Orthonormal Bases of Compactly Supported Wavelets

INGRID DAUBECHIES
AT&T Bell Laboratories

Abstract

We construct orthonormal bases of compactly supported wavelets, with arbitrarily high regularity. The order of regularity increases linearly with the support width. We start by reviewing the concept of multiresolution analysis as well as several algorithms in vision decomposition and reconstruction. The construction then follows from a synthesis of these different approaches.

1. Introduction

In recent years, families of functions $h_{a,b}$,

$$h_{a,b}(x) = |a|^{-1/2} h\left(\frac{x-b}{a}\right), \qquad a, b \in \mathbb{R},\ a \neq 0, \tag{1.1}$$

generated from one single function h by the operation of dilations and translations, have turned out to be a useful tool in many different fields of mathematics, pure as well as applied. Following Grossmann and Morlet [1], we shall call such families "wavelets".

Techniques based on the use of translations and dilations are certainly not new. They can be traced back to the work of A. Calderón [2] on singular integral operators, or to renormalization group ideas (see [3]) in quantum field theory and statistical mechanics. Even in these two disciplines, however, the explicit introduction of special families of wavelets seems to have led to new results (see, e.g. [4], [5], [6]). Moreover, wavelets are useful in many other applications as well. They are used for e.g. sound analysis and reconstruction in [7], and have led to a new algorithm, with many attractive features, for the decomposition of visual data in [8]. They seem to hold great promise for the detection of edges and singularities; see [9]. It is therefore fair to surmise that they will have applications in yet other directions.

Depending on the type of application, different families of wavelets may be chosen. One can choose, e.g., to let the parameters a, b in (1.1) vary continuously on their range $\mathbb{R}^* \times \mathbb{R}$ (where $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$). One can then, for instance, represent functions $f \in L^2(\mathbb{R})$ by the functions Uf,

$$(Uf)(a,b) = \langle h_{a,b}, f \rangle = |a|^{-1/2} \int dx\, \overline{h\left(\frac{x-b}{a}\right)} f(x). \tag{1.2}$$

Communications on Pure and Applied Mathematics, Vol. XLI 909–996 (1988)
 CCC 0010-3640/88/070909-88$04.00

If h satisfies the condition

$$\int d\xi\, |\xi|^{-1} |\hat{h}(\xi)|^2 < \infty, \tag{1.3}$$

where $\hat{}$ denotes the Fourier transform,

$$\hat{h}(\xi) = \frac{1}{\sqrt{2\pi}} \int dx\, e^{ix\xi} h(x),$$

then U (as defined by (1.2)) is an isometry (up to a constant) from $L^2(\mathbb{R})$ into $L^2(\mathbb{R}^* \times \mathbb{R};\ a^{-2}\, da\, db)$. The map U is called the "continuous wavelet transform"; see [1], [10]. In this form, wavelets are closest to the original work of Calderón. The continuous wavelet transform is also closely related to the "affine coherent state representation" of quantum mechanics (first constructed in [11], see also [12]); in fact, for appropriate choices of h, the $h_{a,b}$ are "affine coherent states", and have been used in the study of some quantum mechanics problems in [11], [12].

Note that the "admissibility condition" (1.3) implies, if h has sufficient decay which we shall always assume in practice, that h has mean zero,

$$\int dx\, h(x) = 0. \tag{1.4}$$

Typically, the function h will therefore have at least some oscillations. A standard example is

$$h(x) = (2/\sqrt{3})\pi^{-1/4}(1 - x^2)e^{-x^2/2}. \tag{1.5}$$

For other applications, including those in signal analysis, one may choose to restrict the values of the parameters a, b in (1.1) to a discrete sublattice. In this case one fixes a dilation step $a_0 > 1$, and a translation step $b_0 \neq 0$. The family of wavelets of interest becomes then, for $m, n \in \mathbb{Z}$,

$$h_{mn}(x) = a_0^{-m/2} h(a_0^{-m}x - nb_0). \tag{1.6}$$

Note that this corresponds to the choices

$$a = a_0^m,$$

$$b = nb_0 a_0^m,$$

indicating that the translation parameter b depends on the chosen dilation rate. For m large and positive, the oscillating function h_{m0} is very much spread out, and the large translation steps $b_0 a_0^m$ are adapted to this wide width. For large but

negative m the opposite happens; the function h_{m0} is very much concentrated, and the small translation steps $b_0 a_0^m$ are necessary to still cover the whole range.

A "discrete wavelet transform" T is associated with the discrete wavelets (1.6). It maps functions f to sequences indexed by $\mathbb{Z}^2$,

$$\begin{aligned}(Tf)_{mn} &= \langle h_{mn}, f\rangle \\ &= a_0^{-m/2}\int dx\, \overline{h(a_0^{-m}x - nb_0)}\, f(x).\end{aligned} \tag{1.7}$$

If h is "admissible", i.e., if h satisfies the condition (1.3), and if h has sufficient decay, then T maps $L^2(\mathbb{R})$ into $l^2(\mathbb{Z}^2)$. In general, T does not have a bounded inverse on its range. If it does, i.e., if, for some $A > 0$, $B < \infty$,

$$A\|f\|^2 < \sum_{m,n\in\mathbb{Z}} |\langle h_{mn}, f\rangle|^2 < B\|f\|^2,$$

for all f in $L^2(\mathbb{R})$, then the set $\{h_{mn};\ m, n \in \mathbb{Z}\}$ is called a "frame". In this case one can construct numerically stable algorithms to reconstruct f from its wavelet coefficients $\langle h_{mn}, f\rangle$. In particular,

$$f = \frac{2}{A+B}\sum_{m,n} h_{mn}\langle h_{mn}, f\rangle + R, \tag{1.8}$$

with

$$\|R\| \leqq O\left(\frac{B}{A} - 1\right)\|f\|.$$

For B/A close to 1, which is the case in the decompositions and reconstructions of music and other sound signals, as done by A. Grossmann, R. Kronland and J. Morlet [7], the "error term" R can be omitted. In practice, with e.g. the basic wavelet (1.5), and with $a_0 = 2^{1/4}$, $b_0 = .5$, one finds $B/A - 1 < 10^{-5}$, and the reconstruction formula (1.8) restricted to its first term gives excellent results. In fact, even for the larger value $a_0 = 2$, corresponding to $B/A - 1 \cong .08$, the truncated reconstruction formula, when applied to the wavelet decomposition of speech signals, leads to a clearly understandable reconstruction; see [13].

In the use of wavelet frames for sound analysis, and reconstruction, as studied by the Marseilles group [7], the families of wavelets h_{mn} considered are highly redundant, i.e., they are not independent, in the sense that any finite number of them lies in the closed linear span generated by the others. Consequently, the range of the discrete wavelet transform T is a proper subspace of $l^2(\mathbb{Z}^2)$. The higher the redundancy of the frame, the smaller this subspace, which is a desirable feature for some purposes (e.g. the reduction of calculational noise). If a_0, b_0 are chosen very close to 1, 0, respectively, then the resulting frame is very

redundant and very close to the continuous family of wavelets (1.1); this type of frame was used in the "edge detection" study mentioned earlier; see [9].

For other applications, as e.g. in S. Mallat's vision decomposition algorithm in [8], it is preferable to work with the other extremum, and to reduce redundancy as much as possible. In this case, one can turn to choices of h and a_0, b_0 (typically $a_0 = 2$) for which the h_{mn} constitute an orthonormal basis. This is the case to which we shall be restricting ourselves in the remainder of this paper. For a more detailed study of general (non-orthonormal) wavelet frames, and a discussion of the similarities and the differences between wavelet transform and windowed Fourier transform, the reader is referred to [14], [15].

One example of an orthonormal basis of wavelets[1] for $L^2(\mathbb{R})$ is the well-known Haar basis. For the Haar basis one chooses

$$h(x) = \begin{cases} 1, & 0 \leqq x < \frac{1}{2}, \\ -1, & \frac{1}{2} \leqq x < 1, \\ 0, & \text{otherwise}, \end{cases} \tag{1.9}$$

and $a_0 = 2$, $b_0 = 1$. The resulting h_{mn},

$$h_{mn}(x) = 2^{-m/2} h(2^{-m}x - n), \qquad m, n \in \mathbb{Z}, \tag{1.10}$$

constitute an orthonormal basis for $L^2(\mathbb{R})$. The h_{mn} also constitute an unconditional basis for all $L^p(\mathbb{R})$, $1 < p < \infty$.

Recently, some much more surprising examples of orthonormal wavelet bases have surfaced. The first one was constructed by Y. Meyer [4] in the summer of 1985. He constructed a C^∞-function h of rapid decay (in fact $\hat{h}$, in his example, is a compactly supported C^∞-function) such that the h_{mn}, as defined by (1.10) (i.e., with $a_0 = 2$, $b_0 = 1$), constitute an orthonormal basis for $L^2(\mathbb{R})$. As in the case of the Haar basis, Y. Meyer's basis is also an unconditional basis for all the L^p spaces, $1 < p < \infty$. Much more is true, however. The Meyer basis turns out to be an unconditional basis for all the Sobolev spaces, for the Hardy-Littlewood space H_1, for the Besov spaces, etc.; see [4]. The Meyer basis is therefore a much more powerful tool than the Haar basis.

Some time later, in 1986, another interesting orthonormal basis of wavelets was constructed, independently, by P. G. Lemarié [17] and G. Battle [18]. In their construction the function h is only C^k, but it has exponential decay (as compared

[1] Following Grossmann and Morlet [1] we call "wavelet" any L^2-function h satisfying the admissibility condition (1.3). This is less restrictive than Y. Meyer [16], who, in keeping with the tradition in harmonic analysis, also imposes some regularity. In the terminology of [16], the Haar basis function (1.9) is not a wavelet.

with decay faster than any power in Y. Meyer's case). It also has k vanishing moments, i.e.,

$$\int dx\, x^j h(x) = 0, \qquad j = 0, 1, \cdots, k-1,$$

which makes these h_{mn} an unconditional basis for all the Sobolev spaces $\mathcal{H}_s$, with $s < k - 1$.

In all these constructions the choices $a_0 = 2$, $b_0 = 1$ were made. The choice for b_0 is of course arbitrary, a simple dilation of the function h allows one to fix any non-zero choice for b_0; it is convenient to choose $b_0 = 1$. The choice of a_0 is far less arbitrary. We shall restrict ourselves here to $a_0 = 2$, although it is possible to consider other, though by no means arbitrary, choices for a_0 (see [4], [21]).

In the fall of 1986, S. Mallat and Y. Meyer [16], [19] realized that these different wavelet basis constructions can all be realized by a "multiresolution analysis". This is a framework in which functions $f \in L^2(\mathbb{R}^d)$ can be considered as a limit of successive approximations, $f = \lim_{m \to -\infty} P_m f$, where the different $P_m f$, $m \in \mathbb{Z}$, correspond to smoothed versions of f, with a "smoothing out action radius" of the order of 2^m. The wavelet coefficients $\langle h_{mn}, f \rangle$, with fixed m, then correspond to the difference between the two successive approximations $P_{m-1} f$ and $P_m f$. A more detailed description of multiresolution analysis will be given in Section 2.

The concept of multiresolution analysis plays a central role in S. Mallat's algorithm for the decomposition and reconstruction of images in [8]. In fact, ideas related to multiresolution analysis (a hierarchy of averages, and the study of their differences) were already present in an older algorithm for image analysis and reconstruction, namely the Laplacian pyramid scheme of P. Burt and E. Adelson [20]. The Laplacian pyramid ideas triggered S. Mallat to view the orthonormal bases of wavelets as a vehicle for multiresolution analysis. Together, S. Mallat and Y. Meyer then carried out a more detailed mathematical analysis, showing how all the "accidental" previous constructions found their natural framework in multiresolution analysis; see [16], [19]. By the use of multiresolution analysis and orthonormal wavelet bases, S. Mallat constructed an algorithm that is both more economical and more powerful in its orientation selectivity. On the other hand, by a curious feedback, the combination of Mallat's ideas and of the restrictions on "filters" imposed in [20] led to my construction of orthonormal wavelet bases of compact support, which is the main topic of this paper.

Because of the important role, in the present construction, of the interplay of all these different concepts, and also to give a wider publicity to them, an extensive review will be given in Section 2 of multiresolution analysis (subsection 2A), of the Laplacian pyramid scheme (subsection 2B) and of Mallat's algorithm (subsection 2C).

Sections 3 and 4 contain the new results of this paper. A closer look at Mallat's work shows that he uses the intermediary of orthonormal wavelet bases

for *function* spaces to build an essentially *discrete* algorithm. It seemed therefore natural to wonder whether similar, and as powerful, discrete algorithms could be built *directly*, without using function spaces as an intermediate step. It turns out that it is very easy to write a set of necessary and sufficient conditions, on the "discrete side", ensuring that an algorithm similar to Mallat's works. This is done in subsection 3A. In order to have a useful algorithm, however, an extra regularity condition has to be imposed (this condition is already satisfied in e.g. Burt and Adelsen's Laplacian pyramid scheme). This is done in subsection 3B. The combination of the *discrete* conditions and the regularity condition on the *discrete* algorithm turns out, however, to be strong enough to *impose* an underlying multiresolution analysis of *functions*, with associated orthonormal wavelet basis. Provided the regularity condition is satisfied, there is therefore a one-to-one correspondence between orthonormal wavelet bases and discrete multiresolution decompositions, in the sense of Mallat's algorithm. This equivalence is proved in subsection 3C. Another proof of the same result, using different techniques, can be found in [19]; the proof presented here is more "graphical", and closer to the "filter" point of view of [20].

In Section 4, we exploit the equivalence between discrete and function schemes to build orthonormal bases of wavelets with compact support. Using this equivalence, it turns out that it is sufficient to build a discrete scheme using filters with a finite number of taps. This can be done explicitly, as shown in subsection 4B. As a result one can construct, for any $k \in \mathbb{N}$, a C^k-function ψ with compact support, such that the corresponding ψ_{mn},

$$\psi_{mn}(x) = 2^{-m/2}\psi(2^{-m}x - n),$$

constitute an orthonormal basis. The size of the support increases linearly with the regularity. Moreover, ψ has K consecutive moments equal to zero,

$$\int dx\, x^j \psi(x) = 0, \qquad j = 0, 1, \cdots, K-1,$$

where K also increases linearly with k. All these properties of the construction are proved in subsection 4C. Finally, the "graphical" approach which, as explained in subsection 3B, was the guideline to the proof of the link between the "regularity" condition of Burt and Adelson (see subsection 2B) and multiresolution analysis, can also be used to plot the functions ψ. We conclude this paper with the plots of a few of the compactly supported wavelets constructed here.

2. Multiresolution Analysis and Image Decomposition and Reconstruction

2.A. A review of multiresolution analysis and orthonormal wavelet bases. In this subsection we review the definition of multiresolution analysis, and show how orthonormal bases of wavelets can be constructed starting from a multiresolution analysis. We illustrate this construction with examples. No proofs will be given; for proofs, more details and generalizations we refer to [16], [19] or [21].

The idea of a multiresolution analysis is to write L^2-functions f as a limit of successive approximations, each of which is a smoothed version of f, with more and more concentrated smoothing functions. The successive approximations thus use a different resolution, whence the name multiresolution analysis. The successive approximation schemes are also required to have some translational invariance. More precisely, a multiresolution analysis consists of

(i) a family of embedded closed subspaces $V_m \subset L^2(\mathbb{R})$, $m \in \mathbb{Z}$,

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

such that (ii)

$$\bigcap_{m\in\mathbb{Z}} V_m = \{0\}, \qquad \overline{\bigcup_{m\in\mathbb{Z}} V_m} = L^2(\mathbb{R}), \tag{2.2}$$

and (iii)

$$f \in V_m \Leftrightarrow f(2\,\cdot) \in V_{m-1}; \tag{2.3}$$

moreover, there is a $\phi \in V_0$ such that, for all $m \in \mathbb{Z}$, the ϕ_{mn} constitute an unconditional basis for V_m, that is, (iv)

$$\overline{V_m = \text{linear span } \{\phi_{mn},\ n \in \mathbb{Z}\}} \tag{2.4a}$$

and there exist $0 < A \leqq B < \infty$ such that, for all $(c_n)_{n\in\mathbb{Z}} \in l^2(\mathbb{Z})$,

$$A\sum_n |c_n|^2 \leqq \left\| \sum_n c_n \phi_{mn} \right\|^2 \leqq B \sum_n |c_n|^2. \tag{2.4b}$$

Here $\phi_{mn}(x) = 2^{-m/2}\phi(2^{-m}x - n)$. Let P_m denote the orthogonal projection onto V_m. It is then clear from (2.1), (2.2) that $\lim_{m\to-\infty} P_m f = f$, for all $f \in L^2(\mathbb{R})$. The condition (2.3) ensures that the V_m correspond to different scales, while the translational invariance

$$f \in V_m \to f(\cdot - 2^m n) \in V_m \quad \text{for all} \quad n \in \mathbb{Z}$$

is a consequence of (2.4).

EXAMPLE 2.1. A typical though crude example is the following. Take the V_m to be spaces of piecewise constant functions,

$$V_m = \{ f \in L^2(\mathbb{R});\ f \text{ constant on } [2^m n, 2^m(n+1)[\text{ for all } n \in \mathbb{Z} \}.$$

The conditions (2.1)–(2.3) are clearly satisfied. The projections P_m are defined by

$$P_m f|_{[2^m n, 2^m(n+1)[} = 2^{-m} \int_{2^m n}^{2^m(n+1)} dx f(x).$$

The successive $P_m f$ (as m decreases) do therefore correspond to approximations of f on a finer and finer scale. Finally, we can choose for ϕ the characteristic function of the interval $[0, 1[$,

$$\phi(x) = \begin{cases} 1, & 0 \leqq x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, $\phi \in V_0$ and $V_m = \overline{\text{span}\{\phi_{mn}\}}$.

In what follows, we shall revisit this example to illustrate the construction of an orthonormal wavelet basis from multiresolution analysis.

Note that, in view of (2.3), the condition (2.4a) may be replaced by the weaker condition $V_0 = \text{span}\{\phi_{0n}\}$. Moreover, one may, without loss of generality, assume that the ϕ_{0n} are orthonormal (which automatically implies that the ϕ_{mn} are orthonormal for every fixed m). If the ϕ_{0n} are not orthonormal to start with, then one defines $\tilde{\phi}$ by

$$(\tilde{\phi})\hat{}(\xi) = C\hat{\phi}(\xi)\left(\sum_{k\in\mathbb{Z}} |\hat{\phi}(\xi + 2k\pi)|^2\right)^{-1/2} \tag{2.5}$$

(where we implicitly assume that $\hat{\phi}$ has sufficient decay to make the infinite sum converge). One finds that

$$\overline{\text{span}\{\phi_{0n}\}} = \overline{\text{span}\{\tilde{\phi}_{0n}\}},$$

while, moreover, the $\tilde{\phi}_{0n}$ are orthonormal. See [16] for a detailed proof.

