 617–643.

Mallat, S., & Zhang, Z. 1993. Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, **41**, 3397–3415.

Malvar, H. 1990. Lapped transforms for efficient transform/subband coding. *IEEE Trans. ASSP*, **38**, 969–978.

Meyer, Y. 1985. Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs. *Séminaire Bourbaki*, 1985–1986. nr. 662.

Meyer, Y. 1990. *Ondelettes et opérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III: Opérateurs multilinéaires.* Paris: Hermann. English translation prepared by Cambridge University Press.

Mintzer, F. 1985. Filters for distortion-free two-band multirate filter banks. *IEEE Trans. Acoust. Speech Signal Process.*, **33**, 626–630.

Rokhlin, V. 1985. Rapid solution of integral equations of classical potential theory. *Journal of Computational Physics*, **60**, 187.

Smith, M.J., & Barnwell, D.P. 1986. Exact reconstruction for tree-structured subband coders. *IEEE Trans. ASSP*, **34**, 434–441.

Stein, E. 1993. *Harmonic Analysis: real-variable methods, orthogonality, and oscillatory integrals.* Princeton, NJ: Princeton University Press.

Stromberg, J.O. 1982. A modified Franklin system and higher order spline systems on $\mathbb{R}^n$ as unconditional bases for Hardy spaces. *Pages 475–493 of:* W. Beckner, et al. (ed), *Conference in Harmonic Analysis in Honor of A. Zygmund*, vol. II. Belmont, California: Wadsworth Math. Series.

Vetterli, M. 1986. Filter banks allowing perfect reconstruction. *Signal Proc.*, **10**, 219–244.

Wickerhauser, M.V. 1994. *Adapted wavelet analysis from theory to software.* Wellesley, MA: AK Peters.

Witkin, A. 1983. Scale space filtering. *In: Proc. Int. Joint Conf. Artificial Intell.*

Wornell, G.W., & Oppenheim, A.V. 1992. Wavelet-based representations for a class of self-similar signals with application to fractal modulation. *IEEE Trans. Information Theory*, **38**, 785–800.