Odyssey Enabled Document Delivery

TN: 727710

5/17

Journal: **Proceedings of the 1987 International Workshop on Wavelets and Applications**

Item #:

Date: **1987** 38-66

Vol: **()** Pgs: ~~**All pages**~~

Article Title: Orthonormal bases of wavelets with finite support - connection with discrete filters

Author: I. Daubechies

Imprint:

CUSTOMER HAS REQUESTED:

Turner, Byron
(bct9)

email:byron@math.duke.edu

Phone: **919-660-2844**

Ariel: **152.3.237.118**

Perkins DDS

NDD
Box 90183
Durham, NC 27708

Phone: 919-660-5890
Fax: 919-660-5964

If there are problems with this document:

Missing Page(s):______________________________
Edge(s) Cut Off:______________________________
Unable/Difficult to Read:______________________

Please contact us.

NDD

Orthonormal Bases of Wavelets with Finite Support – Connection with Discrete Filters

*I. Daubechies**

AT&T Bell Laboratories, 600 Mountain Avenue,
Murray Hill, NJ 07974, U.S.A.

Abstract: We define wavelets and the wavelet transform. After discussing their basic properties, we focus on orthonormal bases of wavelets, in particular bases of wavelets with finite support.

Contents:

1. The wavelet transform: Continuous and discrete versions.

As proposed by J. Morlet [1], the wavelet transform of a function f is given by

$$\phi_{\mathrm{Wav},f}(a, b) = \frac{1}{|a|^{1/2}} \int dx\, h\left(\frac{x-b}{a}\right) f(x)\,, \tag{1.1}$$

where h is the basic "wavelet". The parameters a, b can be chosen to vary either continuously ($a, b \in \mathbb{R}$, with $a \neq 0$), or in a discrete way ($a = a_0^m$, $b = nb_0a_0^m$, with $m, n \in \mathbb{Z}$, and $a_0 > 1$, $b_0 > 0$ fixed). J. Morlet proposed the discrete version of (1.1) as an alternative to the windowed Fourier transform, which computes, for a given function f,

$$\phi_{\mathrm{Wind.F.T.},f}(p, q) = \int dx\, e^{-ipx} g(x - q) f(x)\,, \tag{1.2}$$

where g is a fixed window function. In Gabor's approach, the function g is chosen to be Gaussian, $g(x) = \pi^{-1/4} \exp(-x^2/2)$, but many other window functions can be (and are) used. The parameters p, q can again vary either continuously ($p, q \in \mathbb{R}$), or discretely ($p = mp_0, q = nq_0$, with $m, n \in \mathbb{Z}$, and $p_0, q_0 > 0$ fixed). The wavelet transform (1.1) and

---

* "Bevoegdverklaard Navorser" at the Belgium National Foundation for Scientific Research (on leave); on leave also from Department of Theoretical Physics, Vrije Universiteit Brussel, (Belgium).

the windowed Fourier transform (1.2) have many features in common. Provided the basic functions h, g and their Fourier transforms are reasonably well concentrated, the two transforms analyze the frequency content of the signal f, locally in time (if the variable x is to be understood as "time"). This is clear from the fact that (1.1), (1.2) are scalar products of f with $h_{a,b}(x) = |a|^{-1} h\left(\frac{x-b}{a}\right)$, $g_{p,q}(x) = e^{ipx} g(x-q)$ respectively. If g and its Fourier transform $\hat{g}$ are both concentrated around 0 (as for g Gaussian), then $g_{p,q}$ is concentrated around q, while $\hat{g_{p,q}}$ is concentrated around p. The scalar product $\langle g_{p,q}, f\rangle$ therefore analyzes f in a neighborhood of the time-frequency point (q, p). A similar argument holds for the scalar products $\langle h_{a,b}, f\rangle$; note however that the frequency analysis performed by the wavelet transform is different from the windowed Fourier transform (see also below).

Another feature that (1.1) and (1.2) have in common is the reconstruction formula for f from ϕ_f (continuous version). We have

$$f(x) = \frac{1}{2\pi} \int\int dp\, dq\ \phi_{\text{Wind.F.T.},f}(p, q)\, e^{ipx} g(x-q) \tag{1.3}$$

and

$$f(x) = \frac{1}{2\pi C_h} \int\int \frac{da\, db}{a^2}\ \phi_{\text{Wav.T.},f}(a, b)\ |a|^{-1/2} h\left(\frac{x-b}{a}\right). \tag{1.4}$$

In (1.4) the constant C_h is defined by

$$C_h = \int d\xi\ |\hat{h}(\xi)|^2\ |\xi|^{-1}, \tag{1.5}$$

where $\hat{h}(\xi) = (2\pi)^{-1/2} \int dx\, e^{ix.\xi} h(x)$ is the Fourier transform of h. We assume that $|\hat{h}(\xi)| = |\hat{h}(-\xi)|$ for all ξ (otherwise (1.4) has to be replaced by a more complicated formula), and that $C_h < \infty$. For reasonably nice h (e.g. $|h(x)| \le C(1+|x|)^{-\alpha}$, $\alpha > 1$), this amounts to requiring $\int dx\, h(x) = 0$.

The similarity between (1.3) and (1.4) is due to the fact that (1.1)-(1.4) and (1.2)-(1.3) are both special cases of square integrable representations, as pointed out by A. Grossmann. The reconstruction formula (1.4) and the associated reproducing kernel Hilbert space enabled A. Grossmann and J. Morlet to analyze in detail the continuous wavelet transform [2].

The paragraphs above list a few analogies between wavelet transform and windowed Fourier transform. Even more interesting are their crucial differences. They can easily be illustrated by looking at the graphs for $g_{mn}(x) = e^{imp_0x} g(x - nq_0)$ and

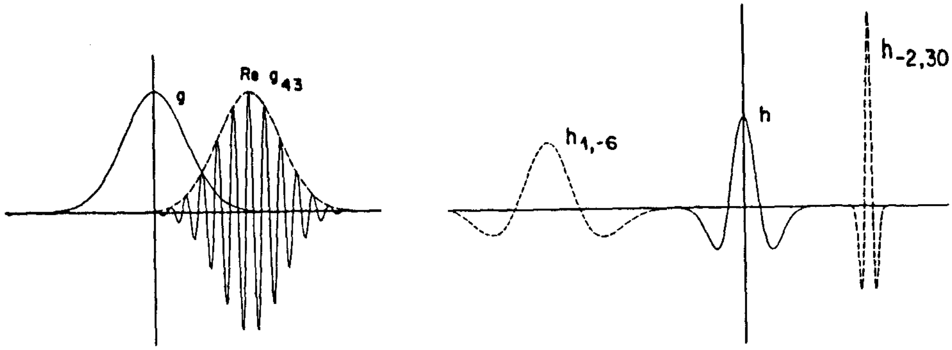

Figure 1. Examples of functions g_{mn}, h_{mn} corresponding to resp. the windowed Fourier transform and the wavelet transform.

$h_{mn}(x) = a_0^{-m/2}\, h(a_0^{-m}x - nb_0)$, corresponding to the discrete versions of both transforms. Figure 1 shows g_{mn}, h_{mn} for a few values of m, n, for the choices $g(x) = \pi^{-1/4} e^{-x^2/2}$, $h(x) = 2/\sqrt{3}\, \pi^{-1/4}(1 - x^2)e^{-x^2/2}$.

In both cases different values of m correspond to different frequency ranges. The high frequency g_{mn} are high frequency oscillations with an amplitude modulation given by $g(x - nq_0)$. The envelope function of all the g_{mn} is therefore always the same function g, translated to the relevant time interval (indexed by n). The high frequency h_{mn} look very different, however. As contracted (and translated) versions of the basic function h they have variable width, adapted to their frequency range: the higher that range, the more narrow they are. This difference in time-resolution for high versus low frequency wavelets, in contrast to the fixed time resolution for all frequency components of the sliding windowed Fourier transform, is illustrated very clearly by Fig. 2. For both transforms, the centers of localization in the time-frequency plane are plotted, corresponding to the g_{mn} or h_{mn}, respectively. For the wavelet transform, this discrete lattice shows the differences in time-resolution as the frequency bounds change. Note that while for higher frequencies the time resolution becomes better, the frequency resolution becomes worse, as was to be expected from the Heisenberg uncertainty principle. The better time resolution for high frequency components enables the wavelet transform to perform better than the sliding window Fourier transform for signals which typically have short-lived high frequency components superposed on longer-lived lower frequency parts, as in e.g. music, or speech. The exponential rather than linear treatment of frequency is also more closely related to our auditory perception.