EXAMPLE 2.1 (continued). In this case the ϕ_{0n} are orthonormal from the start. If we define

$$c_{mn}(f) = \langle \phi_{mn}, f \rangle = 2^{-m/2} \int_{2^m n}^{2^m(n+1)} dx f(x), \tag{2.6}$$

then

$$P_m f = \sum_n c_{mn}(f)\phi_{mn}.$$

Let us look at the difference between $P_m f$ and the next coarser approximation $P_{m+1} f$. One easily checks that

$$\phi_{m+1\,n} = \frac{1}{\sqrt{2}}(\phi_{m\,2n} + \phi_{m\,2n+1});$$

hence

$$c_{m+1\,n}(f) = \frac{1}{\sqrt{2}}\left[c_{m\,2n}(f) + c_{m\,2n+1}(f)\right].$$

This again exhibits $P_{m+1}f$ as an averaged version of $P_m f$, i.e., as a larger scale approximation. The difference between these two successive approximations is given by

$$P_m f - P_{m+1} f = \frac{1}{2} \sum_n \left[c_{m\,2n}(f) - c_{m\,2n+1}(f) \right] \left[\phi_{m\,2n} - \phi_{m\,2n+1} \right].$$

The remarkable fact about this expression is that it can be rewritten under a form very similar to (2.6). Define

$$\psi(x) = \phi(2x) - \phi(2x-1) = \begin{cases} 1, & 0 \leqq x < \frac{1}{2}, \\ -1, & \frac{1}{2} \leqq x < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.7}$$

Then

$$\begin{aligned} \psi_{mn}(x) &= 2^{-m/2} \psi(2^{-m}x - n) \\ &= \frac{1}{\sqrt{2}} \left(\phi_{m-1\,2n} - \phi_{m-1\,2n+1} \right), \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} Q_{m+1} f &= P_m f - P_{m+1} f \\ &= \sum_n d_{m+1\,n}(f) \psi_{m+1\,n}, \end{aligned} \tag{2.9}$$

where

$$d_{m+1\,n}(f) = \langle \psi_{m+1\,n}, f \rangle = \frac{1}{\sqrt{2}} \left[c_{m\,2n}(f) - c_{m\,2n+1}(f) \right].$$

What is so remarkable about this? Note first, as can easily be checked from (2.7), that for fixed m the ψ_{mn} are orthonormal. The decomposition (2.9) is thus the expansion, with respect to an orthonormal basis, of $Q_{m-1}f$, the orthogonal projection of f onto $W_{m+1} = P_m L^2 - P_{m+1} L^2$, i.e., onto the orthogonal complement of V_{m+1} in V_m. The surprising fact is that, as is clear from (2.9), the W_m are also (as are the V_m) generated by the translates and dilates $\cdot \psi_{mn}$ of a single function ψ. Once this is realized, building a wavelet basis becomes trivial. Clearly (2.1)–(2.2), together with $W_m \perp V_m$, $V_{m-1} = V_m \oplus W_m$, imply that the W_m are all mutually orthogonal, and that their direct sum is $L^2(\mathbb{R})$. Since, for each m, the set $\{\psi_{mn};\ n \in \mathbb{Z}\}$ constitutes an orthonormal basis for W_m, it follows that the whole collection $\{\psi_{mn};\ m, n \in \mathbb{Z}\}$ is an orthonormal wavelet basis for $L^2(\mathbb{R})$.

In the example above the function ψ is nothing but the Haar function (see (1.9)), and it is therefore no surprise that the ψ_{mn} constitute an orthonormal

basis. The example does however clearly show how this basis can be constructed from a multiresolution analysis. Let us sketch now how the general case works:

For a multiresolution analysis, i.e., a family of spaces V_m and a function ϕ satisfying (2.1)–(2.4), one defines (as in Example 2.1) W_m as the orthogonal complement, in V_{m-1}, of V_m,

$$V_{m-1} = V_m \oplus W_m, \qquad W_m \perp V_m. \tag{2.10}$$

Equivalently,

$$W_m = Q_m L^2(\mathbb{R}) \quad \text{with} \quad Q_m = P_{m-1} - P_m. \tag{2.11}$$

It follows immediately that all the W_m are scaled versions of W_0,

$$f \in W_m \Leftrightarrow f(2^m \cdot) \in W_0, \tag{2.12}$$

and that the W_m are orthogonal spaces which sum to $L^2(\mathbb{R})$,

$$L^2(\mathbb{R}) = \bigoplus_{m \in \mathbb{Z}} W_m. \tag{2.13}$$

Because of the properties (2.1)–(2.4) of the V_m, it turns out (see [16], [19]) that in W_0 also (as in V_0) there exists a vector ψ such that its integer translates span W_0, i.e.,

$$\overline{\text{span}\{\psi_{0n}\}} = W_0, \tag{2.14}$$

where, as before, $\psi_{mn}(x) = 2^{-m/2}\psi(2^{-m}x - n)$ for $m, n \in \mathbb{Z}$. It follows immediately from (2.12) that then

$$\overline{\text{span}\{\psi_{mn}\}} = W_m,$$

for all $m \in \mathbb{Z}$.

Intuitively one may understand this similarity between W_0 and V_0 by the fact that V_{-1} is "twice as large" as V_0, since V_0 is generated by the integer translates of a single function ϕ_{00}, while V_{-1} is generated by the integer translates of *two* functions, namely ϕ_{-10} and ϕ_{-11}. It therefore seems natural that the orthogonal complement W_0 of V_0 in V_{-1} is also generated by the integer translates of a single function. This hand-waving argument can easily be made rigorous by using group representation arguments. Mere proof of *existence* of a function ψ satisfying (2.14) would however not be enough for practical purposes. A more detailed analysis leads to the following algorithm for the *construction* of ψ (see [16], [19]). We start from a function ϕ such that the ϕ_{0n} are an orthonormal basis for V_0 (if

necessary, we apply (2.5)). Since $\phi \in V_0 \subset V_{-1} = \overline{\text{span}\{\phi(2\cdot - n)\}}$, there exist c_n such that

$$\phi(x) = \sum_n c_n \phi(2x - n). \tag{2.15}$$

Define then

$$\psi(x) = \sum_n (-1)^n c_{n+1} \phi(2x + n). \tag{2.16}$$

The corresponding ψ_{0n} will constitute an orthonormal basis of W_0; see [16], [19]. Consequently the ψ_{mn}, for fixed m, will constitute an orthonormal basis of W_m. It follows then from (2.12) that the $\{\psi_{mn},\ m, n \in \mathbb{Z}\}$ constitute an orthonormal basis of wavelets for $L^2(\mathbb{R})$. This completes the explicit construction, in the general case, of an orthonormal wavelet basis from a multiresolution analysis.

EXAMPLE 2.1 (final visit). As we already noted above, the ϕ_{0n} are orthonormal in this example, and

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

Applying the recipe (2.15)–(2.16) then leads to

$$\psi(x) = \phi(2x) - \phi(2x - 1),$$

which corresponds to (2.7).

Remarks. 1. One can show (see [16]) that the functions ϕ, ψ having all the above properties necessarily satisfy

$$\int dx\, \psi(x) = 0 \tag{2.17}$$

and

$$\int dx\, \phi(x) \neq 0,$$

where we implicitly assume that ϕ, ψ are sufficiently well-behaved for these integrals to make sense (in all examples of practical interest, $\phi, \psi \in L^1$). In fact, one does not even need to assume that the ϕ_{0n} or ψ_{0n} are orthonormal to derive (2.17)–(2.18). In [15] it is shown that (2.17) has to be satisfied even if the ψ_{mn} constitute only a frame (see the introduction). Note also that the transition (2.5) from ϕ to $\tilde{\phi}$, orthonormalizing the ϕ_{0n}, preserves $\int dx\, \phi(x) \neq 0$.

2. If one restricts oneself to the case where ϕ is a *real* function (as in all the examples above), then ϕ is determined uniquely, up to a sign, by the requirement

that the ϕ_{0n} be orthonormal. One then also has $\int dx\, \phi(x) = \pm 1$; we shall fix the sign of ϕ so that

$$\int dx\, \phi(x) = 1. \tag{2.18}$$

In practice, one can often start the whole construction by choosing an appropriate ϕ, i.e., a function ϕ satisfying (2.15) for some c_n. Provided ϕ is "reasonable" (it suffices, e.g., that $\inf_{|\xi| \leq \pi} |\hat{\phi}(\xi)| > 0$ and that $\Sigma_{k \in \mathbb{Z}} |\hat{\phi}(\xi + 2k\pi)|^2$ is bounded), the closed linear spans V_m of the ϕ_{mn} (m fixed) then automatically satisfy (2.1)–(2.4) and there exists an associated orthonormal basis of wavelets. Two typical examples are

EXAMPLE 2.2.

$$\phi(x) = \begin{cases} x, & 0 \leq x \leq 1, \\ 2 - x, & 1 \leq x \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

This is a linear spline function; the spaces V_m consist of continuous, piecewise linear functions. The c_n are given by

$$\phi(x) = \tfrac{1}{2}\phi(2x) + \phi(2x - 1) + \tfrac{1}{2}\phi(2x - 2).$$

EXAMPLE 2.3.

$$\phi(x) = \begin{cases} x^2, & 0 \leq x \leq 1, \\ -2x^2 + 6x - 3, & 1 \leq x \leq 2, \\ (3 - x)^2, & 2 \leq x \leq 3, \\ 0, & \text{otherwise.} \end{cases}$$

This is a quadratic spline function; the spaces V_m consist of C^1, piecewise quadratic functions. The c_n are given by

$$\phi(x) = \tfrac{1}{4}\phi(2x) + \tfrac{3}{4}\phi(2x - 1) + \tfrac{3}{4}\phi(2x - 2) + \tfrac{1}{4}\phi(2x - 3).$$

In these last two examples the corresponding ψ will be, respectively, continuous and piecewise linear, or C^1 and piecewise quadratic. Starting from spline functions one can, in fact, construct orthonormal bases of wavelets with an arbitrarily high number of continuous derivatives. These bases are the Battle-Lemarié bases (see [17], [18], [16]). In these constructions the initial function ϕ is

compactly supported, but the ϕ_{0n} are not orthogonal, as illustrated by the two examples above. One therefore has to apply (2.5) before using (2.15), (2.16); the transition $\phi \to \tilde{\phi}$ in (2.5) leads to a non-compactly supported $\tilde{\phi}$, resulting in a non-compactly supported ψ. Typically the Battle-Lemarié wavelets have exponential decay.

We shall see below that for the construction of orthonormal bases of *compactly supported* wavelets it is more natural to start from the coefficients c_n than from the function ϕ.

Up to now, we have restricted ourselves to one dimension. It is very easy, however, to extend the multiresolution analysis to more dimensions. As pointed out by R. Coifman and Y. Meyer [22], this extension was already inherent in the first construction by P. G. Lemarié and Y. Meyer [23] of an n-dimensional wavelet basis. It becomes much more transparent, however, from the multiresolution analysis point of view. Let us illustrate this for e.g. two dimensions. The case of n dimensions, n arbitrary, is completely similar. Assume that we dispose of a one-dimensional multiresolution analysis, i.e., we have at hand a ladder of spaces V_m, and functions ϕ, ψ satisfying (2.1)–(2.4) and (2.14), where the ϕ_{0n} and the ψ_{0n} are assumed to be orthonormal. Define

$$\mathbf{V}_m = V_m \oplus V_m.$$

Clearly, the $\mathbf{V}_m$ define a ladder of subspaces of $L^2(\mathbb{R}^2)$, satisfying (2.1) and the equivalent, for $\mathbb{R}^2$, of (2.2). Moreover, (2.3) holds, and if we define

$$\Phi(x_1, x_2) = \phi(x_1)\phi(x_2),$$

then this two-dimensional function satisfies the analogue of (2.4),

$$\mathbf{V}_m = \overline{\text{linear span } \{\Phi_{mn};\ n \in \mathbf{Z}^2\}},$$

where Φ_{mn} is defined by

$$\begin{aligned}\Phi_{mn}(x_1, x_2) &= 2^{-m}\Phi(2^{-m}x_1 - n_1, 2^{-m}x_2 - n_2)\\ &= \phi_{mn_1}(x_1)\phi_{mn_2}(x_2).\end{aligned}$$

Note that we use the *same* dilation for both arguments. Because of the definition (2.10) of W_m, we find immediately that

$$\mathbf{V}_{m-1} = \mathbf{V}_m \oplus [(V_m \oplus W_m) \oplus (W_m \oplus V_m) \oplus (W_m \oplus W_m)].$$

This implies that an orthonormal basis for the orthogonal complement $\mathbf{W}_m$ of $\mathbf{V}_m$ in $\mathbf{V}_{m-1}$ is given by the functions $\phi_{mn_1}\psi_{mn_2}$, $\psi_{mn_1}\phi_{mn_2}$, $\psi_{mn_1}\psi_{mn_1}$, with $n_1, n_2 \in \mathbf{Z}$, or equivalently, by the two-dimensional wavelets Ψ^l_{mn},

$$\Psi^l_{mn}(x_1, x_2) = 2^{-m}\Psi^l(2^{-m}x_1 - n_1, 2^{-m}x_2 - n_2), \tag{2.19}$$

where $l = 1, 2, 3$, $n \in \mathbb{Z}^2$, and

$$\Psi^1(x_1, x_2) = \phi(x_1)\psi(x_2), \tag{2.20}$$

$$\Psi^2(x_1, x_2) = \psi(x_1)\phi(x_2), \tag{2.21}$$

$$\Psi^3(x_1, x_2) = \psi(x_1)\psi(x_2). \tag{2.22}$$

It follows that the Ψ^l_{mn}, $l = 1, 2, 3$, $m \in \mathbb{Z}$, $n \in \mathbb{Z}^2$, constitute an orthonormal basis of wavelets for $L^2(\mathbb{R}^2)$.

The above construction shows how any multiresolution analysis + associated wavelet basis in one dimension can be extended to d dimensions. The decomposition + reconstruction algorithm constructed by S. Mallat for visual data in [8] uses such a two-dimensional basis.

2.B. The Laplacian pyramid scheme of P. Burt and E. Adelson. In this subsection we review some aspects, relevant for the present paper, of Burt and Adelson's algorithm for the decomposition and reconstruction of images. For a more detailed presentation, and for applications, we refer the reader to [20].

One of the goals of a decomposition scheme for images is to remove the very high correlations existing between neighboring pixels, in order to achieve data compression. Several different schemes have been proposed to achieve this. Typically they use a prediction method, in which the value at a pixel is predicted (by a weighted average) from either previously encoded or neighboring pixels, and only the difference between the actual pixel value and the predicted value is encoded. Using the neighboring pixels for prediction is more natural and should lead to more accurate prediction (and therefore to greater data compression), but is much harder to implement than the easy causal prediction scheme, using only previously encoded pixels. The scheme proposed by Burt and Adelson combines the ease of computation of a causal scheme with the advantages and elegance of a neighborhood-based (noncausal) scheme. The result is—we quote directly from [20a]—

> "$\cdots$ a technique for image encoding in which local operators of many scales but identical shape serve as the basis functions."

The analogy with multiresolution analysis is evident from this quote.

Images are two-dimensional, and the Laplacian pyramid scheme is a two-dimensional algorithm. For simplicity, the review below will be restricted to one dimension. This does not really matter, except in details (which will be pointed out). Moreover, the two-dimensional schemes used in [20] are in fact obtained (for simplicity reasons) as a tensor product of two one-dimensional schemes.

Our presentation will be already adapted to later use in this paper, and slightly different in notation from [20]. Except for this difference, what follows is the construction in [20].

The original (one-dimensional) data can be represented as a sequence of real numbers, $(c_n)_{n\in\mathbf{Z}}$, representing the pixel values. For later convenience, we give this sequence the index 0,

$$c_n^0 = c_n.$$

The main idea is to decompose c^0 into different sequences corresponding to distinct ranges of spatial frequency. The highest level, with only the high frequency content of c^0, is obtained by computing the difference between c^0 and a blurred version $\tilde{c}^0$. The remainder, i.e., the blurred version, contains only lower spatial frequencies, and can therefore be sampled more sparsely than c^0 itself, without loss of information. The Laplacian pyramid algorithm provides an elegant and easily implementable scheme for doing all this. The whole process is repeated several times in order to achieve the desired decomposition.

One starts by transforming the sequence c^0 into a sequence c^1 by means of an operator which both averages and decimates,

$$c_k^1 = \sum_n w(n - 2k)c_n^0. \tag{2.23}$$

The weighing coefficients $w(n)$ are always real; in [20] they are chosen to be symmetric and normalized, i.e.,

$$\begin{aligned} &w(n) = w(-n), \\ &\sum_n w(n) = 1. \end{aligned} \tag{2.24}$$

They are also required to satisfy an "equal contribution constraint", stipulating that the sum of all the contributions from a given node n is independent of n, i.e., all the nodes contribute the same total amount,

$$\sum_k w(n - 2k) \text{ is independent of } n.$$

This can be rewritten as

$$\sum_n w(2n) = \sum_n w(2n + 1). \tag{2.25}$$

We shall come back below to the mathematical significance of this requirement. Examples given in [20] are

$$\begin{aligned} &w(n) = 0 \quad \text{if} \quad |n| > 2, \\ &w(2) = w(-2) = \tfrac{1}{4} - \tfrac{1}{2}a, \\ &w(1) = w(-1) = \tfrac{1}{4}, \\ &w(0) = a; \end{aligned} \tag{2.26}$$

the different values considered for a are $a = .6, .5, .4$ and $.3$.

The sequence c^1 plays a double role: it will serve as the input sequence (instead of c^0) for the next level of the pyramid, and it is also an intermediate step for the computation of the blurred version $\tilde{c}^0$ to be subtracted from c^0. (Note that we cannot have c^1 itself as this blurred version: c^1 and c^0 "live" on different scales—see Figure 1). More precisely,

$$\tilde{c}_n^0 = \sum_k w(n-2k)c_k^1, \tag{2.27}$$

or

$$\tilde{c}_n^0 = \sum_l \tilde{w}(n,l)c_l^0,$$

where

$$\tilde{w}(n,l) = \sum_k w(n-2k)w(l-2k). \tag{2.28}$$

Notice that this does not quite define a convolution; from (2.28) one sees that $\tilde{w}(n,l) = \tilde{w}(n-2, l-2)$, but $\tilde{w}(n,l) \neq \tilde{w}(n-1, l-1)$ in general. The sequence $\tilde{c}^0$ is clearly a blurred version of c^0; one then defines the difference d^0 by

$$d_n^0 = c_n^0 - \tilde{c}_n^0. \tag{2.29}$$

Knowing this difference sequence (the high spatial frequency content of c^0) and c^1 (a low-pass filtered version of c^0, sampled at a sparser rate) is clearly sufficient to reconstruct the data c^0, since

$$c_n^0 = d_n^0 + \sum_k w(n-2k)c_k^1.$$

The whole process is then iterated. From c^1 one computes c^2 and $\tilde{c}^1$, d^1 is the difference $c^1 - \tilde{c}^1$, etc. A graphical representation of the transitions $c^0 \to c^1 \to c^2 \to \cdots$ and $c^1 \to \tilde{c}^0$ is given in Figures 1a and 1b.

A more condensed notation for all the above is the following. Define the operator F: $l^2(\mathbf{Z}) \to l^2(\mathbf{Z})$ (F for "filter") by

$$(Fa)_k = \sum_n w(n-2k)a_k. \tag{2.30}$$

Then (2.23), (2.27) and (2.29) become

$$c^1 = Fc^0, \tag{2.31}$$

$$\tilde{c}^0 = F^*c^1 = F^*Fc^0, \tag{2.32}$$

$$d^0 = c^0 - \tilde{c}^0 = (\mathbf{1} - F^*F)c^0. \tag{2.33}$$

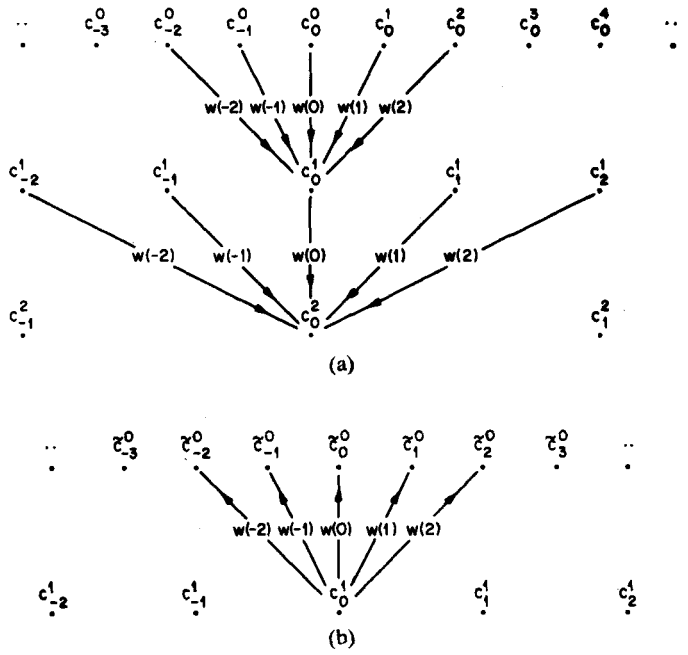


Figure 1. Graphical representation of the Laplacian pyramid scheme (redrawn from [20]).

a. The transition $c^0 \to c^1 \to c^2$. For simplicity's sake we have restricted ourselves to the case $w(n) = 0$ if $|n| > 2$, and only the computation of c_0^1 and c_0^2 are depicted.

b. The transition $c^1 \to \tilde{c}^0$.

Here we use the standard notation F^* for the adjoint of the (bounded) operator F. Note that we implicitly assume that $c^0 \in l^2(\mathbb{Z})$, or, in signal analysis terms, that the data sequence c^0 has finite energy. In practice, the sequence c^0 is finite, and this constraint does not matter.