It should be noted that techniques related to the wavelet transform, based on the use of dilations and translations, have been used in many different fields. Written in a different way, the reconstruction formula (1.4) appears in the pioneering work of A. Calderón in harmonic

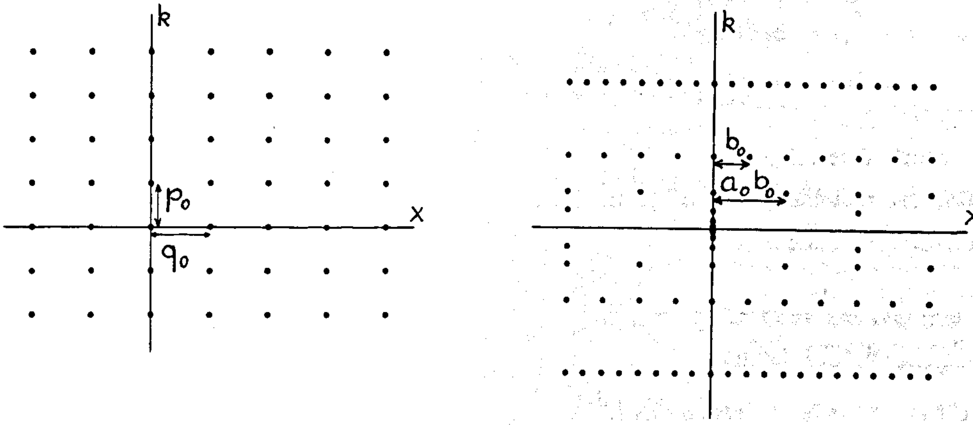

Figure 2. The centers of localization in the time-frequency plane (x = time, k = frequency) corresponding to the g_{mn} (windowed Fourier transform), resp. h_{mn} (wavelet transform).

analysis [3]. In this field also, the possibility to zoom in on short-lived high frequency phenomena was important, for applications to the study of singular integral operators (see, e.g. [4]). Both the wavelet transform and the windowed Fourier transform, with their respective reconstruction formulas (1.3), (1.4), are examples of coherent state decompositions used in quantum physics (for a review, see [5]). The affine coherent states, as the wavelets are called in this framework, were first introduced in [6]. They are shown to play a particular role for the hydrogen atom in [7]. The idea of decomposing into building blocks of constant "shape" but different size (and this is essentially what the wavelet transform does) is also central to the renormalization group theory, important in statistical mechanics and quantum field theory [8]. It is therefore not altogether surprising that new developments in wavelets have led to an elegant application in quantum field theory [9]. Finally, the same ideas are also related to certain filter banks used in acoustic signal analysis. We shall come back to this later.

The reconstruction formulas (1.3), (1.4) use the continuously labelled windowed Fourier transform or wavelet transform of f in order to reconstruct f. When discretely labelled g_{mn} or h_{mn} are used, different reconstruction algorithms apply. In both cases, we define the map

$$C : L^2(\mathbb{R}) \to \ell^2(\mathbb{Z}^2)$$

by

$$(Cf)_{m,n} = \langle \phi_{mn}, f \rangle ,$$

where ϕ_{mn} is either g_{mn} (windowed Fourier transform) or h_{mn} (wavelet transform), and where $\langle \ , \ \rangle$ denotes the standard L^2-inner product

$$\langle f, g\rangle = \int dx\, \overline{f(x)}\, g(x) .$$

The map C depends of course on the chosen function h or g, and on the parameters a_0, b_0 or p_0, q_0 which determine the density of the lattices in Figure 2. If h or g and its Fouier transform are reasonably well concentrated (i.e. in all the cases of practical interest), then the operator C is bounded. In order to have a "good" characterization of signals f by their coefficients $(Cf)_{mn}$, we require that

1. C is one-to-one: if $f_1 \neq f_2$, then $Cf_1 \neq Cf_2$

2. C has a bounded inverse on its range: if Cf_1 and Cf_2 are "close", then so are f_1 and f_2.

This means that there exist $A > 0$, $B < \infty$ such that, for all $f \in L^2(\mathbb{R})$,

$$A \int dx\, |f(x)|^2 \leq \sum_{m,n} |\langle \phi_{mn}, f\rangle|^2 \leq B \int dx\, |f(x)|^2 . \tag{1.6}$$

The set of vectors $\{\phi_{mn}; m,n \in \mathbb{Z}\}$ is then called a "frame" for $L^2(\mathbb{R})$. Note that a frame is not necessarily a basis; in many cases it is "overcomplete". A simple example in $\mathbb{R}^2$ (not related to wavelets or windowed Fourier transform) illustrates this. Define, in $\mathbb{R}^2$, $e_1 = (1,0)$, $e_2 = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$, $e_3 = \left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)$. It is easy to check that for all $v \in \mathbb{R}^2$, $\sum_{j=1}^{3} |\langle e_j, v\rangle|^2 = \frac{3}{2}\|v\|^2$. The $\{e_j; j = 1,2,3\}$ constitute therefore a frame for $\mathbb{R}^2$, while they are clearly not a basis, since they are not linearly independent.

If the two frame bounds A, B in (1.5) are equal, then we call the frame "tight". For a tight frame we have

$$\langle f, \left(\sum_{m,n} \phi_{mn} \langle \phi_{mn}, f\rangle\right)\rangle = \sum_{m,n} |\langle \phi_{mn}, f\rangle|^2 = A\langle f, f\rangle ,$$

hence

$$f = A^{-1} \sum_{m,n} \phi_{mn}(Cf)_{mn} .$$

For tight frames, this is the desired inversion formula, allowing to reconstruct f from the $(Cf)_{mn}$. For general, non-tight frames, one can still write

$$f = 2(A + B)^{-1} \sum_{m,n} \phi_{mn}(Cf)_{mn} + Rf , \tag{1.7}$$

where the remainder term Rf is bounded by

$$\int dx\ |(Rf)(x)|^2 \leq \frac{BA^{-1}-1}{BA^{-1}+1} \int dx\ |f(x)|^2 .$$

Since $(BA^{-1}-1)/(BA^{-1}+1) < 1$, it is clear that (1.6) can be iterated to obtain a reconstruction formula for f from the $(Cf)_{mn}$ with any desired accuracy. For a given precision, the number of terms required depends on $BA^{-1}-1$. The closer the two frame bounds are to each other, the more "snugly" the frame fits, and the fewer iterations are required. It is therefore important to have good estimates for A, B. Such estimates can be computed by means of the Poisson sum formula. The following table lists values of the frame bounds for the wavelet transform, for various choices of a_0 and b_0.

Frame bounds for the wavelet transform, with $h(x) = 2/\sqrt{3}\,\pi^{-1/4}(1-x^2)\,e^{-x^2/2}$.

a_0	b_0	A	B	$BA^{-1}-1$
2.	.5	6.546	7.092	0.083
	1.	3.223	3.596	0.116
$\sqrt{2}$	.5	13.637	13.639	0.0002
	1.	6.768	6.870	0.015

It is clear that even for relatively large values of a_0, b_0 the frame constants can be so close that one can drop the remainder term Rf in (1.6) to obtain an approximate reconstruction formula that is extremely accurate. For other examples, variations on the same theme, and more details, we refer the reader to [10]. Note that other inversion formulas than (1.6) may be used. One easily checks that

$$\int da\, a^{-3/2}\, \langle h_{a,b}, f\rangle = f(b) \cdot \int dt\, t^{-1}\hat{h}(t) .$$

For values of a_0, b_0 close to 1,0 respectively, one can then use a discrete approximation of this formula as a reconstruction algorithm for f. Regardless of which algorithm is used, the condition (1.5) is necessary to ensure that the inversion is numerically stable.

The concepts of frame and frame bounds, and formula (1.6) can also be applied to the windowed Fourier transform. In fact, frames are essential in windowed Fourier transform analysis if one is also interested in good time-frequency localization. This is due to the following theorem, first stated, in a more restricted version, by Balian [11] and Low [12], and extended and rigorously proved by Coifman and Semmes [10]. A simpler and much more elegant proof for bases was subsequently found by Battle [13].

Theorem: If $g \in L^2(\mathbb{R})$, $p_0, q_0 > 0$ are such that the g_{mn} constitute a frame, and $p_0 \cdot q_0 = 2\pi$, then either xg or $g' \notin L^2(\mathbb{R})$.

In practical applications, good time and frequency localization of g is required (in fact, much stronger conditions than $xg, g' \in L^2$ are usually satisfied). In order to have numerically stable inversion formulas recovering f from the $\langle g_{mn}, f \rangle$, one is therefore forced to consider time-frequency lattices with mesh size $p_0 \cdot q_0 < 2\pi$. Lattices with $p_0 \cdot q_0 = 2\pi$ correspond to "sampling" with the Nyquist density, which is the reason why such lattices have been proposed for windowed Fourier transform analysis (starting with Gabor [14], who proposed $g(x) = \pi^{-1/4} e^{-x^2/2}$, $p_0 = 2\pi$, $q_0 = 1$). The above theorem tells us that for reasonable g, $p_0 \cdot q_0 = 2\pi$ always leads to numerically unstable inversion algorithms. This confirms the fact that the Gabor transform has no bounded inverse [15] [16]. Figure 3 illustrates this fact in a dramatic way. For the windowed Fourier transform (1.6) can be written as

$$f = \sum_{m,n} \tilde{g}_{mn} \langle g_{mn}, f \rangle, \tag{1.8}$$

where $\tilde{g}_{mn}(x) = e^{imp_0x} \tilde{g}(x - nq_0)$, and $\tilde{g}$ is a "dual" function, completely determined by g, p_0 and q_0, $\tilde{g} = 2(A + B)^{-1} g + O(BA^{-1} - 1)$. Figure 3 shows this dual function $\tilde{g}$, for $g(x) = \pi^{-1/4} e^{-x^2/2}$ and $p_0 = q_0 = \sqrt{2\pi\lambda}$, for the different values $\lambda = .25, .5, .75, .95$ and 1. For $\lambda = 1$, the g_{mn} do not constitute a frame, which is expressed by the singularities of $\tilde{g}$;

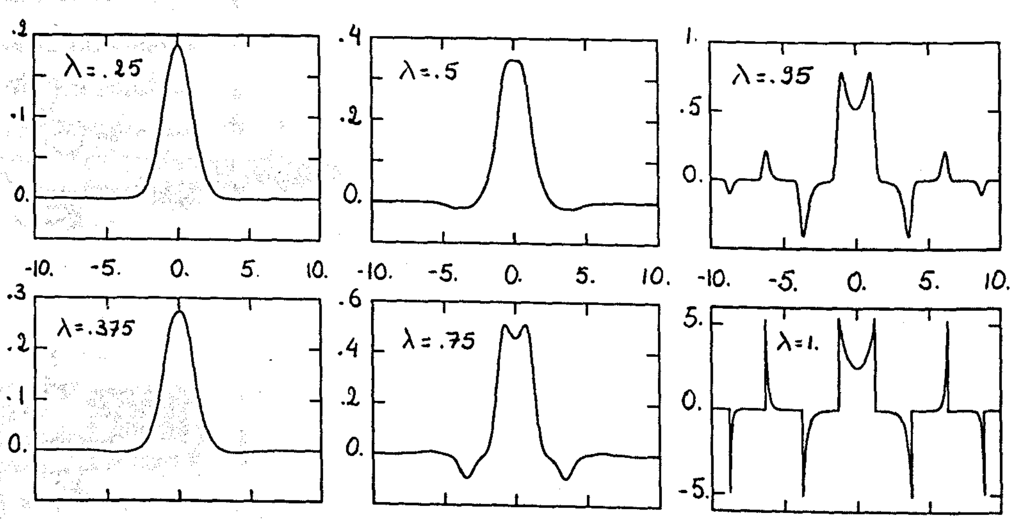

Figure 3. The dual function $\tilde{g}$ corresponding to the lattice g_{mn}, with $g(x) = \pi^{-1/4} \exp(-x^2/2)$, and $p_0 = q_0 = (2\pi\lambda)^{1/2}$, for $\lambda = .25, .375, .5, .75, .95$ and the singular case $\lambda = 1$.