The total decomposition consists thus in L consecutive steps (in practice $L = 5$ or 6), with

$$c^l = Fc^{l-1}, \qquad l = 1, \cdots, L,$$
$$d^{l-1} = c^{l-1} - F^* c^l = (\mathbf{1} - F^*F)c^{l-1}. \tag{2.34}$$

From the sequences $d^0, \cdots, d^{L-1}, c^L$ one then reconstructs c^0 recursively by

$$c^{l-1} = d^{l-1} + F^*c^l. \tag{2.35}$$

At every step, in the decomposition (2.34) as well as in the reconstruction (2.35) the same filter coefficients are used, and all the operations involved are direct and linear (no solving of complicated systems of equations!). This makes

this algorithm very easy to implement. The decimation aspect in the operator F, which reduces the number of entries in the c^l by a factor 2 at every step, makes the whole decomposition algorithm as fast as a fast Fourier transform (see [20]).

Let N be the total number of non-zero entries in c^0. Then the total number of entries in $d^0, \cdots, d^{L-1}, c^L$ (except for edge effects) is

$$N + N/2 + \cdots + N/2^{L-1} + N/2^L = 2N(1 - 2^{-L-1}).$$

After the Laplacian pyramid decomposition there is thus a larger number of entries (almost twice as many) than in the original data sequence. However, it turns out that, because of the removal of correlations, the decomposed data can be greatly compressed (see [20a]). The net effect is still an appreciable data compression. We shall not go into this here, however. Note that the increase of the number of entries is less pronounced in two dimensions (a factor $\frac{4}{3}$ instead of 2).

The similarity between the Laplacian pyramid algorithm and a multiresolution analysis is now clear. In both approaches, the data (a function in multiresolution analysis, a sequence in the Laplacian pyramid) are decomposed into a "pyramid" of approximations, corresponding to less and less detail. Moreover, the differences between each two successive approximations are computed (corresponding to the wavelet decomposition in the multiresolution analysis). However, it is also clear that the schemes are quite different in the details of the computation of the decomposition. The algorithm developed by S. Mallat, described in the next subsection, retains the attractive features of the Laplacian pyramid scheme, but is much closer to the analysis described in subsection 2A.

The filter coefficients $w(n)$, or equivalently the filter operator F, are associated in [20] with "equivalent weighting functions". Only the limit of these functions will be relevant for us; we conclude this subsection by its definition and a few of its properties. One may wonder which kind of input sequence c^0 corresponds to the "simplest" decomposition sequence, i.e., to $d^0 = \cdots = d^{L-1} = 0$, and $(c^L)_n = \delta_{n0}$. The answer is obviously (use the reconstruction algorithm)

$$c^0 = (F^*)^L e, \tag{2.36}$$

where e is the sequence $e_n = \delta_{n0}$. If, e.g., $L = 1$, then the entries of c^0 are exactly the $w(n)$. Since any sequence can be considered as a sum of translated versions of e, the sequence $c^0 = (F^*)^L e$ gives the basic building block for the subspace $(F^*)^L l^2(\mathbb{Z})$, i.e., for the L-th level component sequences. It is therefore important that these sequences c^0 do not look messy, which they well might, for L large enough (for a "messy" example, see Figure 4 in subsection 3B). One can make a graphical representation of the c^0 defined by (2.36), for successive L. We represent the sequence e by a simple histogram, with value 1 for $-\frac{1}{2} \leqq x < \frac{1}{2}$, 0 otherwise (see Figure 2). The sequence F^*e "lives" on a scale twice as small, and will therefore be represented by a histogram with step widths $\frac{1}{2}$ (as opposed to 1 for e); its different amplitudes are given by $2(F^*e)_n = 2w(n)$. Similarly, $(F^*)^l e$ is

represented by a histogram with step width 2^{-l}; the successive amplitudes are given by $2^l((F^*)^l e)_n$. We have introduced an extra factor 2 at every step in our representation for normalization purposes: the area under each histogram is always 1. This normalization will be convenient in Section 3. The example plotted in Figure 2 corresponds to the $w(n)$ given by (2.26), with $a = .375$. Plots for $a = .6, .5, .4$ and $.3$, with slightly different conventions, can be found in [20]; our choice $a = .375$ shall become clear below. One finds a very rapid convergence of these histograms to a rather nice function. This surprising feature is in fact due, in large part, to the special form (2.26) of the coefficients $w(n)$, and in particular, to condition (2.25). The following argument shows why. The "representation" of e in Figure 2 is the characteristic function of the interval $[-\frac{1}{2}, \frac{1}{2}[$, which we

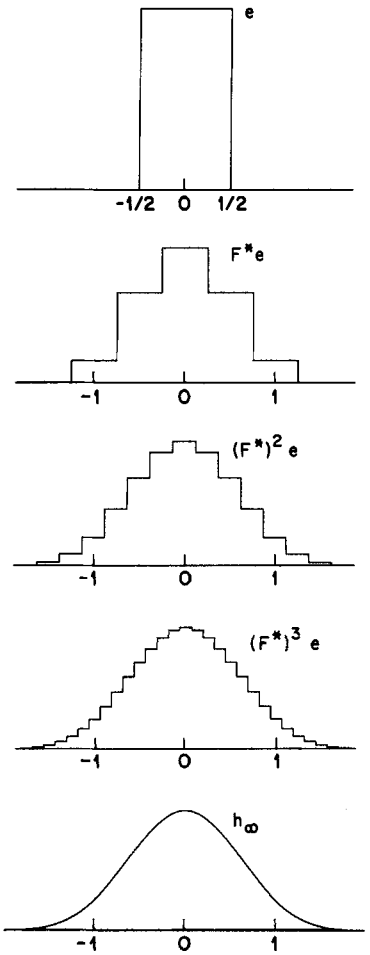


Figure 2. The successive sequences e, F^*e, $(F^*)^2e$, $(F^*)^3e$ represented by histograms, and the limit function h_∞ (see text). We have taken the $w(m)$ as defined by (2.26), with $a = .375$.

denote by h_0. The representations h_1, h_2 of, respectively, $F^*e, (F^*)^2e$ are given by

$$h_1(x) = 2\sum_n w(n)\chi_{[-1/4,1/4[}(x - \tfrac{1}{2}n) \tag{2.37}$$

and

$$h_2(x) = 4\sum_n w(n)\left[\sum_m w(m)\chi_{[-1/8,1/8[}(x - \tfrac{1}{2}n - \tfrac{1}{4}m)\right]. \tag{2.38}$$

To make the transition from h_{j-1} to h_j one

(i) divides h_{j-1}, a step function with step width $2^{-(j-1)}$, into its components

$$h_{j-1} = \sum_k a_{j-1,k}\chi_{[2^{-(j-1)}(k-1/2),2^{-(j-1)}(k+1/2)[} \tag{2.39}$$

(see Figure 3c),

(ii) replaces every component by a suitably scaled and recentered version of h_1,

$$\chi_{[2^{-j+1}(k-1/2),2^{-j+1}(k+1/2)[} \to h_1(2^{j-1}x - k) = 2\sum_n w(n)\chi_{[-1/4,1/4[}(2^{j-1}x - k - \tfrac{1}{2}n),$$

(see Figure 3d),

(iii) sums it all up,

$$h_j(x) = 2\sum_k a_{j-1,k}\sum_n w(n)\chi_{[2^{-j}(2k+n-1/2),2^{-j}(2k+n+1/2)[}$$

(see Figure 3e).

These different steps are illustrated by Figure 3. The construction amounts to defining

$$h_j = \tilde{T}_j h_{j-1} \quad \text{or} \quad h_j = \tilde{T}_j\tilde{T}_{j-1}\cdots\tilde{T}_1 h_0, \tag{2.40}$$

where

$$(\tilde{T}_l f)(x) = 2\sum_k\sum_n w(n)(\chi_{[2^{-l+1}(k-1/2),2^{-l+1}(k+1/2)[}f)(2x - 2^{-l+1}(k+n)). \tag{2.41}$$

The iterative algorithm (2.40), (2.41) is extremely easy to implement numerically.

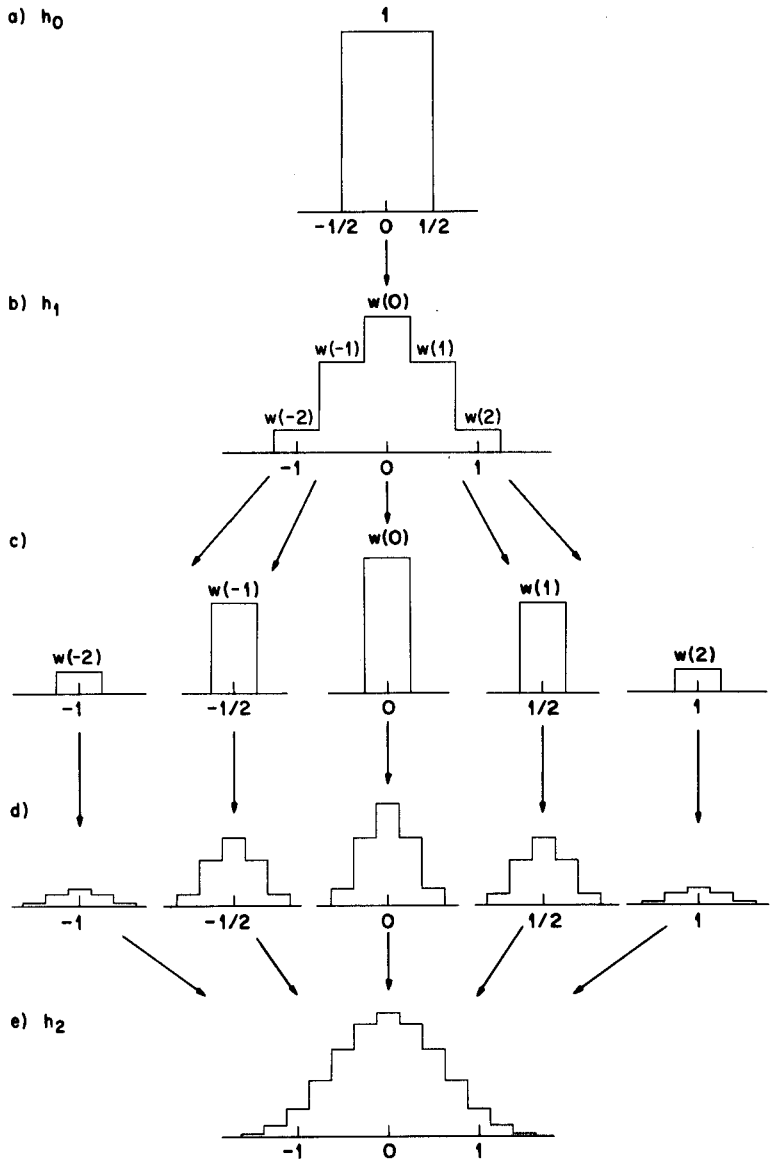


Figure 3. a) $h_0(x) = \chi_{[-1/2,1/2[}(x)$;

b) $h_1(x) = 2\Sigma w(n)\chi_{[n/2-1/4,\, n/2+1/4]}(x)$,

c) h_1 is decomposed into its "components"; each component is a multiple of the characteristic function of an interval of length $\frac{1}{2}$. (The "components" of h_j would have width 2^{-j}).

d) Each "component" is replaced by a proportional version of h_1, centered around the same point as the component, and scaled down by a factor $\frac{1}{2}$. (This scaling factor would be 2^{-j} for h_j).

e) The functions in d) are added to constitute h_2 (or h_{j+1}, if one starts from h_j in c)).

The h_j can, however, also be written differently. Let us go back to expressions (2.36), (2.37) for h_1, h_2. These can be rewritten as

$$h_1(x) = 2\sum_n w(n) h_0(2x - n), \tag{2.42}$$

$$\begin{aligned} h_2(x) &= 4\sum_n w(n) \sum_m w(m) \chi_{[-1/4,1/4[}(2x - n - \tfrac{1}{2}m) \\ &= 2\sum_n w(n) h_1(2x - n). \end{aligned} \tag{2.43}$$

This suggests

$$h_j = Th_{j-1} = \cdots = T^j h_0, \tag{2.44}$$

where

$$(Tf)(x) = 2\sum_n w(n) f(2x - n). \tag{2.45}$$

The following argument shows that (2.44) is indeed true:

$$\begin{aligned} &\big(T\tilde{T}_l f\big)(x) \\ &\quad = 4 \sum_{m,n} w(m) w(n) \\ &\qquad \cdot \sum_k \big(\chi_{[2^{-l+1}(k-1/2),\, 2^{-l+1}(k+1/2)[} f\big)\big(4x - 2m - 2^{-l+1}(n + k)\big), \end{aligned}$$

$$\begin{aligned} &\big(\tilde{T}_{l+1} T f\big)(x) \\ &\quad = 4 \sum_{m,n} w(m) w(n) \sum_k \chi_{[2^{-l}(k-1/2),\, 2^{-l}(k+1/2)[}\big(2x - 2^{-l}(n + k)\big) \\ &\qquad \cdot f\big(4x - m - 2^{-l+1}(n + k)\big). \end{aligned}$$

Substituting $k = k' + 2^{l-1}m$ into this last sum, we find

$$\begin{aligned} &\big(\tilde{T}_{l+1} T f\big)(x) \\ &\quad = 4 \sum_{m,n} w(m) w(n) \sum_{k'} \Big[\chi_{[2^{-l}(k'-1/2),\, 2^{-l}(k'+1/2)[}\big(2x - 2^{-l}(n + k') - m\big) \\ &\qquad\qquad \cdot f\big(4x - 2m - 2^{-l+1}(n + k')\big)\Big] \\ &\quad = \big(T\tilde{T}_l f\big)(x). \end{aligned}$$

Since (see (2.42), (2.43)) $h_1 = Th_0$, $h_2 = T^2h_0$, it follows that

$$\begin{aligned} h_j &= \tilde{T}_j \cdots \tilde{T}_3 h_2 = \tilde{T}_j \cdots \tilde{T}_3 T h_1 \\ &= \tilde{T}_j \cdots \tilde{T}_3 T \tilde{T}_1 h_0 = \tilde{T}_j \cdots \tilde{T}_4 T \tilde{T}_2 \tilde{T}_1 h_0 \\ &= \cdots = T\tilde{T}_{j-1} \cdots \tilde{T}_1 h_0 = T h_{j-1}, \end{aligned}$$

which proves (2.44).

We have thus two different ways, (2.44) and (2.40), to compute the h_j. Figure 2 shows that, at least for some choices of the $w(n)$, the functions h_j converge, for $j \to \infty$, to a "nice" function h_∞. The explicit proofs which will be given in subsection 3B show that, at least for the examples (2.26) with $.125 < a < .625$, the function h_∞ is continuous (see (2.46) below), has compact support, and that the convergence $h_j \to h_\infty$ is uniform. Let us just accept these facts for the moment.

The two formulas (2.40) and (2.44) are both extremely useful in the study (and the proof) of this convergence. The construction of $h_\infty = \lim_{j\to\infty} h_j$ via (2.40) has the following nice localization feature. To compute the value of $h_{j+1}(x)$ the recursion $h_{j+1} = \tilde{T}_j h_j$ uses only values $h_j(y)$ for $|y - x| \leqq 2^{-j-1}(N + 1)$, where we assume $w(n) = 0$ for $n \geqq 2N$. Consequently, the value of $h_\infty(x)$ can be computed using only the values of $h_j(y)$ for $|y - x| \leqq 2^{-j}(N + 1)$. For increasing j, this lends a "zoom-in" quality to the graphical construction of which Figure 2 is an example. This is extremely useful when one wants to focus on details of the behavior of h_∞ (see, e.g. Figure 6 in subsection 4B). This localization feature is not present in (2.44). The formula $h_j(x) = (Th_{j-1})(x)$ uses values of h_{j-1} at points which stay at fixed distance of each other (i.e., $2x$, $2(x \pm \frac{1}{2})$, $2(x \pm 1), \cdots$), independently of how large j is. The usefulness of (2.44) is therefore not "graphical". It is, however, this less local formula which will be most useful in proving convergence of the h_j, continuity of h_∞, etc.

Introducing Fourier transforms, (2.45) can be rewritten as

$$(Tf)\hat{}\,(\xi) = W(\tfrac{1}{2}\xi)\hat{f}(\tfrac{1}{2}\xi),$$

where

$$W(\xi) = \sum_n w(n) e^{in\xi}.$$

Consequently, from (2.45), one obtains

$$\hat{h}_l(\xi) = (2\pi)^{-1/2}\left[\prod_{j=1}^{l} W(2^{-j}\xi)\right]\frac{\sin(2^{-l-1}\xi)}{2^{-l-1}\xi}.$$

For $l \to \infty$, this converges, pointwise, to

$$\hat{h}_\infty(\xi) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} W(2^{-j}\xi),$$

provided this infinite product makes sense. (We shall come back to this, and other convergence problems, in subsection 3B. It turns out that, for $w(n)$ as chosen in (2.26), the convergence $h_l \to h_\infty$ holds in all L^p-spaces, $1 \leqq p \leqq \infty$.) Because of the constraint (2.25), one finds that $W(\xi)$ is divisible by $(1 + e^{i\xi})$,

$$\begin{aligned} W(\xi) &= \tfrac{1}{2}(1 + e^{i\xi})Q(\xi) \\ &= e^{i\xi/2}\cos\tfrac{1}{2}\xi Q(\xi). \end{aligned}$$

Combining this with

$$\prod_{j=1}^{\infty} \cos(2^{-j}\xi) = \frac{\sin\xi}{\xi},$$

we find

$$\hat{h}_\infty(\xi) = (2\pi)^{-1/2} e^{i\xi/2} \frac{2\sin\frac{1}{2}\xi}{\xi} \prod_{j=1}^{\infty} Q(2^{-j}\xi).$$

The constraint (2.25) leads thus to a factor ξ^{-1} in $\hat{h}_\infty$, i.e., to some regularity in h_∞! Without this constraint, as can be easily checked, the graphical procedure in Figure 2 can lead to rather horrible (fractal) functions h_∞ (see e.g. Figure 4). In fact, for the examples (2.26) one even finds two factors $\cos\frac{1}{2}\xi$,

$$W(\xi) = (\cos\tfrac{1}{2}\xi)^2\left[(8a - 3) + 4(1 - 2a)(\cos\tfrac{1}{2}\xi)^2\right].$$

Using an estimation technique due to P. Tchamitchian (see Lemma 3.2 below) one finds that this leads to

$$|\hat{h}_\infty(\xi)| \leqq C(1 + |\xi|)^{-2 + \log_2[\max(1, |8a-3|)]}. \tag{2.46}$$

For $.125 < a < .625$, which includes all the choices in [20], this implies that h_∞ is continuous. For $a = \frac{3}{8} = .375$ (the example chosen in Figure 2), the decay of $\hat{h}_\infty$ is even stronger,

$$W(\xi) = (\cos\tfrac{1}{2}\xi)^4,$$

$$\hat{h}_\infty(\xi) = (2\pi)^{-1/2}\left(\frac{\sin\frac{1}{2}\xi}{\frac{1}{2}\xi}\right)^4.$$

In this case, h_∞ is thus a fourth-order convolution of $\chi_{[0,1[}$ with itself, which results in a $C^{3-\varepsilon}$ function.

The above remarks show that constraints on the $w(n)$, corresponding to divisibility of $W(\xi)$ by $(1 + e^{i\xi})$, result in regularity of h_∞. Constructions similar to (2.44) will be used in Section 3, where the above "trick" for imposing regularity on h_∞ will turn up again.

This concludes our review of the Laplacian pyramid scheme. The above is by no means a complete review; only those aspects relevant to the present paper have been highlighted. For more details, and especially for applications (data compression, image splining) the reader should consult [20a] and [20b].