(1.7) has to be understood in a distributional sense in this case. In fact $\tilde{g}$ is not even in L^2 for $\lambda = 1$ [16]. For the four smaller values of λ, $\tilde{g}$ is still C^∞, with gaussian decay; it is clear how $\tilde{g}$ evolves, for increasing λ, from a gaussian profile to the singular function for $\lambda = 1$.

On the other hand, one can easily show that a family of g_{mn} can only constitute an orthonormal basis if $p_0 \cdot q_0 = 2\pi$. The Balian-Low theorem excludes therefore the existence of any orthonormal basis for the windowed Fourier transform with reasonable localization properties. This leads us to another fundamental difference between the windowed Fourier transform and the wavelet transform: there *do* exist suitably chosen functions h and constants a_0, b_0 such that both h and its Fourier transform have fast decay (e.g. faster than any inverse polynomial power) and such that the corresponding h_{mn} constitute an orthonormal basis of $L^2(\mathbb{R})$. The first example of such an orthonormal basis was constructed by Y. Meyer [17], and extended to higher dimensions by P. G. Lemarié and Y. Meyer [18]. Other examples were constructed shortly after by G. Battle [19] and P. G. Lemarié [20]. These first constructions were rather mysterious, and relied on "miraculous" cancellations. The introduction of multiresolution analysis by S. Mallat and Y. Meyer led to a deeper understanding of these bases.

2. Orthonormal bases of wavelets and multiresolution analysis.

The papers by S. Mallat and Y. Meyer in this same volume no doubt discuss multiresolution analysis in greater detail than possible here. This paragraph is therefore restricted to a short review.

A "multiresolution analysis" of a function f consists in a hierarchy of approximations of f, defined as averages on different scales. The finer the scale, the better the approximation. More precisely, one has a hierarchy of subspaces of $L^2(\mathbb{R})$

$$\ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \tag{2.1}$$

such that

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \quad \overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R}) .$$

The "scaling" aspect is translated by the condition

$$g \in V_j \Leftrightarrow g(2.) \in V_{j+1} . \tag{2.2}$$

The space V_0 thus determines the whole ladder of spaces. A typical but not very sophisticated example is the case where V_0 consists of piecewise constant functions,

$$V_0 = \{g \in L^2(\mathbb{R});\ g \text{ is constant on every half open interval} \tag{2.3}$$
$$[k, k+1[\,,\ \text{ for all } k \in \mathbb{Z}\}\ .$$

The spaces V_j then contain functions which are piecewise constant on the intervals $[k2^j, (k+1)2^j[$; they clearly satisfy (2.1).

We also impose some translation invariance:

$$g \in V_0 \Leftrightarrow g(\cdot - k) \in V_0 \ \text{ for all } k \in \mathbb{Z}\ . \tag{2.4}$$

The final requirement is that there exists $\phi \in V_0$ such that its integer translates are an orthonormal basis for V_0, i.e.

$$\text{for all } g \in V_0,\ \int dx\, |g(x)|^2 = \sum_k \left|\int dx\, \overline{\phi(x-k)}\, g(x)\right|^2 . \tag{2.5}$$

(In fact, it is sufficient to require that the $\phi(\cdot - k)$ constitute a Riesz basis for V_0; one can prove that this entails the existence of $\tilde{\phi} \in V_0$ satisfying (2.5) [21]). In the example (2.3) above, one can choose $\phi(x) = 1$ if $0 \le x < 1$, 0 otherwise; this clearly satisfies (2.5).

For a given function f the successive multiresolution approximations are defined as the orthogonal projections onto the V_j

$$P_j f = \sum_{k \in \mathbb{Z}} \phi_{jk} \langle \phi_{jk}, f \rangle ,$$

where $\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k)$; the ϕ_{jk} constitute an orthonormal basis for V_j by (2.5) and (2.2).

The "difference in information" between two successive approximations $P_j f$ and $P_{j+1} f$ is given by the orthogonal projection $Q_j f$ of f onto the orthogonal complement W_j of V_j in V_{j+1},

$$W_j \perp V_j$$
$$V_j \oplus W_j = V_{j+1}$$
$$Q_j f = P_{j+1} f - P_j f\ .$$

The four requirements (2.1), (2.2), (2.4), (2.5) imply that the spaces W_j are also scaled versions of one space W_0,

$$g \in W_j \Leftrightarrow g(2^{-j}\cdot) \in W_0\ , \tag{2.6}$$

that they are translation invariant for the discrete translations $2^{-j}\mathbb{Z}$,

$$g \in W_0 \Leftrightarrow g(\cdot - k) \in W_0$$

and that they are mutually orthogonal spaces generating all of $L^2(\mathbb{R})$,

$$W_j \perp W_{j'} \quad \text{for } j \neq j'$$
$$\bigoplus_{j \in \mathbb{Z}} W_j = L^2(\mathbb{R}) \ . \tag{2.7}$$

Moreover [21] there exists a function $\psi \in W_0$ such that the $\psi(\cdot - k)$ constitute an orthonormal basis for W_0,

$$\text{for all } \ g \in W_0 \, , \ \int dx \, |g(x)|^2 = \sum_{k \in \mathbb{Z}} \ |\int dx \, \overline{\psi(x-k)} \, g(x)|^2 \ . \tag{2.8}$$

By (2.6) it follows that the $\psi_{jk}(x) = 2^{j/2} \ \psi(2^j x - k)$, for fixed j, constitute an orthonormal basis for W_j. Hence by (2.7), the $\{\psi_{jk}, j, k \in \mathbb{Z}\}$ are an orthonormal basis of wavelets for $L^2(\mathbb{R})$.

In the example (2.3) it is easy to guess ψ. The space W_0 is constituted by those functions that are piecewise constant on the intervals $[k/2, (k+1)/2[$, and are orthogonal to the functions constant on $[\ell, \ell + 1[$. It is easy to convince oneself that W_0 is therefore spanned by the function $\psi(x) = 1$ for $0 \leq x < 1/2$, -1 for $1/2 \leq x < 1$, 0 otherwise, and its integer translates. Since the integer translates of ψ are mutually orthogonal, ψ satisfies (2.8), and the ψ_{jk} are the orthonormal wavelet basis associated with the multiresolution analysis defined by (2.3). This basis was in fact well-known long before wavelets existed: it is called the Haar basis, and is known to be an unconditional basis for all L^p-spaces, $1 < p < \infty$. Multiresolution analysis allows one to construct orthonormal bases, such as the Battle-Lemarié bases, which generalize the Haar basis, but are smoother (the Battle-Lemarié bases are C^k and have exponential decay). They are therefore suitable for other spaces (Sobolev, Besov, ...) than only the L^p-spaces.

There exists an explicit algorithm for the construction of ψ, once ϕ is known. Since $\phi \in V_0 \subset V_1$, there exist c_n such that

$$\phi(x) = \sum c_n \, \phi(2x - n) \ . \tag{2.9}$$

(Since the ϕ_{1n} are an orthonormal basis in V_1, one has $c_n = \sqrt{2} \, \langle \phi_{1n}, \phi \rangle$). Then [21]

$$\psi(x) = \sum (-1)^n \ c_{n+1} \, \phi(2x + n) \ . \tag{2.10}$$

3. Stéphane Mallat's algorithm: the connection with discrete filters.

S. Mallat uses multiresolution analysis and orthonormal bases of wavelets in a discrete algorithm applied to vision analysis [22]. In fact the concept of multiresolution analysis owes it existence to inspiration drawn from concrete schemes for analysis and reconstruction of vision data. To illustrate how related ideas were being used in vision analysis, we sketch one such scheme, the "Laplacian pyramid" of Burt and Adelson [23].

While visual data are of course 2-dimensional, we shall restrict our discussion to 1 dimension; the generalization to higher dimensions is easy. The initial data are a sequence of numbers $(c_n)_{n\in\mathbb{Z}}$, indicating the grey levels at different, equally spaced grid points. We shall identify the original data with the subscript 0, $(c_n^0)_{n\in\mathbb{Z}}$. The idea behind the Laplacian pyramid scheme is to define successively more and more blurred versions c^ℓ of the data c^0 (corresponding to low-pass filters applied to c^0), and to encode the "difference in information" between successive c^ℓ. The scheme proposed by Burt and Adelson [23] achieves this goal in the following elegant way. Define

$$c_k^1 = \sum_n w(n-2k)\, c_n^0 , \tag{3.1}$$

where the $w(j)$ are weighting factors satisfying $\sum_j w(j) = 1$. Typically only a small number of $w(j)$ are different from zero; an example used in [23] is

$$w(0) = a,\ w(1) = w(-1) = .25,\ w(2) = w(-2) = .25 - a/2,\ w(j) = 0 \text{ otherwise}, \tag{3.2}$$

with $a \in\ ].125, .625[$ (see [24]).

The operation (3.1) consists in a convolution and a decimation; the resulting c^1 "lives" on a larger scale than c^0, containing less information. To encode the "difference in information" between c^0 and c^1, we first compute, from c^1, a sequence $\tilde{c}^0$ that lives on the same scale as c^0, using the same weight coefficients,

$$\tilde{c}_n^0 = \sum_k w(n-2k)\, c_k^1 , \quad \text{(see Figure 4b)} \tag{3.3}$$

and we subtract,

$$d_n^0 = c_n^0 - \tilde{c}_n^0 . \tag{3.4}$$

The information in c^0 is thus split up into d^0 and c^1, and is fully recoverable from these two new sequences. The same process, using the sequence c^1 as starting point rather than c^0, can be repeated, leading to sequences c^2 and d^1. After L iterations we have decomposed c^0 into $d^0, d^1, \ldots, d^L$ and c^{L+1}.

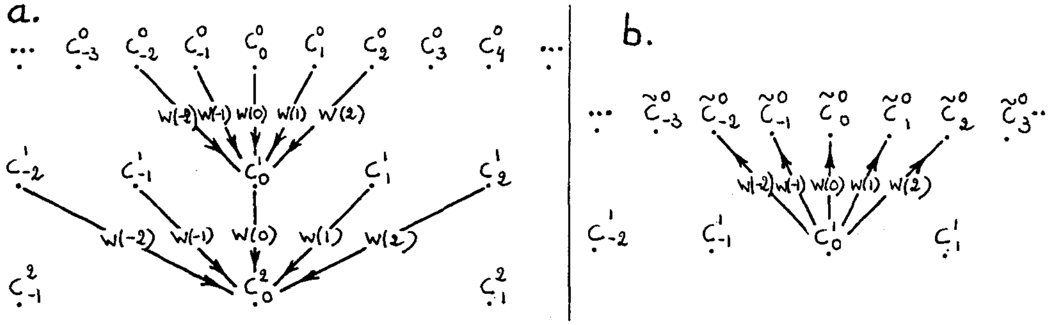

Figure 4. Schematic representation of the operations (3.1) (a) and (3.3) (b).

Let us introduce the filter operator $F : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ defined by

$$(Fa)_k = \sum_n w(n - 2k)\, a_n \ . \tag{3.5}$$

F consists of a convolution, followed by a decimation (only one out of every two samples is retained after the convolution). We also define the dual operation F^* by

$$\sum_n (F^* b)_n\, \overline{a_n} = \sum_n b_n\, \overline{(Fa)_n} \, ,$$

where a, b are arbitrary sequences in $\ell^2(\mathbb{Z})$. Since the $w(n)$ are real, this leads to

$$(F^* b)_n = \sum_k w(n - 2k)\, b_k \ . \tag{3.6}$$

The equations (3.1), (3.3) and (3.4) and their iterates can then be rewritten as

$$c^\ell = F\, c^{\ell-1} \, , \quad \tilde{c}^{\ell-1} = F^* c^\ell$$

$$d^{\ell-1} = c^{\ell-1} - \tilde{c}^{\ell-1} = (Id - F^* F)\, c^{\ell-1}$$

The filters F, F^* are extremely easy to implement numerically, and the decomposition of c^0 into the "difference" sequences $d^0, \dots, d^L$ and a much smoothed out version c^{L+1} can be computed as fast as an FFT (fast Fourier transform). Reconstruction of c^0 from the $d^0, \dots, d^L$ and c^{L+1} is just as easy, since one only needs to iterate

$$c^{\ell-1} = d^{\ell-1} + F^* c^\ell \, , \tag{3.7}$$

starting from $\ell = L$.

One last remark concerning the Laplacian pyramid scheme. The reconstruction formula (3.7) effectively writes c^0 as

$$c^0 = F^* d^1 + (F^*)^2 d^2 + \cdots + (F^*)^L d^L + (F^*)^{L+1} c^{L+1} .$$

Since the filter operator F^* is iterated, it is worthwhile to ensure that it does not look "messy" when applied to, e.g., a sequence with only one non-zero entry. The filters (3.2) satisfy this requirement, as shown in Figure 5. In this figure we visualize the successive sequences e (with $e_0 = 1$, $e_n = 0$ otherwise), $F^* e$, $(F^*)^2 e$,... by piecewise constant functions, with levels given by the entries of the sequence. The stepwidth of the intervals is adapted to the iteration level ℓ in $(F^*)^\ell e$, since b and $F^* b$ "live" on different scales (see Fig. 4b). It is clear that the $(F^*)^\ell e$ converge to a "nice" function in Fig. 5; for a mathematical proof of this convergence we refer to [24].

S. Mallat's algorithm defines "averages" c^ℓ and "differences" d^ℓ, from an initial sequence c^0, via multiresolution analysis in the following way. From c^0, he defines $f_0 \in V_0$ by

$$f_0 = \sum_k c_k^0 \, \phi_{0k} \, . \tag{3.8}$$

As an element of $V_0 = V_{-1} \oplus W_{-1}$, f can be decomposed into $f_{-1} \in V_{-1}$ and $g_{-1} \in W_{-1}$,

$$f_0 = f_{-1} + g_{-1} = \sum_k c_k^1 \, \phi_{-1k} + \sum_k d_k^1 \, \psi_{-1k} , \tag{3.9}$$

where we have used that the ϕ_{jk}, ψ_{jk} are orthonormal bases for V_j, W_j respectively (see §2). Here f_{-1} corresponds to an "averaged" version of f_0, and g_{-1} to the difference in information between f_0 and this average. Since (3.8), (3.9) are expansions with respect to orthonormal basis, the coefficients c_k^1, d_k^1 are easy to compute,

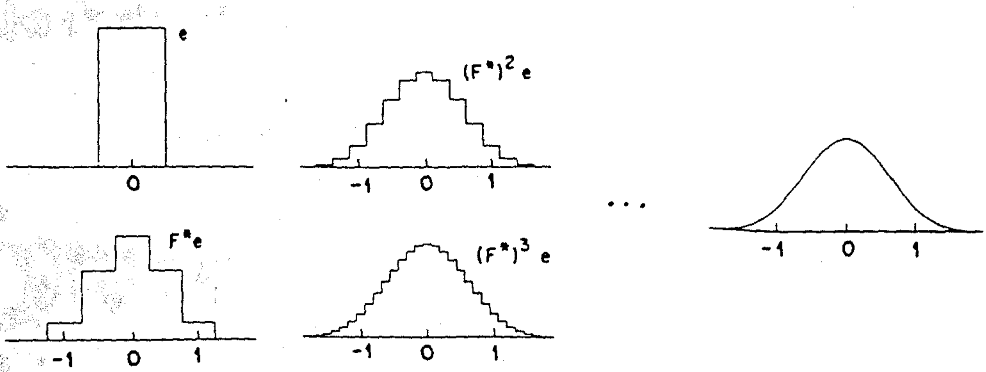

Figure 5. Representation of the sequences $e, F^* e, (F^*)^3 e$ by piecewise constant functions, and the limit function. The stepwidth for each $(F^*)^\ell e$ is $2^{-\ell}$. The filter coefficients for this figure are given by (3.2), with $a = .375$.

$$c_k^1 = \langle \phi_{-1k}, f \rangle = \sum_n c_n^0 \langle \phi_{-1k}, \phi_{0n} \rangle = \sum_n h(n - 2k)\, c_n^0$$

$$d_k^1 = \langle \psi_{-1k}, f \rangle = \sum_n c_n^0 \langle \psi_{-1k}, \phi_{0n} \rangle = \sum_n g(n - 2k)\, c_n^0 \,,$$

where $h(n) = \langle \phi_{-10}, \phi_{0n} \rangle = 2^{-1/2} \int dx\, \phi(x/2)\, \phi(x - n)$, $g(n) = 2^{-1/2} \int dx\, \psi(x/2)\, \phi(x - n)$. Consequently both c^1 and d^1 are derived from c^0 by the application of filters H, G of type (3.5),

$$c^1 = Hc^0 \,, \quad d^1 = Gc^0 \,. \tag{3.10}$$

On the other hand

$$\begin{aligned} c_n^0 &= \langle \phi_{0n}, f_0 \rangle = \langle \phi_{0n}, f_{-1} + g_{-1} \rangle \\ &= \sum_k c_k^1 \langle \phi_{0n}, \phi_{-1,k} \rangle + \sum_k d_k^1 \langle \phi_{0n}, \psi_{-1,k} \rangle \\ &= \sum_k \left[h(n - 2k)\, c_k^1 + g(n - 2k)\, d_k^1 \right], \end{aligned}$$

or

$$c^0 = H^* c^1 + G^* d^1 \,, \tag{3.11}$$

with H^*, G^* defined analogously to (3.6). The whole decomposition process can of course be iterated: f_{-1} decomposes into $f_{-2} + g_{-2}$, corresponding to sequences $c^2 = Hc^1$ and $d^2 = Gc^1$, etc.

Schematically, decomposition and reconstruction can be represented as in Figure 6.

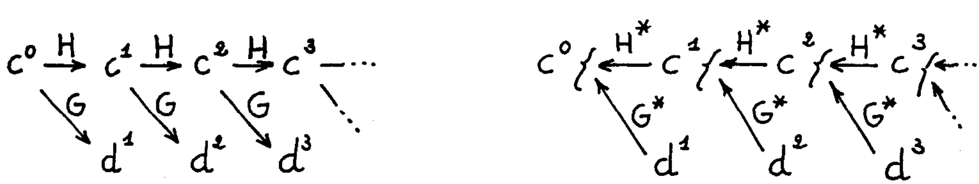

Figure 6. Schematic representation of the tree-algorithms for the decomposition and reconstruction in S. Mallat's scheme.