Remark. During the last revision of this paper before publication, Y. Meyer drew my attention to related work by G. Deslauriers and S. Dubuc [29]. They are interested in functions defined recursively by the following interpolation scheme. At the l-th step, the values of f at the points $2^{-l}(2k + 1)$, $k \in \mathbf{Z}$, are computed from the $f(k2^{-l+1})$ via the formula

$$f\big((2k + 1)2^{-l}\big) = \sum_{m \in \mathbf{Z}} a_m f\big((k - m)2^{-l+1}\big).$$

In many applications considered by Deslauriers and Dubuc, the interpolation procedure is symmetric, i.e., $a_{-m} = a_{m+1}$ for all $m \in \mathbf{Z}$. For suitable choices of the a_m, the functions f constructed via this dyadic interpolation scheme, starting from the $f(k)$, $k \in \mathbf{Z}$, are continuous, and are therefore completely characterized by their values at the dyadic rational points $x = k2^{-l}$, $k \in \mathbf{Z}$, $l \in \mathbf{N}$. A typical function f can be written as

$$f(x) = \sum_{k \in \mathbf{Z}} f(k)g(x - k),$$

where g is the function obtained by interpolation from $g(0) = 1$, $g(k) = 0$ for $k \in \mathbf{Z} \setminus \{0\}$. The definition of g via the dyadic interpolation scheme is exactly the same as our "graphical recursion" (2.40), with the choice $w(0) = \frac{1}{2}$, $w(2n) = 0$ for $n \neq 0$, $w(2n + 1) = \frac{1}{2}a_{n+1}$, $n \in \mathbf{Z}$. The analysis of the properties of g in [29] is then carried out by means of the same correspondence between "graphical recursion" and the iterative formula (2.44). Imposing $\Sigma a_m = 1$ (i.e., $\Sigma w(n) = 1$) immediately leads to $w(\xi) = \big[\frac{1}{2}(1 + e^{i\xi})\big]^2 Q(\xi)$, which is then exploited, in [29], to impose continuity on g. There is therefore a clear similarity between the techniques used here and those exposed in [29]. The applications are different, however. Moreover, the proofs given in Section 3 apply to more general cases than those in [29], since we do not impose $w(2n) = 0$ for $n \neq 0$, nor $w(2n + 1) = w(-2n - 1)$.

2.C. The wavelet based decomposition and reconstruction algorithm of S. Mallat. In [8], Stéphane Mallat exploits the attractive features of multiresolution analysis to construct a decomposition and reconstruction algorithm for

2-*d*-images that has the same philosophy as the Laplacian pyramid scheme, but is more efficient and orientation selective. It is interesting to remark that the development of the concept of multiresolution analysis was triggered by the multiresolution methods, and in particular by the Laplacian pyramid. The full mathematical study of the concept, by S. Mallat and Y. Meyer, was done more or less simultaneously with the practical development, by S. Mallat, of his algorithm for vision analysis and reconstruction. This is not the only instance in which theoretical developments concerning wavelets find their inspiration in applications: the last few years have seen a constant feedback between theory and applications. In fact, this paper is another such instance.

Let us start by a review of the algorithm in one dimension. As in the previous subsection, we want to decompose a sequence $c^0 = (c_n^0)_n \in l^2(\mathbb{Z})$ into levels corresponding to different spatial frequency bands. To achieve this, we shall use a multiresolution analysis, which can be chosen freely (as long as (2.1)–(2.4) are satisfied), but has to be kept fixed for the whole algorithm. We suppose thus that we have chosen spaces V_m and a function ϕ such that (2.1)–(2.4) are satisfied. We assume (if necessary, we apply (2.5) first) that the ϕ_{0n} are orthonormal. Let $\{\psi_{mn};\ m, n \in \mathbb{Z}\}$ be the associated orthonormal wavelet basis (we shall keep the same notations as in subsection 2A). The multiresolution analysis and orthonormal basis chosen in [8] is one in which the V_m consist of cubic spline functions (cf. Examples 2.2 and 2.3, corresponding to linear and quadratic splines, respectively); the corresponding orthonormal basis is one of the Battle-Lemarié bases. In what follows we shall assume that both ϕ and ψ are real, as they are in [8] and indeed in most practical examples.

Form the data sequence $c^0 \in l^2(\mathbb{Z})$ we construct a function f,

$$f = \sum_n c_n^0 \phi_{0n},$$

or

$$f(x) = \sum_n c_n^0 \phi(x - n).$$

This function is clearly an element of V_0. We can now use the whole multiresolution analysis apparatus on this function. We shall compute the successive $P_j f$, corresponding to more and more "blurred" versions of f (and hence of the data sequence c^0), and also the $Q_j f$, corresponding to the difference in information between the "versions" of f at two successive resolution levels. Eventually, of course, this has to be translated back to a "sequence" (as opposed to a "function") language, but this turns out to be very easy.

As element of $V_0 = V_1 \oplus W_1$, f can be decomposed into its components along V_1 and W_1,

$$f = P_1 f + Q_1 f.$$

Each of these components can be expanded with respect to the orthonormal bases ϕ_{1n}, ψ_{1n}, respectively,

$$P_1 f = \sum_k c_k^1 \phi_{1k},$$

$$Q_1 f = \sum_k d_k^1 \psi_{1k}.$$

The sequence c^1 represents a smoothed version of the original data sequence c^0, while d^1 represents the difference in information between c^0 and c^1 (cf. the discussion of P_j, Q_j in subsection 2A). The sequences c^1, d^1 can be computed as a function of c^0 in the following way. Since the ϕ_{1n} are orthonormal bases of V_1, one has

$$\begin{aligned} c_k^1 &= \langle \phi_{1k}, P_1 f \rangle = \langle \phi_{1k}, f \rangle \\ &= \sum c_n^0 \langle \phi_{1k}, \phi_{0n} \rangle, \end{aligned}$$

where

$$\begin{aligned} \langle \phi_{1k}, \phi_{0n} \rangle &= 2^{-1/2} \int dx\, \phi\left(\tfrac{1}{2}x - k\right) \phi(x - n) \\ &= 2^{-1/2} \int dx\, \phi\left(\tfrac{1}{2}x\right) \phi(x - (n - 2k)). \end{aligned}$$

This can be rewritten as

$$c_k^1 = \sum_n h(n - 2k) c_n^0 \tag{2.47}$$

with

$$h(n) = 2^{-1/2} \int dx\, \phi\left(\tfrac{1}{2}x\right) \phi(x - n).$$

Note that these $h(n)$ are, up to a normalization factor $2^{-1/2}$, exactly the coefficients $c(n)$ appearing in (2.15). Similarly,

$$d_k^1 = \sum_n g(n - 2k) c_n^0 \tag{2.48}$$

with

$$g(n) = 2^{-1/2} \int dx\, \psi\left(\tfrac{1}{2}x\right) \phi(x - n).$$

It follows that the expressions for c^1, d^1 as a function of c^0 are of *exactly* the same type as (2.23) in the Laplacian pyramid scheme. The main difference between the two schemes is that *both* the blurred, lower resolution c^1 and the "difference" sequence d^1 are now obtained via a filter of type (2.23). The filter coefficients $h(n)$, $g(n)$ are fixed by the chosen multiresolution analysis framework. It turns out that the $h(n)$ have many properties in common with the $w(n)$ in subsection 2B; for instance, the $h(n)$ satisfy a normalization condition, i.e., $\Sigma_n h(n) = \sqrt{2}$ (see subsection 3A for an explanation of the difference in normalization with the $w(n)$). The requirement $\Sigma_n h(2n) = \Sigma_n h(2n+1)$ is also satisfied by most interesting examples, and in particular in [8] (we shall come back to this later). The filter coefficients $g(n)$ are of a different nature, as one would expect; in particular, $\Sigma_n g(n) = 0$.

Introducing a shorthand notation similar to (2.30), we rewrite (2.47), (2.48) as

$$c^1 = Hc^0,$$

$$d^1 = Gc^0,$$

where H, G are bounded operators from $l^2(\mathbb{Z})$ to itself,

$$\begin{aligned} (Ha)_k &= \sum_n h(n-2k)a_n, \\ (Ga)_k &= \sum_n g(n-2k)a_n. \end{aligned} \tag{2.49}$$

The procedure can now be iterated; since $P_1 f \in V_1 = V_2 \oplus W_2$, we have

$$P_1 f = P_2 f + Q_2 f,$$

$$P_2 f = \sum_k c_k^2 \phi_{2k},$$

$$Q_2 f = \sum_k d_k^2 \psi_{2k}.$$

One finds then

$$\begin{aligned} c_k^2 &= \langle \phi_{2k}, P_{2f} \rangle = \langle \phi_{2k}, P_1 f \rangle \\ &= \sum_n c_n^1 \langle \phi_{2k}, \phi_{1n} \rangle. \end{aligned}$$

It is very easy to check, however, that

$$\langle \phi_{j+1\,k}, \phi_{jn} \rangle = h(n-2k),$$

independently of j. It follows that

$$c_k^2 = \sum_n h(n - 2k)c_n^1$$

or

$$c^2 = Hc^1.$$

Similarly,

$$d^2 = Gc^1.$$

Clearly this can now be iterated as many times as wanted. At every step one finds

$$P_{j-1}f = P_jf + Q_jf$$

$$= \sum_k c_k^j \phi_{jk} + \sum_k d_k^j \psi_{jk}$$

with

$$c^j = Hc^{j-1}, \tag{2.50}$$

$$d^j = Gc^{j-1}. \tag{2.51}$$

This is the desired decomposition. The successive c^j are lower and lower resolution versions of the original c^0, each sampled twice as sparsely as their predecessor (due to the factor 2 in the filter coefficients in (2.47)), and the d^j contain the difference in information between c^{j-1} and c^j. Moreover, the c^j, d^j are computed via a tree algorithm (2.50), (2.51). This computation is therefore as easy to implement as the Laplacian pyramid scheme.

Note that Mallat's algorithm is more economical than the Laplacian pyramid scheme. In practice, one will again stop the decomposition after a finite number L of steps, i.e., c^0 will be decomposed into $d^1, \cdots, d^L$ and c^L. If c^0 has initially N non-zero entries, then (neglecting edge effects) the total number of non-zero entries in the decomposition is $N/2 + N/4 + \cdots + N/2^{L-1} + N/2^L + N/2^L = N$. This shows that, unlike the Laplacian pyramid scheme (see subsection 2B), Mallat's algorithm preserves, at every step, the number of non-zero entries (as was to be expected from an algorithm based on an orthonormal basis decomposition).

So far we have only described the decomposition part of the algorithm. The reconstruction part is just as easy. Suppose we know c^j and d^j. Then

$$P_{j-1}f = P_jf + Q_jf$$

$$= \sum_k c_k^j \phi_{jk} + \sum_k d_k^j \psi_{jk},$$

and hence

$$\begin{aligned} c_n^{j-1} &= \langle \phi_{j-1\,n}, P_{j-1} f \rangle \\ &= \sum_k c_k^j \langle \phi_{j-1\,n}, \phi_{jk} \rangle + \sum d_k^j \langle \phi_{j-1\,n}, \psi_{jk} \rangle \\ &= \sum_k h(n-2k) c_k^j + \sum_k g(m-2k) d_k^j, \end{aligned}$$

or

$$c^{j-1} = H^* c^j + G^* d^j. \tag{2.52}$$

The reconstruction algorithm is therefore also a tree algorithm, using the same filter coefficients as the decomposition.

Remark. In fact, the transition $c^{j-1} \to c^j, d^j$ corresponds to a change of basis in V^{j-1}, $\{\phi_{j-1\,k};\, k \in \mathbb{Z}\} \to \{\phi_{jk}, \psi_{jk};\, k \in \mathbb{Z}\}$. Because of the underlying wavelet structure the orthogonal matrix associated to this basis change has a peculiar structure. The transition $c^j, d^j \to c^{j-1}$ is given by the transposed matrix; this is the reason why the adjoints H^*, G^* of H and G turn up in (2.52).

All the above is one-dimensional. As an image decomposition and reconstruction algorithm, Mallat's scheme is of course two-dimensional, and corresponds to a two-dimensional multiresolution analysis (see subsection 2A). Since the corresponding wavelet basis vectors can all be written as products of one-dimensional ψ_{jk}, ϕ_{jk} (see (2.19)–(2.22)), the two-dimensional algorithm itself can also be generated by a "tensor product" of the one-dimensional algorithm (see [8]). More specifically, the sequences to be decomposed are now elements of $l^2(\mathbb{Z}^2)$,

$$c^0 = \left(c_{mn}^0\right)_{m,n \in \mathbb{Z}},$$

and one defines G_r, H_r and G_c, H_c as the filters G, H defined by (2.49), but acting only on the first, respectively, the second, coefficient (r for "rows", c for "columns"). Then c^0 is decomposed into c^1 and *three* difference sequences (corresponding to the Ψ^j, $j = 1, 2, 3$,—see (2.20)–(2.22)) $d^{1,1}$, $d^{1,2}$ and $d^{1,3}$,

$$\begin{aligned} c^1 &= H_c H_r c^0, \\ d^{1,1} &= G_c H_r c^0, \\ d^{1,2} &= H_c G_r c^0, \\ d^{1,3} &= G_c G_r c^0. \end{aligned}$$

The operator $G_c H_r$ "smooths" over the column index, and looks at the "difference" ($\to$ high frequency information) for the row index; typically, $d^{1,1}$ will

be large when a horizontal edge is present. Similarly, $d^{1,2}$ detects vertical edges. It follows that, at no extra cost, Mallat's algorithm is orientation sensitive, which the two-dimensional Laplacian pyramid scheme of [20] was not. In [8], S. Mallat gives a very striking graphical representation of the whole two-dimensional scheme, illustrated with several examples, which clearly show, in particular, the orientation specificity of his algorithm.

3. Equivalence Between Mallat's Discrete Algorithm and Multiresolution Analysis

3.A. Weaning Mallat's algorithm from its multiresolution parent. Ultimately, Mallat's decomposition and reconstruction algorithm, i.e., (2.50), (2.51) and (2.52), deals only with sequences; the underlying multiresolution analysis is only used in the computation of the filter operators H and G. In this subsection we extract the properties of H and G that make the scheme work, without reference to multiresolution analysis.

These properties are very easy to deduce from subsection 2C. First of all, we impose

$$\begin{aligned} &\sum_n |h(n)| < \infty, \\ &\sum_n |g(n)| < \infty. \end{aligned} \tag{3.1}$$

This implies that the operators H, G, defined by

$$(Ha)_k = \sum_n h(n-2k)a_n,$$

$$(Ga)_k = \sum_n g(n-2k)a_n,$$

are bounded operators on $l^2(\mathbf{Z})$. This condition is satisfied by the $h(n)$, $g(n)$ in subsection 2C; it corresponds to a rather weak decay condition on ϕ. At later stages, we shall impose much stronger decay conditions on the $h(n)$.

A second condition follows from the decomposition formulas (2.50), (2.51) and the reconstruction formula (2.52). The scheme will only work if

$$H^*H + G^*G = \mathbb{1}. \tag{3.2}$$

The third condition expresses orthogonality. Essentially, the decomposition splits the original $l^2(\mathbf{Z})$ into a sum of subspaces. After the first step, we have

$$l^2(\mathbf{Z}) = H^*l^2(\mathbf{Z}) \oplus G^*l^2(\mathbf{Z});$$

after L iterations, one finds

$$l^2(\mathbf{Z}) = \bigoplus_{j=0}^{L-1} (H^*)^j G^* l^2(\mathbf{Z}) + (H^*)^L l^2(\mathbf{Z}).$$

In order to make the decomposition as sharp as possible, i.e., to remove correlations in the original sequence as much as possible, we require that these subspaces be orthogonal. That is, we require

$$HG^* = 0. \tag{3.3}$$

This condition is verified by the filter operators in subsection 2C. One finds

$$\begin{aligned}(HG^*)_{kl} &= \sum_n h(n-2k)g(n-2l) \\ &= \sum_n \langle \phi_{1k}, \phi_{0n}\rangle\langle \phi_{0n}, \psi_{1l}\rangle \\ &= \langle \phi_{1k}, \psi_{1l}\rangle = 0.\end{aligned}$$

So far, H and G play symmetrical roles in our conditions. The final condition will break that symmetry, and identify G as a "difference" operator, and H as an "averaging" operator. Let a be the sequence

$$a_n = \begin{cases} 1 & \text{for} \quad |n| \leqq N, \\ 0 & \text{for} \quad |n| > N, \end{cases}$$

where N is large compared to n_0, with

$$\sum_{|n| \geqq n_0} |h(n)| \leqq \varepsilon,$$

$$\sum_{|n| \geqq n_0} |g(n)| \leqq \varepsilon,$$

for some small ε. If H averages, i.e., corresponds to a low pass filter, and G corresponds to a band pass filter, then we expect (in regions away from the "edges" of a)

$$(Ha)_k \approx \begin{cases} C & \text{for} \quad |k| \leqq \tfrac{1}{2}N - n_0, \\ 0 & \text{for} \quad |k| \geqq \tfrac{1}{2}N + n_0, \end{cases}$$

$$(Ga)_k \approx 0 \quad \text{for} \quad |k| \leqq \tfrac{1}{2}N - n_0 \quad \text{and} \quad \text{for} \quad |k| \geqq \tfrac{1}{2}N + n_0.$$

This implies that we require

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

$$\sum_n h(n) = C.$$

The constant C can be determined as follows. For $N \to \infty$, the edge effects become negligible, and

$$\|Ga\|^2/\|a\|^2 \to 0,$$

$$\|Ha\|^2/\|a\|^2 \sim \sum_{|k| \leqq N/2} C^2/2N \to \tfrac{1}{2}C^2.$$

But

$$\|Ha\|^2 + \|Ga\|^2 = \langle a, (H^*H + G^*G)a \rangle = \|a\|^2,$$

hence $C = \sqrt{2}$. Thus our final conditions read

$$\begin{aligned} \sum_n h(n) &= \sqrt{2}, \\ \sum_n g(n) &= 0. \end{aligned} \tag{3.4}$$

These conditions are satisfied in subsection 2C. One has

$$\phi_{10} = \sum_n h(n)\phi_{0n};$$

hence, by integration,

$$2^{-1/2} \int dx\, \phi(\tfrac{1}{2}x) = \left[\sum_n h(n)\right] \int dx\, \phi(x),$$

or

$$\sum_n h(n) = \sqrt{2}, \quad \text{since} \quad \int dx\, \phi(x) \neq 0$$

(see (2.18)). Similarly,

$$\psi_{10} = \sum_n g(n)\phi_{0n};$$

since (see (2.17)) $\int dx\, \psi(x) = 0$, it follows that $\Sigma_n g(n) = 0$.

We have identified four conditions, (3.1)–(3.4), which guarantee that an algorithm "à la Mallat" works, and corresponds to averaging, respectively difference operations, followed by exact reconstruction. In terms of the $h(n)$, $g(n)$, conditions (3.2) and (3.3) can be rewritten as

$$\sum_k [h(m - 2k)h(n - 2k) + g(m - 2k)g(n - 2k)] = \delta_{mn} \tag{3.5}$$

and

$$\sum_n h(n-2k)g(n-2l) = 0. \tag{3.6}$$

In the remainder of this subsection, we shall rewrite the conditions (3.1)–(3.4) in various ways which make them more tractable to analysis.

In order to get rid of the factors 2 in (3.5), (3.6), we define

$$\begin{aligned} a(n) &= h(2n), \\ b(n) &= h(2n+1), \\ c(n) &= g(2n), \\ d(n) &= g(2n+1). \end{aligned} \tag{3.7}$$

Rewriting (3.5), (3.6) in terms of functions of a, b, c, d leads to

$$\sum_k [a(m-k)a(n-k) + c(m-k)c(n-k)] = \delta_{mn}, \tag{3.8a}$$

$$\sum_k [b(m-k)b(n-k) + d(m-k)d(n-k)] = \delta_{mn}, \tag{3.8b}$$

$$\sum_k [a(m-k)b(n-k) + c(m-k)d(n-k)] = 0, \tag{3.8c}$$

$$\sum_n [a(n-k)c(n-l) + b(n-k)d(n-l)] = 0. \tag{3.8d}$$

In this form the conditions are completely expressed in terms of convolutions of the sequences a, b, c, d. It is therefore natural to introduce the 2π-periodic functions

$$\begin{aligned} \alpha(\xi) &= \sum_n a(n)e^{in\xi}, \\ \beta(\xi) &= \sum_n b(n)e^{in\xi}, \\ \gamma(\xi) &= \sum_n c(n)e^{in\xi}, \\ \delta(\xi) &= \sum_n d(n)e^{in\xi}, \end{aligned} \tag{3.9}$$

and to rewrite the conditions in terms of these functions. We obtain

$$|\alpha(\xi)|^2 + |\gamma(\xi)|^2 = 1, \tag{3.10a}$$

$$|\beta(\xi)|^2 + |\delta(\xi)|^2 = 1, \tag{3.10b}$$

$$\alpha(\xi)\overline{\beta(\xi)} + \gamma(\xi)\overline{\delta(\xi)} = 0, \tag{3.10c}$$

$$\alpha(\xi)\overline{\gamma(\xi)} + \beta(\xi)\overline{\delta(\xi)} = 0. \tag{3.10d}$$

These conditions are obviously not independent. Except for trivial solutions, which would be in contradiction with (3.4), i.e., with

$$\alpha(0) + \beta(0) = \sqrt{2}, \tag{3.11a}$$

$$\gamma(0) + \delta(0) = 0, \tag{3.11b}$$

we find from (3.10c) and (3.10d)

$$\begin{aligned} \gamma(\xi) &= e^{i\lambda(\xi)}\overline{\beta(\xi)}, \\ \delta(\xi) &= -e^{i\lambda(\xi)}\overline{\alpha(\xi)}, \end{aligned} \tag{3.12}$$

where λ is a real function such that $\lambda(\xi + 2\pi) - \lambda(\xi) \in 2\pi\mathbb{Z}$ for all ξ. For the sake of simplicity we shall restrict ourselves to $\lambda(\xi) = 0$ for the moment. We thus choose

$$\begin{aligned} \gamma(\xi) &= \overline{\beta(\xi)}, \\ \delta(\xi) &= -\overline{\alpha(\xi)}. \end{aligned} \tag{3.13}$$

The only equation remaining from the system (3.10) is then

$$|\alpha(\xi)|^2 + |\beta(\xi)|^2 = 1. \tag{3.14}$$

The choice (3.13), together with (3.11b), also implies

$$\alpha(0) - \beta(0) = 0.$$

Hence (from (3.11a)),

$$\alpha(0) = \beta(0) = 2^{-1/2}, \tag{3.15}$$

which agrees with (3.14) for $\xi = 0$. It follows that *any* choice of 2π-periodic functions α and β satisfying (3.14), (3.15) and $\Sigma|a_n| < \infty$, $\Sigma_n|b_n| < \infty$, leads, via (3.13), (3.9) and (3.7), to two filter operators H and G satisfying (3.1)–(3.4). These filter operators can then be used for a decomposition and reconstruction algorithm "à la Mallat", without reference to multiresolution analysis.

Remarks. 1. The system of equations (3.10) can also be rewritten as one matrix equation. If we define the 2×2 matrix-valued 2π-periodic function $M(\xi)$ by

$$M(\xi) = \begin{pmatrix} \alpha(\xi) & \gamma(\xi) \\ \beta(\xi) & \delta(\xi) \end{pmatrix}, \tag{3.16}$$

then (3.10) states that $M(\xi)$ should be unitary, for all ξ.

2. Note that, in view of (3.9) and (3.7), the choice (3.13) is equivalent to

$$g(n) = (-1)^n h(-n+1). \tag{3.17}$$

The equations (3.14) and (3.15) involve only the $h(n)$. They can be rewritten as

$$\sum_n h(n-2k)h(n-2l) = \delta_{kl} \tag{3.18}$$

and

$$\sum_n h(2n) = \sum_n h(2n+1) = 2^{-1/2}.$$

This last condition is implied by (3.18) and

$$\sum h(n) = 2^{1/2}. \tag{3.19}$$

3. If one introduces the 2π-periodic function $H(\xi)$,

$$H(\xi) = \sum_n h(n) e^{in\xi},$$

then the conditions (3.14), (3.15) can also be written in terms of H. Clearly,

$$H(\xi) = \alpha(2\xi) + e^{i\xi}\beta(2\xi),$$

or

$$\alpha(2\xi) = \tfrac{1}{2}[H(\xi) + H(\xi+\pi)],$$

$$\beta(2\xi) = \tfrac{1}{2}e^{-i\xi}[H(\xi) - H(\xi+\pi)].$$

Then (3.14), (3.15) are equivalent with

$$|H(\xi)|^2 + |H(\xi + \pi)|^2 = 2 \tag{3.20}$$

and

$$H(0) = \sqrt{2}\,. \tag{3.21}$$

Under the form (3.20) this condition is not new. It can be found in [16], where $m_0(\xi) = 2^{-1/2}H(\xi)$ is used, rather than H. While this paper was being written, S. Mallat pointed out to me that (3.20) is very similar to a condition derived by M. Smith and T. Barnwell [24] in the construction of "conjugate quadrature filters". In fact, (3.20) is identical to their condition. Smith and Barnwell were looking for, and found, a tree-structured two-band coding scheme with exact reconstruction, which is exactly what this subsection is about! The constructions given later (at least insofar as they describe discrete filters) are therefore, in fact, special cases of their construction. Ultimately, however, our aim here is to construct orthonormal wavelet bases of compact support, which is a very different point of view. Even from the filter point of view, our results go further than Smith and Barnwell's, in that we give complete characterization of the possible filters. We shall however not go into this here.

4. Similarly one can introduce $G(\xi) = \sum_n g(n)e^{in\xi}$. The matrix statement (3.16) is then equivalent to the requirement that the matrix

$$\frac{1}{\sqrt{2}}\begin{pmatrix} H(\xi) & G(\xi) \\ H(\xi + \pi) & G(\xi + \pi) \end{pmatrix} \tag{3.22}$$

be unitary. This is the form under which this requirement appears in [16]. Depending on what one wants to do, (3.22) and (3.20) may or may not be more useful than (3.16) and (3.14). The advantage of (3.14), (3.16) is that no correlations are introduced, as in (3.20), (3.22), linking the behavior of H at $\xi + \pi$ with its values at ξ. The conditions (3.16) or (3.22) can be generalized to situations where three or more band filters are considered (corresponding to decimations with factors $3, 4, \cdots$ rather than 2), or even more complicated structures, in more than one dimension (associated with lattices in $\mathbf{Z}^d$; see [21]). It was pointed out to me by P. Auscher [25] that in these cases the generalization of (3.16) is more useful, for practical construction, than the generalization of (3.22), precisely because it avoids introducing correlations.

5. Note that $\sum_n h(2n) = \sum_n h(2n + 1)$, which is a consequence of (3.18)–(3.19) (see Remark 2 above) implies that *all* the possible $H(\xi)$, satisfying all the above conditions, necessarily are divisible by $(1 + e^{i\xi})$ (see subsection 2B).

Finally, let us conclude this subsection with some simple examples.

EXAMPLE 3.1. The simplest possible example is

$$\alpha(\xi) = \beta(\xi) = 2^{-1/2},$$

corresponding to

$$h(0) = 2^{-1/2}, \qquad g(0) = 2^{-1/2},$$

$$h(1) = 2^{-1/2}, \qquad g(1) = -2^{-1/2},$$

all the other $h(n)$, $g(n)$ being zero.

EXAMPLE 3.2. The next simplest example is

$$\alpha(\xi) = 2^{-1/2}\left[\nu(\nu - 1) + (\nu + 1)e^{i\xi}\right]/(\nu^2 + 1),$$

$$\beta(\xi) = 2^{-1/2}\left[(1 - \nu) + \nu(\nu + 1)e^{i\xi}\right]/(\nu^2 + 1),$$

where ν is an arbitrary real number. This corresponds to

$$h(0) = 2^{-1/2}\nu(\nu - 1)/(\nu^2 + 1), \qquad g(0) = 2^{-1/2}\nu(\nu + 1)/(\nu^2 + 1),$$

$$h(1) = 2^{-1/2}(1 - \nu)/(\nu^2 + 1), \qquad g(1) = -2^{-1/2}(\nu + 1)/(\nu^2 + 1),$$

$$h(2) = 2^{-1/2}(\nu + 1)/(\nu^2 + 1), \qquad g(2) = 2^{-1/2}(1 - \nu)/(\nu^2 + 1),$$

$$h(3) = 2^{-1/2}\nu(\nu + 1)/(\nu^2 + 1), \qquad g(3) = -2^{-1/2}\nu(\nu - 1)/(\nu^2 + 1),$$

all the other $h(n)$, $g(n)$ being zero.

Note. We have here taken

$$g(n) = (-1)^n h(3 - n)$$

rather than (3.17); this shift corresponds simply to choosing $\lambda(\xi) = \xi$ instead of 0 in (3.12).

3.B. Introducing a regularity condition. In the preceding subsection we derived and discussed a set of necessary and sufficient conditions, directly on the filter operators, for Mallat's algorithm to work. All these conditions concerned only what happened in one step of decomposition/reconstruction. In the discussion, in subsection 2B, of the Laplacian pyramid scheme, we saw that it is also important that the iterated reconstruction, applied to a sequence consisting of only one non-zero entry, looks still reasonably nice, even after several iterations.

In Mallat's algorithm, a sequence c^0 is decomposed into $d^1, \cdots, d^L, c^L$, with $d^j = GH^{j-1}c^0$, and $c^L = H^L c^0$; the reconstruction formula is then (cf. (2.52))

$$c^0 = \sum_{j=1}^{L} (H^*)^{j-1} G^* d^j + (H^*)^L c^L.$$

The iterated filter operator is thus H^*. It is therefore important (see subsection 2B) to study the behavior of $(H^*)^l e$, for large l, where e is a sequence with only one non-zero entry, e.g. $e_n = \delta_{n0}$. Ideally we want the graphical representation (with histograms—see Figures 2, 3 in subsection 2B) of $(H^*)^l e$ to look "nice", which expresses itself by convergence, for $l \to \infty$, to a reasonably regular function.

To show that this is a genuine concern, we have plotted, in Figure 4, the histogram representation of $(H^*e), \cdots, (H^*)^6 e$, for H^* chosen as in Example 3.2

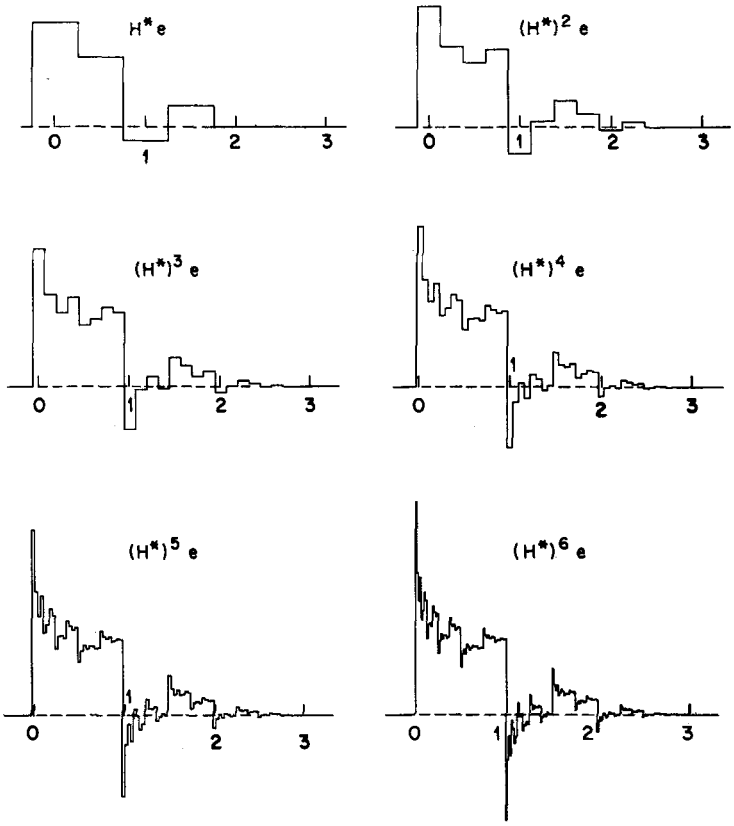


Figure 4. The histogram representations of $(H^*)^j e$, $j = 1, \cdots, 6$, for $h(n)$ which do not satisfy a regularity condition (see text).

(subsection 3A), with $\nu = -1.5$. For increasing l, $(H^*)^l e$ becomes increasingly messy; in fact, $(H^*)^l e$ converges, for $l \to \infty$, to a discontinuous, fractal function.

As in subsection 2B, we represent $(H^*)^l e$ by a histogram η_l with step width 2^{-l}, and with amplitudes given by the successive $2^{l/2}((H^*)^l e)_n$. The normalization, different from that in subsection 2B (because $\Sigma h(n) = \sqrt{2}$ and not 1), is again chosen so that the area under the histogram remains 1 for every l. The stepfunction η_l can be written as (see subsection 2B)

$$\eta_l(x) = \left(T_H^l \chi_{[-1/2, -1/2[}\right)(x), \tag{3.23}$$

where

$$(T_H f)(x) = \sqrt{2} \sum_n h(n) f(2x - n). \tag{3.24}$$

By taking Fourier transforms, (3.23) and (3.24) lead to

$$\hat{\eta}_l(\xi) = (2\pi)^{-1/2} \left[\prod_{j=1}^{l} m_0(2^{-j}\xi)\right] \frac{\sin(2^{-l-1}\xi)}{2^{-l-1}\xi}, \tag{3.25}$$

where $m_0(\xi) = 2^{-1/2} \Sigma_n h(n) e^{in\xi}$. Hence, at least in a formal sense, $\eta_l \to \eta_\infty$ for $l \to \infty$, with

$$\hat{\eta}_\infty(\xi) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} m_0(2^{-j}\xi). \tag{3.26}$$

The following lemma ensures that $\hat{\eta}_\infty$ is well defined, i.e., that the infinite product in (3.26) converges, at least pointwise.

LEMMA 3.1. *Suppose that, for some* $\varepsilon > 0$,

$$\sum_n |h(n)| |n|^\varepsilon < \infty. \tag{3.27}$$

Then (3.26) *converges pointwise, for all* $\xi \in \mathbb{R}$. *The convergence is uniform on compact sets.*

Proof: Since $\Sigma h(n) = \sqrt{2}$, we have $m_0(\xi) = 1 + 2^{-1/2} \Sigma_n h(n)(e^{in\xi} - 1)$, hence $|m_0(\xi) - 1| \leq \sqrt{2} \Sigma_n |h(n)| |\sin \frac{1}{2} n\xi|$. For any $0 < \delta \leq 1$ there exists C_δ such that, for all $\alpha \in \mathbb{R}$, $|\sin \alpha| \leq C_\delta |\alpha|^\delta$. It follows that

$$|m_0(\xi) - 1| \leq C \left[\sum_n |h(n)| |n|^{\min(1, \varepsilon)}\right] \cdot |\xi|^{\min(1, \varepsilon)};$$

hence

$$|m_0(2^{-j}\xi) - 1| \leq C\lambda^{-j} |\xi|^{\min(1, \varepsilon)}, \tag{3.28}$$

where $\lambda = 2^{\min(1,\varepsilon)} > 1$. This is sufficient to ensure convergence of (3.26), for any $\xi \in \mathbb{R}$. It immediately follows from (3.28) that the convergence is uniform on compact sets.

Remark. While being more restrictive than (3.1), the condition (3.27) is still very mild. In practice one requires much stronger decay for the $h(n)$. For filter construction purposes, one even restricts oneself to the case where only finitely many $h(n)$ are different from zero.

It is however not sufficient to know that $\hat{\eta}_\infty$ is well defined. In order to avoid situations such as depicted in Figure 4, we require that (i) $\hat{\eta}_\infty$ has sufficient decay, so that η_∞ is sufficiently regular (at least continuous), and (ii) η_l converges to η_∞, pointwise, for $l \to \infty$.

To ensure the decay, for $|\xi| \to \infty$, of $\hat{\eta}_\infty(\xi)$, we shall use the same trick as in subsection 2B, i.e., we shall require that $m_0(\xi)$ is divisible by $(1 + e^{i\xi})^N$, for some $N > 0$. The precise statement is given in the following lemma, using an estimation technique of P. Tchamitchian [5].

LEMMA 3.2. *If* $m_0(\xi) = (1 + e^{i\xi})]^N \mathscr{F}(\xi)$, *where* $\mathscr{F}(\xi) = \sum_n f(n) e^{in\xi}$ *satisfies*

$$\sum |f(n)|\,|n|^\varepsilon < \infty \quad \text{for some} \quad \varepsilon > 0 \tag{3.29}$$

and

$$\sup_{\xi \in \mathbb{R}} |\mathscr{F}(\xi)| = B, \tag{3.30}$$

then there exists $C > 0$ *such that, for all* $\xi \in \mathbb{R}$,

$$\left| \prod_{j=1}^{\infty} m_0(2^{-j}\xi) \right| \leq C(1 + |\xi|)^{-N + (\log B)/(\log 2)}. \tag{3.31}$$

Remarks. 1. It follows from (3.31) that η_∞ is continuous if H satisfies all the above conditions, and if $B < 2^{N-1}$.

2. The condition (3.29) will automatically be satisfied if

$$\sum_n |h(n)|\,|n|^{N+\varepsilon} < \infty. \tag{3.32}$$

Proof: Since $\prod_{j=1}^{\infty}\cos(2^{-j}x) = x^{-1}\sin x$, we have

$$\prod_{j=1}^{\infty} m_0(2^{-j}\xi) = \left[e^{i\xi/2}\prod_{j=1}^{\infty}\cos(2^{-j-1}\xi)\right]^N \prod_{j=1}^{\infty}\mathscr{F}(2^{-j}\xi)$$

(3.33)

$$= e^{iN\xi/2}\left(\frac{\sin\frac{1}{2}\xi}{\frac{1}{2}\xi}\right)^N \prod_{j=1}^{\infty}\mathscr{F}(2^{-j}\xi),$$

where the right-hand side converges uniformly on compact sets because of (3.29). There exists therefore a constant C such that, for all $|\xi| \leqq 1$,

$$\left|\prod_{j=1}^{\infty} m_0(2^{-j}\xi)\right| \leqq C. \tag{3.34}$$

Take now $|\xi| > 1$. Determine $j_0 \in \mathbb{N}$ such that

$$2^{-j_0}|\xi| < 1 \leqq 2^{-j_0+1}|\xi|,$$

i.e.,

$$\log|\xi|/\log 2 < j_0 \leqq 1 + \log|\xi|/\log 2.$$

Then

$$\left|\prod_{j=1}^{\infty}\mathscr{F}(2^{-j}\xi)\right| = \left|\prod_{j=1}^{j_0}\mathscr{F}(2^{-j}\xi)\right| \cdot \left|\prod_{j=1}^{\infty}\mathscr{F}(2^{-j}2^{-j_0}\xi)\right|$$

(3.35)

$$\leqq B^{j_0}\prod_{j=1}^{\infty}\left(1 + 2^{-j\varepsilon}\sum_n |f(n)|\,|n|^{\varepsilon}\right)$$

$$\leqq C\exp\{\log B \cdot \log|\xi|/\log 2\},$$

To estimate $\prod_{j=1}^{\infty}\mathscr{F}(2^{-j}2^{-j_0}\xi)$ we have used the same argument as in the proof of Lemma 3.1. This is allowed since $\Sigma f(n) = \mathscr{F}(0) = m_0(0) = 1$. Together, (3.35), (3.34) and (3.33) imply (3.31).

In our search for "regularity" we have, so far, only used one of the special conditions on the $h(n)$, derived in subsection 3A, namely (3.4), $\Sigma_n h(n) = \sqrt{2}$. And even that has not played a critical role, since it was only used for normalization purposes, and we could have as easily normalized by any other constant which happened to be the sum of the $h(n)$. For our last step, the proof that the histograms η_l converge pointwise to the continuous function η_∞ (assuming B is not too large), we need an extra ingredient, namely $|m_0(\xi)| \leqq 1$. Since,

however (see (3.20)), as a consequence of (3.2)–(3.3), $|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1$, this condition is automatically fulfilled for $h(n)$ satisfying (3.18)–(3.19).