The tree-structure, together with the easy convolution and decimation structure of H, G, makes that this algorithm works very fast; the whole decomposition can be done as fast as an FFT. Note that at every level c^ℓ is replaced by a roughly equivalent number of entries: if the c_n^ℓ are zero except for N consecutive entries, then, apart from edge effects, only $N/2$ entries of $c^{\ell+1}$, $d^{\ell+1}$ will be non-vanishing. The total number of relevant entries in $d^1, d^2, \ldots, d^L, c^L$ is therefore essentially the same as in the original sequence c^0. This is in opposition to the

Laplacian pyramid scheme, where the total number of entries after L iterations is essentially $(2 - 2^{-L})$ times the original number. The multiresolution-based algorithm of S. Mallat is thus more efficient than the Laplacian pyramid scheme. Moreover, when generalized to 2 dimensions, it also turns out to be orientation selective, at no extra cost, which is another advantage over the Laplacian pyramid scheme [22].

In fact, for the implementation of S. Mallat's algorithm, one only needs the two filters G, H; their multiresolution analysis origins are not used explicitly. One may therefore try to isolate the relevant properties of the filters, and design filters satisfying all these properties directly, without multiresolution analysis. From (3.10) and (3.11) a first condition can be derived.

(C1) $$H^*H + G^*G = Id .$$

The orthogonality of V_{-1} and W_{-1} results in

(C2) $$HG^* = 0;$$

this expresses the fact that the two terms in the decomposition (3.11) of c^0 are always orthogonal sequences. A third condition expresses the fact that H is an "averaging operator", i.e. a low pass filter, while G measures the difference between a sequence and its average, and is therefore a band pass filter. This results in

(C3) $$\sum_n g(n) = 0$$

$$\sum_n h(n) = \sqrt{2}$$

where the $\sqrt{2}$-normalization is due to the decimation 2:1 in the definition of the filter H (see [24]). Finally, we also impose a regularity condition, similar to the condition on the Laplacian pyramid scheme. The complete reconstruction formula for c^0 from $d^1, d^2, \ldots, d^L, c^L$ is

$$c^0 = G^*d^1 + H^*G^*d^2 + \cdots + (H^*)^{L-1}G^*d^L + (H^*)^Lc^L .$$

We shall therefore require that the operator H^* which is iterated satisfies

(C4) the piecewise constant functions representing $(H^*)^\ell e$

(where $e_n = 0$ for $n \neq 0, e_0 = 1$) converge to a "nice" function as $\ell \to \infty$.

For a more precise formulation of this condition, see [24]. Filters H, G which are derived from a multiresolution analysis automatically satisfy conditions C1-C3. Moreover, one can show that in this case the piecewise constant functions representing $(H^*)^\ell e$ converge to the averaging function ϕ itself [24], so that C4 is also satisfied. It is possible to construct filters H, G of type (3.5) which satisfy C1-C3, but not C4. An example is given in Figure 7. In this case the $(H^*)^\ell e$ converge, for $\ell \to \infty$, to a distribution which is singular at every dyadic rational between 0 and 3, i.e. every point of the form $k2^{-m}$, with $0 \le k < 3\,2^m$. This example shows that condition C4 is necessary to avoid "messy" iterations. It turns out [24] that conditions C1-C4 ensure that the filters H, G are associated to a multiresolution analysis. The "averaging function" ϕ of that multiresolution analysis is exactly the "nice" function to which the $(H^*)^\ell e$-piecewise constant functions converge. The proof in [24] of this equivalence between filters and orthonormal bases of wavelets essentially uses this "graphical" construction of ϕ as a limit of piecewise constant functions representing sequences. The orthogonality of the $\phi(\cdot - k)$ e.g. can easily be proved in the following way. Let T be the translation by one unit step for sequences, $(Ta)_n = a_{n-1}$. Then e and $T^\ell e$ are obviously orthogonal if $\ell \neq 0$,

$$\sum_k e_k (T^\ell e)_k = 0 \, .$$

On the other hand, the special structure (3.5) of the filters H, G together with the conditions C1-C2 imply $HH^* = Id = GG^*$. It follows that $(H^*)^m e$ and $(H^*)^m T^\ell e$ are orthogonal for all

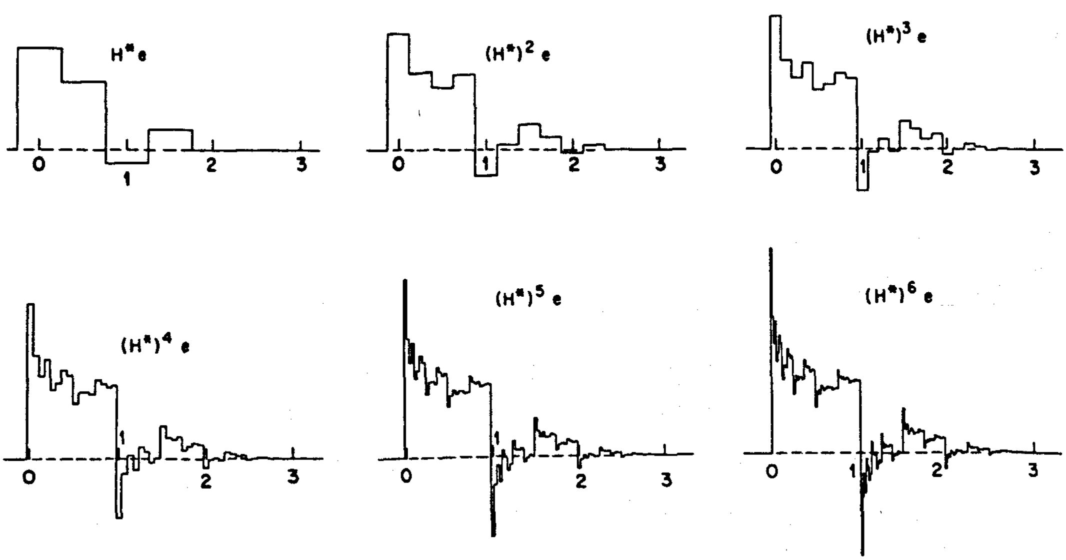

Figure 7. An example of a pair of filters H, G which do not satisfy the regularity condition C4. In this case $h(0) = 15/(13\sqrt{2})$, $h(1) = 10/(13\sqrt{2})$, $h(2) = -2/(13\sqrt{2})$, $h(3) = 3/(13\sqrt{2})$, with all the other $h(n) = 0$. The $g(n)$ are defined by (3.12), $g(n) = (-1)^n h(-n+1)$. One readily checks that C1-C3 are satisfied.

$m \geq 0$. The piecewise constant functions representing them are therefore also orthogonal for all m, and so are their limits $\phi(x)$ and $\phi(x-e)$. Defining $\psi(x) = \sqrt{2} \sum_n g(n) \phi(2x-n)$, one proves similarly that the $\psi(\cdot - k)$ are orthogonal, and that the $\phi(\cdot - k)$ are orthogonal to the $\psi(\cdot - \ell)$. The property

$$\phi(x) = \sqrt{2} \sum_n h(n)\, \phi(2x-n)$$

follows immediately from the construction, while C1 implies that for all $f \in L^2(\mathbb{R})$, and all $J > 0$

$$\sum_n |\langle f, \phi_{J+1,m}\rangle|^2 = \sum_k |\langle f, \phi_{-J,k}\rangle|^2 + \sum_{j=-J}^{J} \sum_k |\langle f, \psi_{jk}\rangle|^2$$

It is not hard to prove that $\sum_k |\langle f, \phi_{-J,k}\rangle|^2 \to 0$, while $\sum_n |\langle f, \phi_{J+1,m}\rangle|^2 \to \int dx\ |f(x)|^2$ for $J \to \infty$. It then follows that the ψ_{jk} are an orthonormal basis of wavelets [24].

Remarks.

1. The conditions C1-C2 are, in a different form, the "unitarity conditions" imposed by Y. Meyer (see e.g. [25], or his paper in this volume).

2. A different proof of the equivalence filters — wavelet bases can be found in [22], where the regularity condition C4 is replaced by the condition

$$\inf_{\xi \in \mathbb{R},\ |\xi| \leq \pi/2} \left|\sum_n h(n)\, e^{in\xi}\right| > 0\,.$$

   This positivity condition is sufficient to ensure that the ψ_{mn} are an orthonormal basis; it does not guarantee any regularity for ϕ or ψ, however. The "messy" example in Fig. 7, e.g., satisfies this positivity condition.

3. Using the special form (3.5) of H and G one can show [24] that C2 is already implied by C1.

4. The condition $C1$ can easily be rewritten in z-transform-language. Let us associate, to any sequence $c = (c_n)_{n \in \mathbb{Z}}$, the function $c(z) = \sum_n c_n z^n$. Then the definition of the operator H as a convolution followed by a decimation implies

$$(Hc)(z^2) = \frac{1}{2}\left[\mathcal{H}(z)c(z) + \mathcal{H}(-z)c(-z)\right],$$

   where $\mathcal{H}(z) = \sum_n h(-n)z^n$.

Similarly the z-transform of H^*c is

$$(H^*c)(z) = \overline{\mathcal{H}(z)}\, c(z^2) .$$

We also define $\mathcal{G}(z) = \sum_n g(-n)z^n$. Condition C1 gives then

$$c(z) = (H^*Hc + G^*Gc)(z)$$

$$= \frac{1}{2}[|\mathcal{H}(z)|^2 + |\mathcal{G}(z)|^2]c(z) + [\overline{\mathcal{H}(z)}\,\mathcal{H}(-z) + \overline{\mathcal{G}(z)}\,\mathcal{G}(-z)]c(-z),$$

or

$$|\mathcal{H}(z)|^2 + |\mathcal{G}(z)|^2 = 2, \quad \overline{\mathcal{H}(z)}\,\mathcal{H}(-z) + \overline{\mathcal{G}(z)}\,\mathcal{G}(-z) = 0 .$$

This amounts to requiring that the 2×2 matrix

$$\begin{pmatrix} \frac{1}{\sqrt{2}}\mathcal{H}(z) & \frac{1}{\sqrt{2}}\mathcal{H}(-z) \\ \frac{1}{\sqrt{2}}\mathcal{G}(z) & \frac{1}{\sqrt{2}}\mathcal{G}(-z) \end{pmatrix}$$

is unitary. Note that this implies $\overline{\mathcal{H}(z)}\,\mathcal{G}(z) + \overline{\mathcal{H}(-z)}\,\mathcal{G}(-z) = 0$, which can easily be shown to be equivalent with C2. This is another way of proving that C1 implies C2.

5. Filters satisfying conditions C1-C3 had been constructed before by Smith and Barnwell [26]. They call these filters "conjugate quadrature filters" (CQF) as a special case of the "quadrature mirror filters" (QMF) of Esteban and Galand [27]. CQF give exact reconstruction, without any aliasing, as all QMF do, but also without any amplitude or phase distortion. For their purposes, they do not impose the regularity condition C4, and their filters are therefore not equivalent to an orthonormal wavelet basis in general.

6. While other solutions to C1 exist, it is convenient to choose the $g(n)$ such that

$$g(n) = (-1)^n h(-n + 1) . \tag{3.12}$$

This choice reduces condition C1 to an equation for $\sum_n h(n)e^{in\xi}$ [24]. It is the analog of the correspondence (2.9)-(2.10) between ϕ and ψ.

7. One way to ensure that the regularity condition C4 is satisfied is to impose that, for some $N \geq 2$,

$$\sum_n h(n)\, e^{in\xi} = \left(\frac{1 + e^{in\xi}}{2}\right)^N Q(\xi), \tag{3.13}$$

where $\sup_{\xi \in \mathbb{R}} |Q(\xi)| < 2^{N-1/2}$. In order for (3.12) to satisfy C1-C3 it is necessary and sufficient that

$$|Q(\xi)|^2 = \sum_{j=0}^{N-1} \binom{N-1+j}{j} (\sin^2 \xi/2)^j + (\sin^2 \xi/2)^N R(\frac{1}{2}\cos\xi), \tag{3.14}$$

where R is a real, odd function (see [24]). In order to derive (3.14) we assume that the $g(n)$ are defined by (3.12).

4. Orthonormal bases of wavelets with finite support.

In the preceding paragraph we saw that every CQF, i.e. every pair of filters satisfying C1-C3, which also satisfies the regularity condition C4, automatically defines an associated orthonormal wavelet basis for $L^2(\mathbb{R})$; where the function ϕ is the limit of the piecewise constant functions representing the $(H^*)^\ell e$, for $\ell \to \infty$.

On the other hand, it is clear from Figure 5 that ϕ will have a compact support, i.e. will vanish outside a finite interval, if only finitely many $h(n)$ are different from zero. More concretely, if the $h(n)$ are $\neq 0$ only for $0 \leq n \leq N$, then the piecewise constant function representing H^*e will be concentrated on $[-1/4, N/2 + 1/4]$, the piecewise constant function for $(H^*)^2 e$ on $[-1/8, N/2 + N/4 + 1/8], \ldots$. In general $(H^*)^\ell e$ corresponds to a piecewise constant function on $[-2^{-\ell-1}, N(1 - 2^{-\ell}) + 2^{-\ell-1}]$; the limit function ϕ is therefore concentrated on $[0, N]$ (see Fig. 5). If we define the $g(n)$ by (3.12), then only finitely many $g(n)$ will be nonvanishing as well, implying that ψ, as a finite linear combination of the compactly supported $\phi(2x - n)$, has compact support too. One checks that if $h(n) \neq 0$ only for $0 \leq n \leq N$, then ψ is concentrated on $[(1 - N)/2, (1 + N)/2]$ (see [24]).

In order to construct orthonormal bases of wavelets with finite support, it is therefore sufficient to construct filters with only finitely many coefficients or "taps" which satisfy the conditions C1-C4. A family of examples is given by the $h(n)$ defined by (3.13), with $R \equiv 0$ in (3.14). One also needs a procedure to determine the polynomial $Q(\xi)$ from its squared modulus $|Q(\xi)|^2$ (3.14); this procedure is given by a lemma of Riesz' [28][24].

Figure 8 shows a few examples of compactly supported wavelet bases obtained in this way. In each case both ϕ and ψ are plotted. The figure shows clearly that ϕ, ψ become more

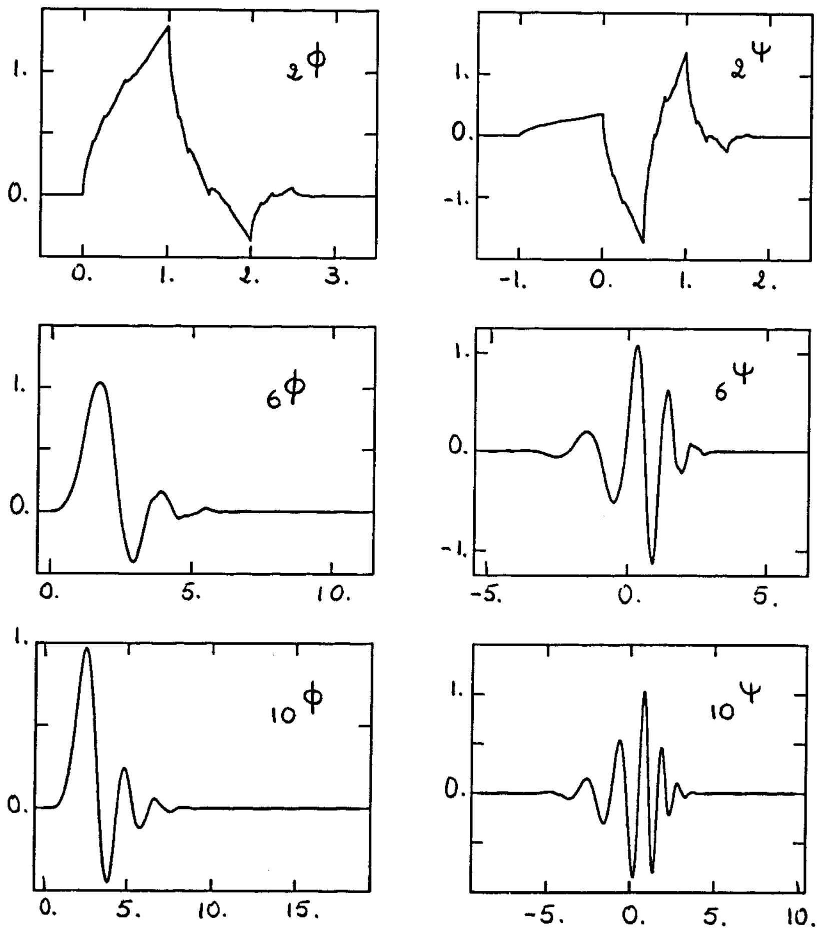

Figure 8. A few examples of functions ϕ, ψ giving rise to orthonormal bases of wavelets, corresponding to (3.13), (3.14), with $R \equiv 0$. It is clear that $_N\phi, _N\psi$ become more regular for larger values of N.

regular as N (see (3.13)) increases. The case $N = 1$ (not plotted) corresponds to the discontinuous Haar basis, where $\psi(x) = 1$ for $0 \leq x < 1/2$, -1 for $1/2 \leq x < 1$, 0 otherwise. The next case, $N = 2$, leads to $h(0) = (1 + \sqrt{3})/4\sqrt{2}$, $h(1) = (3 + \sqrt{3})/4\sqrt{2}$, $h(2) = (3 - \sqrt{3})/4\sqrt{2}$, $h(3) = (1 - \sqrt{3})/4\sqrt{2}$, and all other $h(n) = 0$. The corresponding ϕ and ψ are continuous but not C^1; they turn out to be Hölder continuous with exponent $\gamma = \ln(1 + \sqrt{3})/\ln 2 = .5500...$ (see § below). One can prove [24] that the regularity of ϕ, ψ in this family of examples increases linearly with N, i.e. there exists $\mu > 0$ such that $_N\phi, _N\psi \in C^{\mu N}$ for all $N \geq 2$.

In this family of examples the size of the support of ϕ, ψ is thus determined by the desired regularity. It turns out that this is a general feature, and that a linear relationship between these two quantities (regularity and support width) is the best one can hope for. More precisely, one can prove [29a]

Theorem:

If $\phi \in C^k$, support $\phi \subset [0, N]$ and $\phi(x) = \sum_{n=0}^{N} a_n \phi(2x - n)$, then $N \geq k + 2$.

The proof is so simple that we include it here.

Proof.

1. Let $v_0 \in \mathbb{R}^{N-1}$ be the vector $(v_0)_j = \phi(j)$. The equation for ϕ implies the existence of a matrix A, completely determined by the a_n, such that $v_0 = Av_0$.

2. Since $\phi'(x) = 2 \sum_{n=0}^{n} a_n \phi'(2x - n)$, the vector v_1 defined by $(v_1)_j = \phi'(j)$ satisfies $v_1 = 2Av_1$. Analogously one defines $v_2, \ldots, v_k$, each satisfying $v_j = 2^j Av_j$. Moreover none of the v_j can be zero, since $v_j = 0$ would imply $\phi^{(j)}(x) = 0$ for all x of the type $k2^{-\ell}$ (by iteration of the equation for ϕ), which leads to $\phi^{(j)} \equiv 0$. This is, however, incompatible with the finite support condition on ϕ.