PROPOSITION 3.3. *Define* $m_0(\xi) = 2^{-1/2}\sum_n h(n)e^{in\xi}$, *where the* $h(n)$ *satisfy* (3.18), (3.19). *Suppose moreover that*

$$m_0(\xi) = \left[\tfrac{1}{2}(1 + e^{i\xi})\right]^N \mathcal{F}(\xi), \tag{3.36}$$

with $\mathcal{F}(\xi) = \sum_n f(n)e^{in\xi}$ *such that*

$$\sum_n |f(n)|\,|n|^\varepsilon < \infty \quad \text{for some} \quad \varepsilon > 0 \tag{3.37}$$

and

$$\sup_{\xi \in \mathbf{R}} |\mathcal{F}(\xi)| = B < 2^{N-1}. \tag{3.38}$$

Then the piecewise constant functions η_l, *defined recursively by*

$$\eta_l(x) = \sqrt{2}\sum_n h(n)\eta_{l-1}(2x - n), \tag{3.39}$$

with

$$\eta_0(x) = \chi_{[-1/2,1/2[}(x),$$

converge pointwise to the continuous function η_∞ *defined by*

$$\hat{\eta}_\infty(\xi) = (2\pi)^{-1/2}\prod_{j=1}^{\infty} m_0(2^{-j}\xi).$$

Proof: 1. As an intermediate step, we prove $\mu_l \to \eta_\infty$, pointwise, where the μ_l are defined in the same recursive way as the η_l, but starting from a different initial function,

$$\mu_0(x) = \begin{cases} 1 + x, & -1 \leqq x \leqq 0, \\ 1 - x, & 0 \leqq x \leqq 1, \\ 0, & \text{otherwise.} \end{cases}$$

2. Taking Fourier transforms, we find

$$\hat{\mu}_l(\xi) = (2\pi)^{-1/2}\left[\prod_{j=1}^{l} m_0(2^{-j}\xi)\right]\left[\frac{\sin(2^{-l-1}\xi)}{2^{-l-1}\xi}\right]^2.$$

From Lemma 3.1 it follows that $\hat{\mu}_l \to \hat{\eta}_\infty$, uniformly on compact sets. This

implies that, for all $\delta > 0$, and for all $R > 0$, we can find l_0 such that, for all $l \geqq l_0$,

$$\int_{|\xi| \leqq R} d\xi \, |\hat{\mu}_l(\xi) - \hat{\eta}_\infty(\xi)| \leqq \delta.$$

On the other hand, $\hat{\eta}_\infty \in L^1$ since $B < 2^{N-1}$. It follows that for all $\delta > 0$ there exists R such that

$$\int_{|\xi| \geqq R} d\xi \, |\hat{\eta}_\infty(\xi)| \leqq \delta.$$

L^1-convergence of $\hat{\mu}_l$ to $\hat{\eta}_\infty$, which implies pointwise convergence of μ_l to η_∞, will then follow if we can prove that, for all $\delta > 0$, there exist R and l_0 large enough, so that, for all $l \geqq l_0$,

$$\int_{|\xi| \geqq R} d\xi \, |\hat{\mu}_l(\xi)| \leqq \delta.$$

3. We need thus to evaluate the integral

$$\int_{|\xi| \geqq R} d\xi \, |P_l(\xi)| \left| \frac{\sin(2^{-l-1}\xi)}{2^{-l-1}\xi} \right|^2,$$

where $P_l(\xi) = \prod_{j=1}^{l} m_0(2^{-j}\xi)$. To do this, we split the integral into two parts, namely $|\xi| \geqq 2^l\pi$ and $R \leqq |\xi| \leqq 2^l\pi$. To evaluate these two parts, we shall use the following three properties of P_l:

(i) $|P_l(\xi)| \leqq 1$, $\quad$ (since $|m_0(\xi)| \leqq 1$),

(ii) $$|P_l(\xi)| \leqq \left[\prod_{j=1}^{l} |\cos(2^{-j}\xi)| \right]^N \prod_{j=1}^{l} |\mathscr{F}(2^{-j}\xi)|$$

$$\leqq C \left| \frac{2^{-l} \sin \tfrac{1}{2}\xi}{\sin(2^{-l-1}\xi)} \right|^N (1 + |\xi|)^\beta,$$

where $\beta = \log B / \log 2$ (use the proof of Lemma 3.2) and

(iii) P_l is periodic, with period $2^{l+1}\pi$.

4. We concentrate first on $|\xi| \geq 2^l\pi$. Using the periodicity of P_l, we find

$$\int_{|\xi|\geq 2^l\pi} d\xi\,|P_l(\xi)|\left|\frac{\sin(2^{-l-1}\xi)}{2^{-l-1}\xi}\right|^2$$

$$= \sum_{k\neq 0}\int_{|\xi|\leq 2^l\pi} d\xi\,|P_l(\xi)|\frac{|\sin(2^{-l-1}\xi)|^2}{|2^{-l-1}\xi + k\pi|^2}$$

$$\leq C\int_{|\xi|\leq 2^l\pi} d\xi\,|P_l(\xi)|\,|\sin(2^{-l-1}\xi)|^2.$$

Choose $\lambda = 2^{-\alpha l}$, with $\alpha \in]0,1[$ to be fixed later. Then

$$\int_{|\xi|\leq 2^l\pi} d\xi\,|P_l(\xi)|\,|\sin(2^{-l-1}\xi)|^2$$

(3.40)

$$\leq \tfrac{1}{4}\lambda^2\int_{|\xi|\leq 2^l\lambda} d\xi\,|P_l(\xi)| + \int_{2^l\lambda\leq|\xi|\leq 2^l\pi} d\xi\,|P_l(\xi)|.$$

Now

$$\int_{|\xi|\leq 2^l\lambda} d\xi\,|P_l(\xi)|$$

$$\leq \int_{|\xi|\leq 1} d\xi\,|P_l(\xi)| + C\int_{1\leq|\xi|\leq 2^l\lambda} d\xi\,(1+|\xi|)^\beta|2^l\sin(2^{-l-1}\xi)|^{-N}$$

$$\leq 1 + 2^N C\int_1^\infty dx\,(1+x)^\beta x^{-N} = C_1,$$

where C_1 is finite because $N - \beta > 1$.

On the other hand,

$$\int_{2^l\lambda\leq|\xi|\leq 2^l\pi} d\xi\,|P_l(\xi)|$$

$$\leq C2^{-lN}(1+2^l\pi)^\beta 2^l\int_\lambda^\pi dx\,|\sin\tfrac{1}{2}x|^{-N}$$

$$\leq C_2 2^{l(1+\beta-N)}\lambda^{-N}.$$

Putting it all together, and choosing $\alpha = (N-\beta-1)/(N+2) \in]0,1[$, this implies that (3.40) is

$$\leq C_3 2^{-2l(N-\beta-1)/(N+2)}. \tag{3.41}$$

This clearly tends to zero for $l \to \infty$.

5. We now evaluate the integral of $|\hat{\mu}_l|$ over $R \leqq |\xi| \leqq 2^l\pi$. Since $|\sin x| \geqq 2|x|/\pi$ for $|x| \leqq \frac{1}{2}\pi$, we find

$$\int_{R \leqq |\xi| \leqq 2^l\pi} d\xi\, |P_l(\xi)| \left| \frac{\sin(2^{-l-1}\xi)}{2^{-l-1}\xi} \right|^2$$

$$\leqq 4C \int_{R \leqq |\xi| \leqq 2^l\pi} d\xi\, (1 + |\xi|)^\beta |\xi|^{-2} |\sin \tfrac{1}{2}\xi|^N \left| \frac{2^{-l}}{\sin(2^{-l-1}\xi)} \right|^{N-2}$$

$$\leqq 4C\pi^{N-2} \int_R^\infty dx\, (1 + x)^\beta x^{-N}.$$

Since $N - \beta - 1 > 0$, this tends to zero for $R \to \infty$, uniformly in l. Together with (3.41) this proves that

$$\int_{|\xi| \geqq R} d\xi\, |\hat{\mu}_l(\xi)|$$

can be made as small as wanted, by choosing l and R large enough. As pointed out in point 2, this proves $\|\hat{\mu}_l - \hat{\eta}_\infty\|_{L^1} \to_{l \to \infty} 0$.

6. We have thus proved that $\mu_l \to \eta_\infty$, pointwise. In fact, we can even show a little bit more. The same arguments (points 2 → 5) as above can be stretched a little to prove

$$\int d\xi\, (1 + |\xi|)^\lambda |\hat{\eta}_\infty(\xi)| < \infty$$

and

$$\int d\xi\, (1 + |\xi|)^\lambda |\hat{\eta}_\infty(\xi) - \hat{\mu}_l(\xi)| \xrightarrow[l \to \infty]{} 0,$$

where

$$\lambda = \tfrac{1}{2}(N - \beta - 1) > 0.$$

Consequently, η_∞ is λ-Lipschitz,

$$|\eta_\infty(x) - \eta_\infty(y)| \leqq C|x - y|^\lambda,$$

and the convergence $\mu_l \to \eta_\infty$ is uniform on compact sets.

7. Finally, we only need to show that pointwise convergence of the μ_l implies pointwise convergence of the η_l. The two functions μ_0 and η_0 agree on integers,

$$\mu_0(0) = \eta_0(0) = 1,$$

$$\mu_0(k) = \eta_0(k) = 0 \quad \text{for} \quad k \in \mathbb{Z},\ k \neq 0.$$

Using the recursion relation (3.39), which both the μ_l and the η_l satisfy, one sees that this implies, for all $l \in \mathbb{N}$,

$$\eta_l(2^{-l}k) = \mu_l(2^{-l}k) \quad \text{for all} \quad k \in \mathbb{Z}.$$

Let $x \in \mathbb{R}$ be arbitrary. For any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|x - y| \leqq \delta \Rightarrow |\eta_\infty(x) - \eta_\infty(y)| \leqq \tfrac{1}{2}\varepsilon.$$

There also exists l_0 such that, for all $l \geqq l_0$, and all $y \in [x - \delta, x + \delta]$, one has

$$|\eta_\infty(y) - \mu_l(y)| \leqq \tfrac{1}{2}\varepsilon.$$

Choose $l \geqq l_1 = \max(l_0, -\ln \delta/\ln 2)$. Since η_l is piecewise constant, with step width 2^{-l}, it follows that there exists $k \in \mathbb{Z}$ such that

$$|x - 2^{-l}k| \leqq 2^{-l} \leqq \delta$$

and

$$\eta_l(x) = \eta_l(2^{-l}k) = \mu_l(2^{-l}k).$$

Hence

$$|\eta_l(x) - \eta_\infty(x)| \leqq |\mu_l(2^{-l}k) - \eta_\infty(2^{-l}k)| + |\eta_\infty(2^{-l}k) - \eta_\infty(x)| \leqq \varepsilon.$$

Since ε was arbitrary, this shows that η_l converges pointwise to η_∞, for $l \to \infty$.

Remarks. 1. Using only slightly modified arguments, one proves, under the same conditions (in fact, only $B < 2^{N-1/2}$ is needed) that $\eta_l \to \eta_\infty$ in L^2, for $l \to \infty$. One simply replaces the L^1-estimates for $\eta_\infty - \mu_l$ by L^2-estimates for $\eta_\infty - \eta_l$ (no intermediary μ_l are needed).

2. As noted above, it is sufficient that

$$\sum_n |h(n)||n|^{N+\varepsilon} < \infty$$

to ensure (3.37).

3. The $h(n)$ of Example 3.1 do not satisfy the conditions of the proposition, since in this case

$$m_0(\xi) = \tfrac{1}{2}(1 + e^{i\xi}),$$

hence $N = 1$, $B = |\mathscr{F}(\xi)| = 1$, and therefore $B = 2^{N-1}$. However, in this case one checks directly that

$$\eta_l = \chi_{[-2^{-l-1}, 1-2^{-l-1}[}.$$

The limit η_∞ is not continuous in this case, $\eta_\infty = \chi_{[0,1[}$, but the pointwise convergence $\eta_l \to \eta_\infty$ still holds a.e.

4. The coefficients $h(n)$ defined by

$$h(0) = h(3) = 2^{-1/2},$$

$$h(n) = 0 \quad \text{otherwise},$$

satisfy all the "discrete" conditions of subsection 3A, but do not satisfy the conditions in the last proposition (for the same reason as the $h(n)$ of Example 3.1). In this case, however, the pointwise convergence of the η_l fails on a whole interval. It is easy to check that, for any l, the η_l take only two values, 0 and 1. (The easiest way to check this is to use the "graphical" construction (2.40) of the η_l—see subsection 2B and Figure 3.) On the other hand,

$$m_0(\xi) = \tfrac{1}{2}(1 + e^{3i\xi}),$$

hence

$$\hat{\eta}_\infty(\xi) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} m_0(2^{-j}\xi) = (2\pi)^{-1/2} e^{3i\xi/2} \frac{\sin \frac{3}{2}\xi}{\frac{3}{2}\xi}$$

or

$$\eta_\infty = \tfrac{1}{3}\chi_{[0,3]}.$$

There is therefore no pointwise convergence for any x between 0 and 3. The L^2-convergence fails too, since $\|\eta_\infty\|_{L^2}^2 = \frac{1}{3}$, whereas for all finite l, η_l is the characteristic function of a union of intervals, and hence $\|\eta_l\|_{L^2}^2 = \|\eta_l\|_{L^1} = \hat{\eta}_l(0) = 1$.

5. Only two values of ν, in Example 3.2, lead to coefficients $h(n)$ that satisfy the conditions of the proposition. They correspond to $m_0(\xi)$ divisible by $(1 + e^{i\xi})^2$. As noted above, all $m_0(\xi)$ satisfying the discrete conditions in subsection 3A are divisible by $(1 + e^{i\xi})$ (see Remark 5 at the end of subsection 3A). In Example 3.2, extra divisibility by another factor $(1 + e^{i\xi})$ leads to the condition

$$h(1) - h(3) = 2h(0),$$

or

$$\nu = \pm 1/\sqrt{3}.$$

The corresponding $h(0), \cdots, h(3)$ are

$$\begin{aligned} h(0) &= (1 \mp \sqrt{3})/(4\sqrt{2}), \\ h(1) &= (3 \mp \sqrt{3})/(4\sqrt{2}), \\ h(2) &= (3 \pm \sqrt{3})/(4\sqrt{2}), \\ h(3) &= (1 \pm \sqrt{3})/(4\sqrt{2}). \end{aligned} \tag{3.42}$$

We shall come back to these $h(n)$ later.

With Proposition 3.3 we have completed our program of writing a set of explicit conditions on the $h(n)$, $g(n)$, without reference to a multiresolution analysis background, which make Mallat's algorithm work, and which moreover lead to filters with sufficient "regularity".

In the case where the $h(n)$, $g(n)$ are calculated starting from a multiresolution analysis (see subsection 2C), one has

$$h(n) = \langle \phi_{10}, \phi_{0n} \rangle,$$

or

$$\phi_{10} = \sum_n h(n)\phi_{0n},$$

i.e.,

$$\phi\left(\tfrac{1}{2}x\right) = 2^{1/2} \sum_n h(n)\phi(x - n).$$

This is equivalent to

$$\hat{\phi}(\xi) = 2^{-1/2} \sum_n h(n) e^{in\xi/2}\, \hat{\phi}\left(\tfrac{1}{2}\xi\right) = m_0\left(\tfrac{1}{2}\xi\right)\hat{\phi}\left(\tfrac{1}{2}\xi\right).$$

It follows that

$$\hat{\phi}(\xi) = \left[\prod_{j=1}^{\infty} m_0(2^{-j}\xi)\right]\hat{\phi}(0), \tag{3.43}$$

or, since $\hat{\phi}(0) = (2\pi)^{-1/2} \int dx\, \phi(x) = (2\pi)^{-1/2}$ (see (2.18)),

$$\phi(x) = \eta_\infty(x). \tag{3.44}$$

As pointed out in subsection 2B, the $\eta_l = T^l \chi_{[-1/2,1/2[}$ can also be computed via a different recursion, (2.40), which we shall call the "graphical" recursion, and which lies at the basis of the graphical construction technique illustrated by Figure 3. It follows from (3.44) that, in the case where the $h(n)$ are derived from a multiresolution analysis framework, the graphical construction by iteration (see Figure 3, where the $h(n)$ now play the role of the $w(n)$) is therefore nothing but a reconstruction of the function ϕ; in the limit for $l \to \infty$, finer and finer detail is achieved for increasing l.

3.C. Equivalence between the discrete conditions and multiresolution analysis. So far we have formulated conditions, directly on the $h(n)$, which ensure that S. Mallat's algorithm works (with these coefficients), and has regularity (in the sense given to it at the end of subsection 2B, or in subsection 3B). We have seen for every condition how the coefficients $h(n)$ computed from a multiresolution analysis fit into the picture. The main result of this subsection is that these multiresolution-based examples are the *only* ones. It turns out that *any* sequence

of $h(n)$ satisfying the conditions in subsections 3A and 3B corresponds to a multiresolution analysis. The function η_∞ defined by (3.26) is then exactly the function ϕ from the multiresolution structure.

To prove this equivalence, we start from a sequence $h(n)$ satisfying (3.18), (3.19) and (3.27). We also assume that the function $m_0(\xi) = 2^{-1/2}\sum_n h(n)e^{in\xi}$ satisfies all the conditions in Proposition 3.3. We then define, as in (3.17),

$$g(n) = (-1)^n h(-n+1), \tag{3.45}$$

and, as in (3.44),

$$\phi(x) = \eta_\infty(x),$$

or

$$\hat{\phi}(\xi) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} m_0(2^{-j}\xi). \tag{3.46}$$

From the proof of Proposition 3.3 we know that ϕ is a bounded, uniformly continuous function; since $\hat{\phi} \in L^1 \cap L^\infty$, one also has $\phi \in L^2$. We define, in accordance with (2.16),

$$\psi(x) = \sqrt{2}\sum_n g(n)\phi(2x-n). \tag{3.47}$$

Since $\sum_n |g(n)| = \sum_n |h(n)| < \infty$, it follows that

$$|\hat{\psi}(\xi)| \leqq 2^{-1/2}\sum_n |h(n)| \cdot |\hat{\phi}(\tfrac{1}{2}\xi)|.$$

All the estimates of subsection 3B on η_∞ carry over, therefore, to ψ, and one finds that ψ is a bounded, uniformly continuous L^2-function. As before, we define $\psi_{jk}(x) = 2^{-j/2}\psi(2^{-j}x - k)$, and $\phi_{jk}(x) = 2^{-j/2}\phi(2^{-j}x - k)$. The definitions (3.46) and (3.47) immediately imply

$$\phi_{jk} = \sum_n h(n-2k)\phi_{j-1\,n}, \tag{3.48}$$

$$\psi_{jk} = \sum_n g(n-2k)\phi_{j-1\,n}. \tag{3.49}$$

We shall prove that the ψ_{jk} constitute an orthonormal basis of $L^2(\mathbb{R})$. In a first step we prove some orthogonality relations.

LEMMA 3.4. *Let $h(n)$ satisfy* (3.18), (3.19), (3.28) *and the conditions in Proposition* 3.3. *Let $g(n)$, ϕ, ψ be defined by* (3.45), (3.46), (3.47), *respectively.*

Then $\phi, \psi \in L^2(\mathbb{R})$, *and, for all* $j, k, k' \in \mathbb{Z}$,

$$\langle \psi_{jk}, \psi_{jk'} \rangle = \delta_{kk'}, \tag{3.50}$$

$$\langle \psi_{jk}, \phi_{jk'} \rangle = 0, \tag{3.51}$$

$$\langle \phi_{jk}, \phi_{jk'} \rangle = \delta_{kk'}. \tag{3.52}$$

Remark. Note that (3.50)–(3.52) are restricted to one j-level at a time. The orthogonality between j-levels will follow from Lemma 3.5.

Proof: 1. Let η_l be defined as in Proposition 3.3,

$$\eta_l = T^l \chi_{[-1/2,1/2[},$$

with

$$(Tf)(x) = \sqrt{2} \sum_n h(n) f(2x - n). \tag{3.53}$$

For reasons which will become obvious, we add an index 0 to η_l,

$$\eta_{l,0} = \eta_l.$$

For arbitrary $k \in \mathbb{Z}$, we define

$$\eta_{l,k} = (T_k)^l \chi_{[-1/2+k,1/2+k[}$$

with $(T_k f)(x) = \sqrt{2} \Sigma_n h(n) f(2x - n - k)$. Due to the translations over k, built into $\eta_{0,k}$ as well as into T_k, $\eta_{l,k}$ is just a translated version of $\eta_{l,0}$. This can easily be checked by induction,

$$\eta_{0,k}(x) = \chi_{[-1/2+k,1/2+k[}(x) = \eta_{0,0}(x - k)$$

and

$$\begin{aligned} \eta_{l,k}(x) &= \sqrt{2} \sum_n h(n) \eta_{l-1,k}(2x - n - k) \\ &= \sqrt{2} \sum_n h(m) \eta_{l-1,0}(2x - 2k - n) \\ &= \eta_{l,0}(x - k). \end{aligned}$$

Since (see Remark 1 following Proposition 3.3) $\|\eta_{l,0} - \phi\|_{L^2} \to 0$ for $l \to \infty$, it follows that $\|\eta_{l,k} - \phi_{0k}\|_{L^2} \to 0$ for $l \to \infty$.

2. Since $\hat{\eta}_{l,0}(\xi) = \left[\prod_{j=1}^{l} m_0(2^{-j}\xi)\right] \hat{\eta}_{0,0}(2^{-l}\xi)$, and since $|m_0(\xi)| \leqq 1$ and $\eta_{0,0} \in L^2$, it follows that all the $\eta_{l,k}$ are in L^2.