3. It follows that the $(N - 1) \times (N - 1)$ matrix A has at least the $(k + 1)$ eigenvalues 1, 1/2, , ..., 2^{-k}. Hence $N \geq k + 2$. ■

5. Fractal properties.

The graph of ${}_2\phi$ (see Figure 8) exhibits a certain "jaggedness" that seems to repeat itself in a self-similar way at smaller scales. This is made even clearer by the blowups in Figure 9.

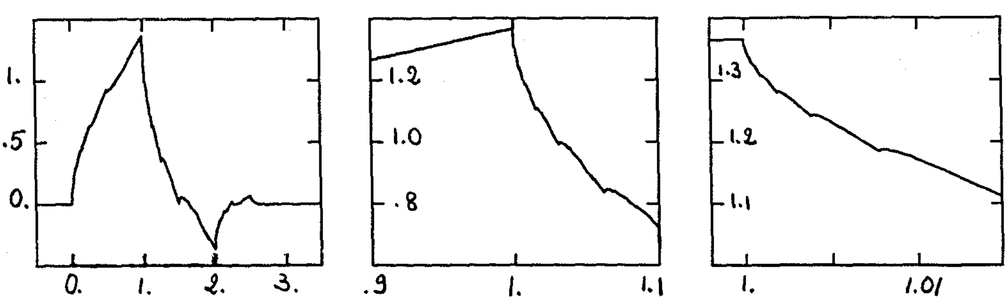

Figure 9. The function ${}_2\phi(x)$ (see Fig. 8) and two successive blow-ups of its behavior around $x = 1$. Analogous self-similar patterns repeat itself, on smaller and smaller scales, near every dyadic rational point, i.e. near every x of the form $k2^{-\ell}$, $0 \leq x < 3$.

A closer study of ${}_2\phi$ reveals a very rich structure, although the function is not C^1, it is differentiable almost everywhere. In fact, if the binary expansion of $x \in [0, 3]$ contains more than (roughly) 25% of digits 1, then ϕ is differentiable in x. (We shall make this statement more precise below). Since almost all numbers have 50% of the digits in their binary expansion equal to 1, this implies that ϕ is almost everywhere differentiable. Let us see how such properties can be derived.

We know that support $\phi = [0, 3]$ and that

$$\phi(x) = a_0\phi(2x) + a_1\phi(2x-1) + a_2\phi(2x-2) + a_3\phi(2x-4), \tag{5.1}$$

where $a_j = \sqrt{2}\,h(j)$, or $a_0 = (1+\sqrt{3})/4$, $a_1 = (3+\sqrt{3})/4$, $a_2 = (3-\sqrt{3})/4$, $a_3 = (1-\sqrt{3})/4$ (see §4). It follows that

$$\begin{cases} a_0 + a_2 = a_1 + a_3 = 1 \\ a_1 + 3a_3 = 2a_2 \,. \end{cases} \tag{5.2}$$

For $x \in [0, 1]$, we define $v(x) \in \mathbb{R}^3$ by

$$v(x) = \begin{pmatrix} \phi(x) \\ \phi(x+1) \\ \phi(x+2) \end{pmatrix}.$$

From (5.1) one easily checks that

$$\begin{aligned} &\text{for } x \le 1/2: \quad v(x) = T_0 v(2x) \\ &\text{for } x \ge 1/2: \quad v(x) = T_1 v(2x-1)\,, \end{aligned} \tag{5.3}$$

where

$$T_0 = \begin{pmatrix} a_0 & 0 & 0 \\ a_2 & a_1 & a_0 \\ 0 & a_3 & a_2 \end{pmatrix} \quad \text{and} \quad T_1 = \begin{pmatrix} a_1 & a_0 & 0 \\ a_3 & a_2 & a_1 \\ 0 & 0 & a_3 \end{pmatrix}.$$

The matrices T_0, T_1 have very special properties. In particular, they both have eigenvalue 1, with a common left eigenvector,

$$(1,1,1)\,T_0 = (1,1,1) = (1,1,1)\,T_1\,. \tag{5.4}$$

Similarly one computes (use (5.2))

$$(1,2,3)T_0 = \frac{1}{2}(1,2,3) + (\frac{1}{2} + a_2)(1,1,1)$$
$$(1,2,3)T_1 = \frac{1}{2}(1,2,3) + a_2(1,1,1)\,. \tag{5.5}$$

This implies that T_0, T_1 both have eigenvalue 1/2, and that they have a common left invariant subspace, associated to the eigenvalues 1 and 1/2.

Note.

The functions ${}_N\phi$ obtained from (3.13) with $R \equiv 0$ in (3.14), are associated to $(2N-1) \times (2N-1)$ matrices T_0, T_1 in exactly the same way as in the case $N = 2$. For general N, the martices T_0, T_1 have N common eigenvalues $1, 1/2, \ldots, 2^{-N+1}$. If we define the row vectors $u_j, j = 0, \ldots, N-1$ by $u_j = (1^j, 2^j, \ldots, (2N-1)^j)$, then the left eigenvectors for both T_0, T_1 for the eigenvalue 2^{-j} always lie in the subspace $U_j = \mathrm{span}\{u_0, u_1, \ldots, u_j\}$. (Full details are given in [29b]).

These spectral properties of T_0, T_1 have several consequences. It follows, e.g., that, for all $x \in [0,1]$,

$$\phi(x) + \phi(x+1) + \phi(x+2) = 1 \tag{5.6}$$

Proof.

1. We prove this only for x of the type $k2^{-\ell}$, with $k, \ell \in \mathbb{N}$. By continuity the result then follows for all x.

2. For $\ell = 0$, we have $\phi(0) = \phi(3) = 0$. Hence

$$\phi(x) + \phi(x+1) + \phi(x+2) = \phi(1) + \phi(2)$$

for all $x = k2^{-\ell} \in [0,1]$, with $\ell = 0$ ($k = 0, 1$ are the only possibilities).

3. Take any $\ell > 0$. Then

$$\phi(x) + \phi(x+1) + \phi(x+2) = u_0 \cdot v(x)$$

$$= \begin{cases} u_0 \cdot T_0 v(2x) = u_0 \cdot v(2x) & \text{if } x \le 1/2 \\ u_0 \cdot T_1 v(2x-1) = u_0 \cdot v(2x-1) & \text{if } x \ge 1/2\,. \end{cases}$$

Since $2x$ or $2x - 1$ are of the type $k2^{-\ell+1}$, we conclude by induction that $u_0 \cdot v(x) = \phi(1) + \phi(2)$ for all x of type $k2^{-\ell}$. By continuity $u_0 \cdot v(x) = \phi(1) + \phi(2)$ for all $x \in [0,1]$.

4. It then follows that $\int_0^3 dx\, \phi(x) = \int_0^1 dx\, [u_0 \cdot v(x)] = \phi(x) + \phi(2)$. Since $\int dx\, \phi(x) = 1$, this implies $\phi(1) + \phi(2) = 1$, hence (5.6). ■

Similarly, using (5.5) one proves, for all $x \in [0, 1]$

$$\phi(x) + 2\phi(x+1) + 3\phi(x+2) = -x + (1 + a_2)\ . \tag{5.7}$$

Note.

In the general case ($N \geq 2$), we find that ${}_N\phi$ satisfies N such sum rules, one for each $u_j \cdot v(x)$, $j = 0, \ldots, N-1$.

One can also use (5.3) to study the local behavior of ϕ in the neighborhood of a point x. For any $x \in [0, 1]$, we write the binary expansion of x, e.g.

$$x = .1011001011100 \cdots$$

Define then τx to be given by the same binary expansion, except for the first digit, which is dropped,

$$\tau x = .011001011100 \cdots$$

It follows that $\tau x = 2x$ if $x < 1/2$ and $\tau x = 2x - 1$ if $x > 1/2$. Consequently (5.3) can be rewritten as

$$v(x) = T_{d_1(x)}\, v(\tau x)\,, \tag{5.8}$$

where $d_j(x)$ denotes the j-th digit in the binary expansion of x. Note that the binary expansion is not ambiguously defined for dyadic rationals x, i.e. for x of the type $k2^{-\ell}$. For $x = 1/2$, e.g., both the expansions .0111111... and .1000000... are admissible. Consequently $\tau 1/2$ is not well-defined, giving the answer 1 or 0 according to the chosen binary expansion. One easily checks, however, that $T_1 v(0) = T_0 v(1)$ (use (5.6) and (5.7)), so that (5.8) holds, even for $x = 1/2$, regardless of the choice of binary expansion. It is easy to convince oneself that (5.8) and its iterates never lead to contradictions at dyadic rationals x.

Iterating (5.8) leads to

$$v(x) = \mathbf{T}_m(x)\, v(\tau^m x)\,, \tag{5.9}$$

where

$$\mathbf{T}_m(x) = T_{d_1(x)} T_{d_2(x)} \cdots T_{d_m(x)}\ .$$

Similarly, for t small enough so that the binary expansion of $x + t$ has the same m first digits as the expansion for x,

$$v(x + t) = \mathbf{T}_m(x)\, v(\tau^m x + 2^m t)\ . \tag{5.10}$$

In order to estimate the difference $v(x + t) - v(x)$, we use the spectral decomposition of $\mathbf{T}_m(x)$. From (5.4) it follows that $u_0 = (1, 1, 1)$ is a left eigenvector for $\mathbf{T}_m(x)$, $e_1(m, x) = u_0$ with eigenvalue 1,

$$e_1(m, x)\, \mathbf{T}_m(x) = e_1(m, x)\ .$$

Because of (5.5) one finds that $\mathbf{T}_m(x)$ also has eigenvalue 2^{-m}. The corresponding left eigenvector $e_2(m, x)$ is a linear combination of u_0 and $u_1 = (1, 2, 3)$,

$$e_2(m, x)\, \mathbf{T}_m(x) = 2^{-m} e_2(m, x)\ ,$$

with

$$e_2(m, x) = u_1 + [(1 - 2^{-m})^{-1} (x - 2^{-m}\tau^m x) - (1 + a_2)]\, u_0\ .$$

The third eigenvalue of $\mathbf{T}_m(x)$ can be computed from its determinant. Defining $r_m(x) = m^{-1} \sum_{j=1}^{m} d_j(x)$ to be the average number of digits 1 in the first m digits of the expansion for x, we find

$$\begin{aligned} \det \mathbf{T}_m(x) &= (\det T_1)^{m r_m(x)}\, (\det T_0)^{m(1-r_m(x))} \\ &= 2^{-m}\, a_3^{m r_m(x)}\, a_0^{m(1-r_m(x))}\ . \end{aligned}$$

It follows that the third eigenvalue of $\mathbf{T}_m(x)$ is $\lambda_m(x) = a_3^{m r_m(x)}\, a_0^{m(1-r_m(x))}$. One can find explicit expressions for the corresponding left eigenvector $e_3(m, x)$, as well for the three right (column) eigenvectors $\tilde{e}_j(m, x)$ of $\mathbf{T}_m(x)$, but these are not really necessary. It is sufficient to know (this is proved in [29b]) that they are all uniformly bounded in m and x. For any $v \in \mathbb{R}^3$ we have

$$\begin{aligned} \mathbf{T}_m(x) v = {} & \tilde{e}_1(m, x)\, [e_1(m, x) \cdot v] + 2^{-m} \tilde{e}_2(m, x) [e_2(m, x) \cdot v] \\ & + \lambda_m(x)\, \tilde{e}_3(m, x)\, [e_3(m, x) \cdot v]\ . \end{aligned}$$

Applying this to (5.8), (5.9), and using (5.6), (5.7) we find thus, for sufficiently small t,

$$v(x) = \tilde{e}_1(m, x) + 2^{-m}[(1 - 2^{-m})^{-1}(x - 2^{-m}\tau^m x) - \tau^m x]\, \tilde{e}_2(m, x)$$

$$+ \lambda_m(x) \tilde{e}_3(m, x) [e_3(m, x) \cdot v(\tau^m x)]$$

$$v(x + t) = \tilde{e}_1(m, x) + 2^{-m}[(1 - 2^{-m})^{-1}(x - 2^{-m}\tau^m x) - \tau^m x - 2^m t] \tilde{e}_2(m, x)$$

$$+ \lambda_m(x) \tilde{e}_3(m, x) [e_3(m, x) \cdot v(\tau^m x + 2^m t)] .$$

Hence

$$v(x + t) - v(x) = -t \tilde{e}_2(m, x) + \lambda_m(x) \tilde{e}_3(m, x) [e_3(m, x) \cdot (v(\tau^m x + 2^m t) - v(\tau^m x))]$$

For all $x \in [0, 1]$ such that there exists a limit for the average incidence of digits 1 in the binary expansion,

$$r(x) = \lim_{m \to \infty} r_m(x)$$

and such that $0 < r(x) < 1$, one can easily show that "sufficiently small t" means $t \leq 2^{-m(1+\epsilon)}$, where $\epsilon > 0$ can be chosen arbitrarily small, for large enough m. Choosing t such that $2^{-(m+1)(1+\epsilon)} \leq t \leq 2^{-m(1+\epsilon)}$, we find then

$$t^{-1}[v(x + t) - v(x)] = -\tilde{e}_2(m, x) + R(m, x) , \tag{5.11}$$

where the remainder term $R(m, x)$ is bounded by

$$\|R(m, x)\| \leq C\, 2^{m(1+\epsilon)} \, |\lambda_m(x)| .$$

If $f(x) > (\log 2a_0)/(\log a_0 - \log |a_3|) \cong .2368 \cdots$, then $2^{m(1+\epsilon)} |\lambda_m(x)| \underset{m \to \infty}{\sim} [2^{1+\epsilon} a_0^{1-r(x)} |a_3|^{r(x)}]^m$ tends to zero for $m \to \infty$, if ϵ is small enough. The second term in (5.11) can therefore be neglected for large enough m. On the other hand $\tilde{e}_2(m, x)$ tends towards a limit as $m \to \infty$ (see [29b]). It follows that $v(x)$ is differentiable for all x such that $r(x)$ is well-defined and $.2368 \cdots < r(x) < 1$, which implies that $\phi(x)$, $\phi(x + 1)$, $\phi(x + 2)$ are differentiable as well.

The same technique, i.e. the spectral analysis of $\mathbf{T}_m(x)$, can be used to prove that ϕ is Hölder continuous, with exponent $2 - \ln(\sqrt{3} + 1)/\ln 2 \cong .5500\ldots$. This exponent is the best possible one. One can also analyze the behavior of ϕ near dyadic rational points. As approached from below, the binary expansion near a dyadic point has a tail of only $1 - s$; as approached from above, the expansion has only $0 - s$ (see above). The result is that ϕ is left-differentiable but not right-differentiable at every dyadic rational point. This can clearly be seen of Figure 9, at $x = 1$.

A similar analysis can be carried out for the ${}_N\phi$ corresponding to higher values of N; see [29b].

References.

1. See e.g. J. Morlet, G. Arens, I. Fourgeau and D. Giard, "Wave propagation and sampling theory," Geophysics **47** (1982) 203-236.
2. A. Grossman and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape", SIAM J. Math. Anal. **15** (1984) 723-736.
   P. Goupillaud, A. Grossmann and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis", Geoexploration **23** (1984) 85.
3. The wavelet transform is implicitly used in A. Calderón, "Intermediate spaces and interpolation, the complex method", Studia Math. **24** (1964) 113-190. It appears more explicitly in e.g. A. Calderón and A. Torchinsky, "Parabolic maximal functions associated to a distribution, I", Adv. Math. **16** (1975) 1-64.
4. An application to a singular integral operator relevant for quantum mechanics can be found in C. Fefferman and R. de la Llave, "Relativistic stability of matter," Rev. Mat. Iberoamericana **2** (1986).
5. J. R. Klauder and B.-S. Skagerstam, "Coherent States", World Scientific (Singapore) 1985.
6. E. W. Aslaksen and J. R. Klauder, "Unitary representations of the affine group," J. Math. Phys. **9** (1968) 206-211; "Continuous representation theory using the affine group", J. Math. Phys. **10** (1969) 2267-2275.
7. T. Paul, "Affine coherent states and the radial Schrödinger equation I. Radial harmonic oscillator and the hydrogen atom", to be published.
8. K. G. Wilson and J. B. Kogut, Physics Reports **12C** (1974) 77.
   J. Glimm and A. Jaffe, "Quantum physics: a functional integral point of view", Springer (New York) 1981.
9. G. Battle and P. Federbush, "Ondelettes and phase cell cluster expansions: a vindication", Comm. Math. Phys. **109** (1987) 417-419.
10. I. Daubechies, "The wavelet transform, time-frequency localisation and signal analysis", to be published in IEEE Trans. Inf. Theory.
11. R. Balian, "Un principe d'incertitude fort en théorie du signal ou en mécanique quantique", C. R. Acad. Sc. Paris **292**, série 2 (1981) 1357-1362.

12. F. Low, "Complex sets of wave-packets" in "A passion for physics — Essays in honor of G. Chew", World Scientific (Singapore) 1985, pp. 17-22.

13. G. Battle, "Heisenberg proof of the Balian-Low theorem", to be published in Lett. Math.Phys.

14. D. Gabor, "Theory of communication", J. Inst. Elec. Eng. (London) **93** III (1946) 429-457.

15. M. J. Bastiaans, "A sampling theorem for the complex spectrogram and Gabor's expansion of a signal in Gaussian elementary signals", Optical Eng. **20** (1981) 594-598.

16. A.J.E. M. Janssen, "Gabor representation of generalized functions," J. Math. Appl. **80** (1981) 377-394.

17. Y. Meyer, "Principe d'incertitde, bases hilbertiennes et algèbres d'opérateurs", Séminaire Bourbaki, 1985-1986, nr.662.

18. P. G. Lemarié and Y. Meyer, "Ondelettes et bases hilbertiennes", Rev. Mat. Iberoamericana **2** (1986) 1-18.

19. G. Battle, "A block spin construction of ondelettes. I: Lemarié functions," Comm. Math. Phys. **110** (1987) 601-615.

20. P. Lemarié, "Ondelettes à localisation exponentielle", to be published in J. de Math. Pures et Appl.

21. S. Mallat, "Multiresolution approximation and wavelets", to be published.

22. S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", to be published in IEEE Trans. on Pattern Analysis and Machine Intelligence.

23. P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code", IEEE Trans. Comm. **31** (1983) 582-540, and "A multiresolution spline with application to image mosaics", ACM Trans. on Graphics, **2** (1983) 217-236.

24. I. Daubechies, "Orthonormal bases of compactly supported wavelets", Comm. Pure & Appl. Math. **49** (1988) 909-996.

25. Y. Meyer, "Ondelettes et fonctions splines", Séminaire E.D.P., Ecole Polytechnique, Paris, France, December 86.

26. M. J. Smith and D. P. Barnwell, "Exact reconstruction techniques for tree-structured subband coders", IEEE Trans. on ASSP **34** (1986) 434-441.

27. D. Esteban and C. Galand, "Application of quadrature minor filters to split band voice coding schemes", Proc. Int. Conf. ASSP (1977) 191-195.

28. G. Polya and G. Szegö, "Aufgaben und Lehrsätse aus der Analysis" Vol. II, Springer (Berlin) 1971.

29a. I. Daubechies and J. Lagarias, "Two-scale difference equations: I. Global regularity of solutions", preprint AT&T Bell Laboratories.

29b. ____, "Two-scale difference equations: II Infinite products of matrices, local regularity and fractals.", preprint AT&T Bell Laboratories.