3. For fixed l, the different $\eta_{l,k}$ are orthonormal. This can again be proved by induction. By translation invariance, it is sufficient to prove that $\langle \eta_{l,k}, \eta_{l,k'} \rangle = \delta_{kk'}$ for $k' = 0$. We have

$$\langle \eta_{0,k}, \eta_{0,0} \rangle = \int_{-1/2}^{1/2} dx\, \chi_{[-1/2+k,1/2+k[}(x) = \delta_{k0}$$

and

$$\begin{aligned}\langle \eta_{l,k}, \eta_{l,0} \rangle &= 2 \sum_{n,m} h(n)h(m) \int dx\, \eta_{l-1,k}(2x - n - k)\eta_{l-1,0}(2x - m) \\ &= 2 \sum_{n,m} h(n)h(m) \int dx\, \eta_{l-1,2k+n-m}(2x)\eta_{l-1}(2x) \\ &= \sum_{n,m} h(n)h(m)\, \delta_{0,2k+n-m} = \sum_n h(n)h(m+2k) \\ &= \delta_{k,0} \qquad \text{(by (3.18))}.\end{aligned}$$

By induction it follows that $\langle \eta_{l,k}, \eta_{l,k'} \rangle = \delta_{kk'}$ for all l, k, k'.

4. It follows immediately that

$$\begin{aligned}\langle \phi_{jk}, \phi_{jk'} \rangle &= 2^{-j} \int dx\, \phi(2^{-j}x - k)\phi(2^{-j}x - k') \\ &= \int dx\, \phi(x)\phi(x - k' + k) \\ &= \lim_{l \to \infty} \langle \eta_{l,0}, \eta_{l,k'-k} \rangle = \delta_{kk'}.\end{aligned}$$

5. With $g(n)$ defined by (3.45), the conditions (3.18), (3.19) on the $h(n)$ imply (see subsection 3A)

$$\sum_n g(n - 2k)h(n - 2l) = 0, \tag{3.54}$$

$$\sum_n g(n - 2k)g(n - 2l) = \delta_{kl}. \tag{3.55}$$

Hence, by (3.48) and (3.49),

$$\begin{aligned}\langle \psi_{jk}, \phi_{jk'} \rangle &= \sum_{n,n'} g(n - 2k)h(n' - 2k')\langle \phi_{j-1\,n}, \phi_{j-1\,n'} \rangle \\ &= \sum_n g(n - 2k)h(n - 2k') = 0,\end{aligned}$$

and

$$\langle \psi_{jk}, \psi_{jk'} \rangle = \sum_{m,n'} g(m-2k)g(n'-2k')\langle \phi_{j-1\,n}, \phi_{j-1\,n'} \rangle$$

$$= \sum_{n} g(n-2k)g(n-2k') = \delta_{kk'}.$$

The "discrete orthogonality condition" (3.18) plays a crucial role in this proof. In the terminology of subsection 3A, (3.18) is equivalent to $HH^* = \mathbf{1}$, where H^* is the bounded l^2-operator (see subsection 3A)

$$(H^*a)^n = \sum_k h(n-2k)a_k.$$

This implies that H^*, as an operator from l^2 to l^2, preserves orthogonality of sequences. The operator T_H defined by (3.24) was in fact constructed to exactly reproduce, when acting on $\chi_{[-1/2,1/2[}$ and its iterates, the action of H^* on the sequence e ($e_n = \delta_{n0}$) and its iterates (see subsection 2B). This implies that repeated application of T_H preserves the orthogonality of the $\eta_{0,k}$. This is what makes the above proof work.

In the following lemma we prove that the ψ_{jk} constitute a tight frame (see Section 1, or (3.57) below). Again, the crucial ingredient will be one of the discrete identities which follow from the conditions on $h(n)$, $g(n)$. From subsection 3A we know that, with $g(n)$ as defined by (3.45), and with $h(n)$ satisfying all the conditions above,

$$\sum_k [h(n-2k)h(m-2k) + g(n-2k)g(m-2k)] = \delta_{mn}$$

(this can also be derived directly from (3.18) and (3.45)). It follows that (use (3.48), (3.49))

$$\sum_k \left[h(m-2k)\phi_{jk} + g(m-2k)\psi_{jk}\right] = \phi_{j-1\,m}. \tag{3.56}$$

This, of course, already points towards multiresolution analysis (see subsection 2A).

LEMMA 3.5. *Let $h(n)$, $g(n)$, ϕ, ψ be as in Lemma 3.4. Then, for all $f \in L^2(\mathbb{R})$,*

$$\sum_{j,k\in\mathbf{Z}} |\langle \psi_{jk}, f \rangle|^2 = \|f\|^2. \tag{3.57}$$

Proof: 1. Take any $f \in C_0^\infty$. Then, since $\phi \in L^2$, $\Sigma_n |\langle \phi_{jn}, f\rangle|^2$ converges, for any $j \in \mathbb{Z}$. Moreover, by (3.56),

$$\begin{aligned}\sum_n |\langle \phi_{j-1\,n}, f\rangle|^2 = \sum_{n,k,l} \big[& h(n-2k)h(n-2l)\langle \phi_{jk}, f\rangle\langle f, \phi_{jl}\rangle \\ & + 2h(n-2k)g(n-2l)\mathcal{R}e\big(\langle \phi_{jk}, f\rangle\langle f, \psi_{jl}\rangle\big) \\ & + g(n-2k)g(n-2l)\langle \psi_{jk}, f\rangle\langle f, \psi_{jl}\rangle\big] \\ = \sum_k \big[& |\langle \phi_{jk}, f\rangle|^2 + |\langle \psi_{jk}, f\rangle|^2\big],\end{aligned}$$

where we have used (3.18), (3.54) and (3.55).

2. By iteration, one has, for all $N \in \mathbb{N}$,

$$\sum_n |\langle \phi_{-Nn}, f\rangle|^2 = \sum_k |\langle \phi_{Nk}, f\rangle|^2 + \sum_{j=-N}^{N} \sum_k |\langle \psi_{jk}, f\rangle|^2. \tag{3.58}$$

In this expression we shall let N tend to ∞.

3. We first concentrate on $\Sigma_k |\langle \phi_{Nk}, f\rangle|^2$. Let us suppose, for the sake of definiteness, that $\operatorname{supp} f \subset [-2^{n_0}, 2^{n_0}]$. Take $N \geqq n_0 + 1$, so that the translation steps in the $\phi_{Nk}(x) = \phi_{N0}(x - 2^N k)$ are larger than $|\operatorname{supp} f|$. On the other hand, for any $\varepsilon > 0$ there exists $k_0 \in \mathbb{N}$ such that

$$\int_{|x| \geqq k_0} dx |\phi(x)|^2 \leqq \varepsilon.$$

Then

$$\begin{aligned}\sum_k |\langle \phi_{Nk}, f\rangle|^2 & \\ = \sum_{|k| \leqq k_0} |\langle \phi_{Nk}, f\rangle|^2 & + \sum_{|k| \geqq k_0+1} |\langle \phi_{Nk}, f\rangle|^2 \\ \leqq (2k_0+1)2^{-N}\|\phi\|_\infty^2 \|f\|_1^2 & + \|f\|_2^2 2^{-N} \sum_{|k| \geqq k_0+1} \int_{|x| \leqq 2^{n_0}} dx |\phi(2^{-N}x - k)|^2 \\ \leqq 2^{-N}(2k_0+1)\|\phi\|_\infty^2 \|f\|_1^2 & + \varepsilon \|f\|_2^2.\end{aligned}$$

By choosing ε and N appropriately, this can be made arbitrarily small. Hence

$$\sum_k |\langle \phi_{Nk}, f\rangle|^2 \xrightarrow[N \to \infty]{} 0. \tag{3.59}$$

4. We now concentrate on $\Sigma_k|\langle\phi_{-Nk}, f\rangle|^2$. By means of the Poisson formula this can be rewritten as

$$\sum_k |\langle\phi_{-Nk}, f\rangle|^2$$

$$= 2\pi \sum_{l\in\mathbf{Z}} \int d\xi\, \hat{\phi}(2^{-N}\xi)\overline{\hat{\phi}(2^{-N}\xi + 2\pi l)}\,\overline{\hat{f}(\xi)}\,\hat{f}(\xi + 2\pi l 2^N)$$

$$= 2\pi \int d\xi |\hat{\phi}(2^{-N}\xi)|^2 |\hat{f}(\xi)|^2 + R. \tag{3.60}$$

Here

$$|R| \leqq \sum_{l\neq 0} \int d\xi\, |\hat{f}(\xi)|\,|\hat{f}(\xi + 2\pi l 2^N)|,$$

because $|\hat{\phi}(\xi)| = (2\pi)^{-1/2}\prod_{j=1}^{\infty}|m_0(2^{-j}\xi)| \leqq (2\pi)^{-1/2}$, since $|m_0(\xi)| \leqq 1$. Since $f \in C_0^\infty$, we can find C such that

$$|\hat{f}(\xi)| \leqq C(1 + |\xi|)^{-3}.$$

An easy estimation then leads to

$$|R| \leqq C' 2^{-3N/2}.$$

This tends to zero for $N \to \infty$.

5. We now examine the first term in (3.60). One has

$$|\hat{\phi}(\xi) - \hat{\phi}(0)| = (2\pi)^{-1/2}\left|\prod_{j=1}^{\infty} m_0(2^{-j}\xi) - \prod_{j=1}^{\infty} m_0(0)\right|$$

$$\leqq (2\pi)^{-1/2} \sum_{j=1}^{\infty} |m_0(2^{-j}\xi) - m_0(0)|,$$

since $|m_0(\zeta)| \leqq 1$ for all $\zeta \in \mathbf{R}$. But

$$|m_0(\zeta) - m_0(0)| \leqq 2^{-1/2} \sum_n |h(n)|\,|e^{in\zeta} - 1|$$

$$\leqq C|\zeta|^\varepsilon,$$

where we have used (3.19) and $|e^{i\alpha} - 1| \leqq C_\varepsilon|\alpha|^\varepsilon$ (we assume $0 < \varepsilon \leqq 1$). Hence,

$$|\hat{\phi}(\xi) - \hat{\phi}(0)| \leqq (2\pi)^{-1/2} C \sum_{j=1}^{\infty} |2^{-j}\xi|^\varepsilon \leqq C'|\xi|^\varepsilon.$$

Consequently, using $\hat{\phi}(0) = (2\pi)^{-1/2}$, we find

$$2\pi \int d\xi \, |\hat{\phi}(2^{-N}\xi)|^2 |\hat{f}(\xi)|^2$$

$$\leqq \int d\xi \, |\hat{f}(\xi)|^2 + 2\pi \int d\xi \left[2C'|2^{-N}\xi|^{\varepsilon} + C'^2|2^{-N}\xi|^{2\varepsilon}\right] |\hat{f}(\xi)|^2$$

$$= \|f\|^2 + C''2^{-N\varepsilon} \int d\xi (1 + |\xi|^{2\varepsilon}) |\hat{f}(\xi)|^2.$$

This converges to $\|f\|^2$ as $N \to \infty$. Hence

$$\sum_k |\langle \phi_{-Nk}, f \rangle|^2 \xrightarrow[N \to \infty]{} \|f\|^2. \tag{3.61}$$

6. Putting together (3.58), (3.60) and (3.61) shows that, for all $f \in C_0^{\infty}(\mathbb{R})$,

$$\sum_{j,k} |\langle \psi_{jk}, f \rangle|^2 = \|f\|_{L^2}^2. \tag{3.62}$$

Since $C_0^{\infty}(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, (3.62) extends to all $f \in L^2(\mathbb{R})$.

Since $\|\psi\| = 1$ (this is a special case of (3.50), with $j = k = k' = 0$), (3.57) implies that the ψ_{jk} constitute an orthonormal basis. This completes the proof of the main theorem of this section.

THEOREM 3.6. *Let $h(n)$ be a sequence such that*

(i) $\Sigma_n |h(n)| \, |n|^{\varepsilon} < \infty$ *for some* $\varepsilon > 0$,

(ii) $\Sigma_n h(n - 2k)h(n - 2l) = \delta_{kl}$,

(iii) $\Sigma h(n) = 2^{1/2}$.

Suppose also that $m_0(\xi) = 2^{-1/2}\Sigma_n h(n)e^{in\xi}$ can be written as

$$m_0(\xi) = \left[\tfrac{1}{2}(1 + e^{i\xi})\right]^N \left[\sum_n f(n)e^{in\xi}\right],$$

where

(iv) $\Sigma_n |f(n)| \, |n|^{\varepsilon} < \infty$ *for some* $\varepsilon > 0$,

(v) $\sup_{\xi \in \mathbb{R}} |\Sigma_n f(n)e^{in\xi}| < 2^{N-1}$.

Define

$$g(n) = (-1)^n h(-n + 1),$$

$$\hat{\phi}(\xi) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} m_0(2^{-j}\xi),$$

$$\psi(x) = 2^{1/2} \sum_n g(n)\phi(2x - n).$$

Then the $\phi_{jk}(x) = 2^{-j/2}\phi(2^{-j}x - k)$ *define a multiresolution analysis (in the sense of subsection* 2A); *the* ψ_{jk} *are the associated orthonormal wavelet basis.*

Remarks. 1. As we already said in the introduction, this theorem is also proved in [19], under slightly different conditions. The growth restrictions (3.37) and (3.38) on the $h(n)$ are replaced, in [19], by the condition that $\inf_{|\xi| \le \pi/2}|m_0(\xi)| > 0$. Together with $|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1$, $m_0(0) = 1$, this condition implies that the ϕ_{jk}, with ϕ defined as above, define a multiresolution analysis. The function ϕ may, however, still be very irregular; the coefficients $h(n)$ used in Figure 4, e.g., satisfy the positivity condition of [19], but are clearly not associated with a regular ϕ. In the present paper, we emphasized regularity of the discrete filters; once regularity is ensured by means of conditions (3.36)–(3.38), equivalence with regular multiresolution analysis follows. Consequently, the techniques of our proofs and the proofs in [19] are quite different. The basic intuition for the present proof was mainly graphical. As explained above, the orthogonality of the ϕ_{0k} follows naturally, given our "graphical" construction, from the discrete conditions. Similarly, (3.60) can be understood graphically.

2. At the end of subsection 3B (Remark 3) we mentioned the link between the present construction and the "conjugated quadrature filters" of Smith and Barnwell [24]. Any of their conjugated quadrature filters will satisfy all the conditions in subsection 3A. Provided they also satisfy the regularity condition in subsection 3B, they can be used to construct orthonormal wavelet bases. Since the goals of [24] are completely different however, most of the examples in [24] do not satisfy our regularity condition.

4. Orthonormal Bases of Wavelets with Compact Support

In subsection 2A we reviewed how orthonormal bases of wavelets can be constructed, starting from a multiresolution analysis framework. The basic ingredient there was a function ϕ such that (2.15) held, for some c_n, without even requiring the ϕ_{0n} to be orthogonal. Theorem 3.6 gives another recipe for constructing an orthonormal basis of wavelets (and the associated multiresolution analysis), this time from a sequence $(h(n))_{n \in \mathbf{Z}}$.

If this sequence has finite length, $h(n) = 0$ for $n < N_-$, or $n > N_+$, then the corresponding basic wavelet has compact support. This can be checked very easily from the graphical construction of ϕ (see Figures 2, 4), or from the recursive definition of the η_l,

$$\phi(x) = \lim_{l \to \infty} \eta_l(x), \tag{4.1}$$

$$\eta_l(x) = \sqrt{2} \sum_n h(n)\eta_{l-1}(2x - n), \tag{4.2}$$

$$\eta_0 = \chi_{[-1/2,1/2[}. \tag{4.3}$$

The recursive definition of the η_l implies that all the η_l have compact support, supp $\eta_l \subset [N_{l,-}, N_{l,+}]$, with $N_{l,-} = \frac{1}{2}(N_{l-1,-} + N_-)$, and $N_{l,+} = \frac{1}{2}(N_{l-1,+} + N_+)$, while $N_{0,-} = -\frac{1}{2}$, $N_{0,+} = \frac{1}{2}$. Hence $H_{l,-} \to N_-$, $N_{l,+} \to N_+$ for $l \to \infty$, which implies that ϕ has compact support $\subset [N_-, N_+]$. Since only finitely many $g(n)$ are non-zero ($g(n) = 0$ for $n < -N_+ + 1$ or $n > -N_- + 1$), ψ also has compact support,

$$\text{supp } \psi \subset \left[\tfrac{1}{2}(1 - N_+ - N_-), \tfrac{1}{2}(1 + N_+ - N_-)\right].$$

In order to construct orthonormal bases of compactly supported wavelets, it suffices, therefore, to construct finite-length sequences $h(n)$ satisfying all the conditions of Theorem 3.6. An example of such a finite-length sequence is Example 3.2, with $\nu = \pm 1/\sqrt{3}$ (see Remark 5 following Proposition 3.3). In this case one finds (see (3.42)) $N_- = 0$, $N_+ = 3$, and

$$m_0(\xi) = \left[\tfrac{1}{2}(1 + e^{i\xi})\right]^2 \tfrac{1}{2}\left[(1 \mp \sqrt{3}) + (1 \pm \sqrt{3})e^{i\xi}\right]. \tag{4.4}$$

Since

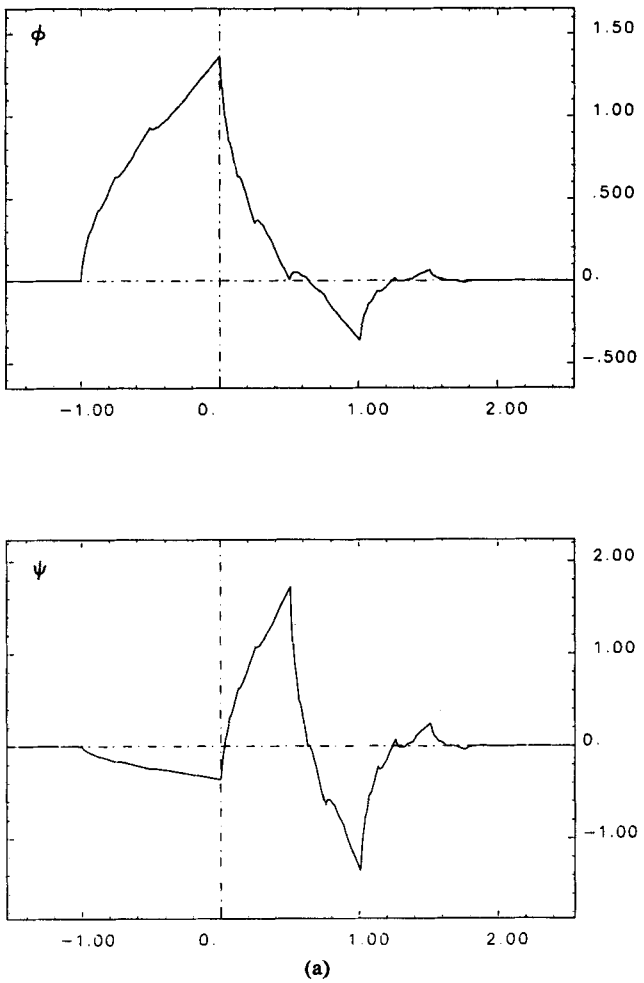
$$\sup_{\xi \in \mathbf{R}} \tfrac{1}{2}\left|(1 \mp \sqrt{3}) + (1 \pm \sqrt{3})e^{i\xi}\right| = \sqrt{3} < 2,$$

the example (3.42) satisfies all the required conditions. The $h(n)$ given by (3.42) correspond, therefore, to an orthonormal basis of continuous wavelets. The basic wavelet has support width equal to $N_+ - N_- = 3$. Figure 5 shows the graphs of ϕ, ψ and their Fourier transforms, for this example. There are several striking features in Figure 5. First of all, it is obvious that even though ϕ and ψ are continuous, they are not very regular. There exist other constructions of compactly supported wavelet bases, in which ϕ and ψ have more regularity, at the cost of larger numbers of non-zero coefficients $h(n)$, which results in larger support widths for ψ, ϕ. For the family of examples we shall examine below, the support width of ψ, ϕ increases linearly with their regularity. Another striking feature of Figure 5 is the lack of any symmetry or antisymmetry axis for ψ, ϕ. This is quite unlike the Meyer wavelets (see [4]) or the Battle-Lemarié wavelets (see [16]). In all these (non-compactly supported) examples, ϕ is an even function, and ψ is symmetric around $x = \frac{1}{2}$. We shall see below that, except for the Haar basis (see (1.9) or Example 3.1), there exist *no* compactly supported wavelet bases in which ϕ is either symmetric or antisymmetric around any axis.

The plots of ψ and ϕ in Figure 5 (and later figures, for other examples) are made by direct implementation of the "graphical recursion algorithm" equivalent with (4.1)–(4.3) (see subsection 2B). This is much more efficient than Fourier transform of the infinite product (3.46) (see [26]). To plot Figure 5, only 8 iterations of type (2.40) were needed (i.e., η_8 is plotted rather than ϕ; the

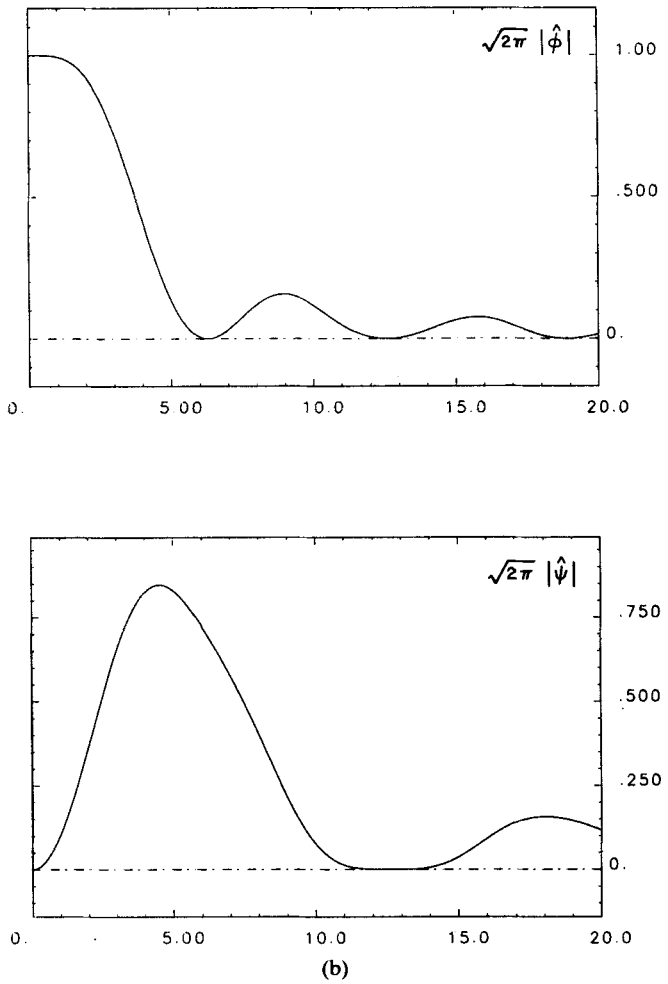
difference is not detectable at the scale of the figure). If more detail is wanted at any point (see Figure 6), it is possible to restrict to a neighborhood, and to locally iterate a few times more to obtain this detail.

In the following subsections we describe families of examples of compactly supported wavelet bases, and their properties. Henceforth, we shall always assume that only finitely many $h(n)$ are non-zero.



(a)

Figure 5. The functions ϕ, ψ, and the modulus of their Fourier transforms, $|\hat{\phi}|$, $|\hat{\psi}|$, for the orthonormal basis of compactly supported wavelets corresponding to the $h(n)$ in (3.42) (see text). Out of the two possibilities in (3.42) we choose the one corresponding to $\nu = -1/\sqrt{3}$ (i.e., $h(0) = (1 + \sqrt{3})/4\sqrt{2}$, etc.)



(b)

Figure 5. Continued

4.A. Lack of symmetry. Here we shall use again the notations $a(n),\cdots,$ $d(n)$ (see (3.7)) and $\alpha(\xi),\cdots,\delta(\xi)$ (see (3.9)) introduced in subsection 3A. Let us define, for any trigonometric polynomial $P(\xi) = \Sigma_n p_n e^{in\xi}$, the two numbers

$$N_+(P) = \max\{n;\, p_n \neq 0\},$$

$$N_-(P) = \min\{n;\, p_n \neq 0\}.$$

One easily checks that

$$N_+\big(|P|^2\big) = -N_-\big(|P|^2\big) = N_+(P) - N_-(P).$$

Since $|\alpha(\xi)|^2 + |\beta(\xi)|^2 = 1$ (see (3.14)), and $\alpha \neq 0$, $\beta \neq 0$ (see (3.15)), this implies

$$N_+(\alpha) - N_-(\alpha) = N_+(\beta) - N_-(\beta). \tag{4.5}$$

On the other hand, the definition (3.7) of the $a(n)$, $b(n)$, gives

$$N_+(m_0) = \max(2N_+(\alpha), 2N_+(\beta) + 1),$$

$$N_-(m_0) = \min(2N_-(\alpha), 2N_-(\beta) + 1).$$

Together with (4.5) this leads to

$$\begin{aligned} &N_+(m_0) - N_-(m_0) \\ &\quad = \max(2N_+(\alpha) - 2N_-(\beta) - 1, 2N_+(\beta) - 2N_-(\alpha) + 1). \end{aligned} \tag{4.6}$$

In any case, $N_+(m_0) - N_-(m_0)$ is an odd number.

If the function ϕ were symmetric around zero, $\phi(x) = \phi(-x)$, then $h(n) = h(-n)$ would follow. This would however imply $N_+(m_0) = -N_-(m_0)$, i.e., $N_+(m_0) - N_-(m_0) = 2N_+(m_0)$ would be even. Since this is in contradiction with (4.6), it follows that the function ϕ, associated with an orthonormal basis of wavelets with compact support, can never be an even function.

What about symmetry with respect to another point $\lambda \neq 0$? Suppose

$$\phi(\lambda + x) = \phi(\lambda - x),$$

where we can, without loss of generality, shift λ to the interval $]0, 1[$. Then it follows that

$$\hat{\phi}(\xi) = e^{2i\lambda\xi}\hat{\phi}(-\xi).$$

Because of the definition of $\hat{\phi}$ as the infinite product (3.46), this implies

$$m_0(\xi) = e^{2i\lambda\xi}m_0(-\xi).$$

Since both $m_0(\xi)$ and $m_0(-\xi)$ are trigonometric polynomials, this leaves only one possible value for λ, namely $\lambda = \frac{1}{2}$. Let us, therefore, assume that ϕ is symmetric with respect to $\frac{1}{2}$,

$$\phi(x + 1) = \phi(-x).$$

Then

$$h(2n + 1) = h(-2n),$$

or

$$b(n) = a(-n).$$

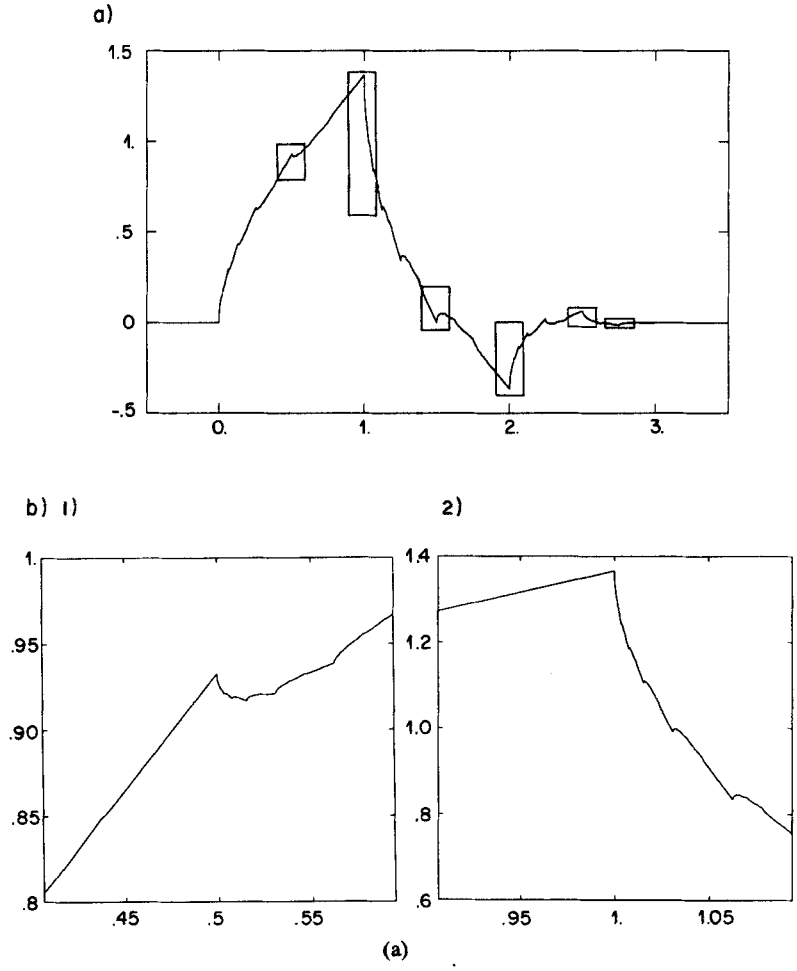


Figure 6. The function ϕ of Figure 5, and 6 local blow-ups
(a) The different zoom-in zones are shown on the graph of ϕ
(b) The blow-ups around 1) $x = .5$, 2) $x = 1$, 3) $x = 1.5$, 4) $x = 2.$, 5) $x = 2.5$, 6) $x + 2.75$.
The detail in these blow-ups illustrates the fractal, self-similar nature of this function ϕ.

Hence

$$\beta(\xi) = \overline{\alpha(\xi)}\,.$$

Together with (3.14) this implies

$$2|\alpha(\xi)|^2 = 1,$$

or

$$a(n) = \pm 2^{-1/2}\delta_{nk} = b(-n) \quad \text{for some} \quad k \in \mathbb{N}.$$

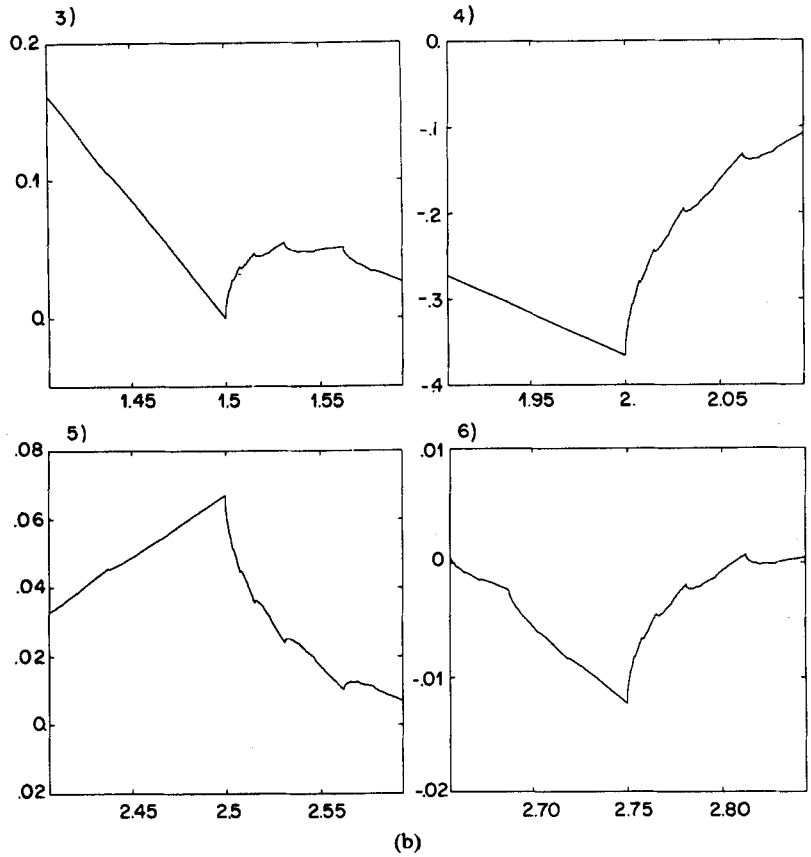


Figure 6. Continued

We can, again without loss of generality, choose $k = 0$ (this amounts to a translation of ϕ by an integer). The corresponding $h(n)$ are then exactly given by Example 3.1, resulting in $\phi = \chi_{[0,1]}$.

All these arguments prove the following proposition.

PROPOSITION 4.1. *The Haar basis* (1.9) *is the only orthonormal basis of compactly supported wavelets for which the associated averaging function* ϕ *has a symmetry axis.*

In the following subsection we explicitly characterize all the functions m_0 corresponding to orthonormal wavelet bases with compactly supported basic wavelet.

4.B. Characterization of all orthonormal, compactly supported wavelet bases. The basic condition (3.18) on the $h(n)$ can be rewritten as (see subsection 3A)

$$|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1. \tag{4.7}$$

On the other hand, we have imposed, in Proposition 3.3, the following structure on m_0:

$$m_0(\xi) = \left[\tfrac{1}{2}(1 + e^{i\xi})\right]^N Q(e^{i\xi}), \tag{4.8}$$

where Q is a polynomial, since only finitely many $h(n)$ are non zero. Moreover, since all the $h(n)$ are real, all the coefficients in Q are real as well. From (4.8) we have

$$|m_0(\xi)|^2 = \left[\cos^2\tfrac{1}{2}\xi\right]^N |Q(e^{i\xi})|^2.$$

Since $\overline{Q(e^{i\xi})} = Q(e^{-i\xi})$, the polynomial $|Q(e^{i\xi})|^2$ can be rewritten as a polynomial in $\cos\xi$, or, equivalently, as a polynomial in $\sin^2\frac{1}{2}\xi$. Introducing the shorthand $y = \cos^2\frac{1}{2}\xi$, (4.7) becomes

$$y^N P(1-y) + (1-y)^N P(y) = 1. \tag{4.9}$$

Any m_0 of type (4.8) which solves (4.7) corresponds therefore to a polynomial P solving (4.9) and satisfying

$$P(y) \geqq 0 \quad \text{for} \quad y \in [0,1]. \tag{4.10}$$

Conversely, every polynomial P satisfying both (4.9) and (4.10) leads to solutions of (4.7), with real coefficients $h(n)$.This is due to the following lemma of Riesz [27].

LEMMA 4.2. *Let A be a positive trigonometric polynomial containing only cosines, $A(\xi) = \sum_{n=0}^{N} a_n \cos n\xi$ (with $a_n \in \mathbb{R}$). Then there exists a trigonometric polynomial B, of order N, $B(\xi) = \sum_{n=0}^{N} b_n e^{in\xi}$, with real coefficients b_n, such that*

$$|B(\xi)|^2 = A(\xi). \tag{4.11}$$

The proof of this lemma (see [27]) is simple and elegant. It constructs B explicitly; this construction is now widely used by engineers when designing filters. We include the proof here, because we shall come back to the construction later.

Proof: To

$$A(\xi) = a_0 + \frac{1}{2}\sum_{n=1}^{N} a_n(e^{in\xi} + e^{-in\xi})$$

$$= e^{-iN\xi}\left[\frac{1}{2}\sum_{n=0}^{N-1} a_{N-n}\, e^{in\xi} + a_0\, e^{iN\xi} + \frac{1}{2}\sum_{n=1}^{N} a_n\, e^{i(N+n)\xi}\right]$$

we associate the polynomial

$$P_A(z) = \frac{1}{2}\sum_{n=0}^{N-1} a_{N-n}z^n + a_0 z^N + \frac{1}{2}\sum_{n=1}^{N} a_n z^{N+n}.$$

This polynomial has $2N$ zeros (counting multiplicity). Since $P_A(e^{i\xi}) = e^{iN\xi}A(\xi)$, it follows that the two polynomials $P_A(z)$ and $z^{2N}P_A(z^{-1})$ agree on the unit circle, and therefore on the whole complex plane. They have therefore the same zeros. This means that if z_0 is a zero of $P_A(z)$, $P_A(z_0) = 0$, then so is z_0^{-1}. On the other hand, since the a_n are real, $P_A(z) = P_A(\bar{z})$. This implies that if z_0 is a zero of $P_A(z)$, then so is its complex conjugate $\bar{z}_0$. The zeros of $P_A(z)$ therefore come in quadruplets, z_0, $\bar{z}_0$, z_0^{-1} and $\bar{z}_0^{-1}$, or (if $z_0 = r_0$ is real) in duplets, r_0, $r_0^{-1} \in \mathbb{R}$. Let z_j, $\bar{z}_j$, z_j^{-1}, $\bar{z}_j^{-1}$ be the quadruplets of complex zeros of $P_A(z)$, and r_k, r_k^{-1} the real duplets,

$$P_A(z) = \frac{1}{2}a_N\left[\prod_{k=1}^{K}(z - r_k)\left(z - r_k^{-1}\right)\right]$$

$$\cdot\left[\prod_{j=1}^{J}(z - z_j)(z - \bar{z}_j)\left(z - z_j^{-1}\right)\left(z - \bar{z}_j^{-1}\right)\right].$$

For $z = e^{i\xi}$ on the unit circle, one finds

$$\left|\left(e^{i\xi} - z_0\right)\left(e^{i\xi} - \bar{z}_0^{-1}\right)\right| = |z_0|^{-1}\left|\left(e^{i\xi} - z_0\right)\left(\bar{z}_0 - e^{-i\xi}\right)\right|$$

$$= |z_0|^{-1}|e^{i\xi} - z_0|^2.$$

Consequently,

$$A(\xi) = |A(\xi)| = \left|P_A(e^{i\xi})\right|$$

$$= \left[\frac{1}{2}|a_N|\prod_{k=1}^{K}|r_k|^{-1}\prod_{j=1}^{J}|z_j|^{-2}\right]\left|\prod_{k=1}^{K}\left(e^{i\xi} - r_k\right)\prod_{j=1}^{J}\left(e^{i\xi} - z_j\right)\left(e^{i\xi} - \bar{z}_j\right)\right|^2$$

$$= |B(\xi)|^2,$$

where

$$B(\xi) = \left[\frac{1}{2}|a_N|\prod_{k=1}^{K}|r_k|^{-1}\prod_{j=1}^{J}|z_j|^{-2}\right]^{1/2}$$

$$\cdot\prod_{k=1}^{K}\left(e^{i\xi} - r_k\right)\prod_{j=1}^{J}\left(e^{2i\xi} - 2e^{i\xi}\,\mathrm{Re}\, z_j + |z_j|^2\right) \tag{4.12}$$

is clearly a trigonometric polynomial of order N with only real coefficients.

Remarks. 1. Note that B is generally not unique. Out of any quadruplet of zeros z_0, $\bar{z}_0$, z_0^{-1}, $\bar{z}_0^{-1}$ one can choose the pair of zeros to retain, for the construction of B, in four different ways. For every duplet of real zeros of P_A two choices are possible. This results in 2^N different possibilities for B.

2. All these different possibilities, corresponding to different choices of the zeros of P_A to retain for B, constitute, however, the only solutions to (4.11). One can show (see [27]) that, up to an arbitrary phase factor $\pm e^{iK\xi}$, $K \in \mathbb{Z}$, *all* the polynomials B satisfying (4.11) are necessarily of the form (4.12).

If P is a polynomial satisfying (4.9) and (4.10), then Lemma 4.2 tells us that there exists a trigonometric polynomial of the same order such that

$$|Q(e^{i\xi})|^2 = P(\sin^2 \tfrac{1}{2}\xi) = P(\tfrac{1}{2}(1 - \cos \xi)).$$

It follows that $m_0(\xi) = [\frac{1}{2}(1 + e^{i\xi})]^N Q(e^{i\xi})$ satisfies (4.7). If, moreover,

$$\sup_{\xi} |Q(e^{i\xi})| = \sup_{y \in [0,1]} |P(y)|^{1/2} < 2^{N-1},$$

then all the conditions of Theorem 3.6 are satisfied, and there exists an associated orthonormal wavelet basis.

To construct compactly supported orthonormal wavelet bases, with m_0 of type (4.8), it is therefore necessary and sufficient to find polynomials P solving (4.9) and (4.10), which are moreover strictly bounded above by $2^{2(N-1)}$.

The following two combinatorial lemmas allow one to "guess" a particular solution of (4.9).

LEMMA 4.3.

$$\sum_{j=0}^{k} \binom{n+j}{j} = \binom{n+k+1}{k}.$$

Proof: Define $S_{n,k} = \sum_{j=0}^{k} \binom{n+j}{j}$. Then

$$S_{n+1,k+1} = \frac{(k+n+2)!}{(k+1)!(n+1)!} + \sum_{j=0}^{k} \frac{(n+j)!}{(n+1)!j!}(n+j+1)$$

$$= \binom{k+n+2}{k+1} + S_{n,k} + \sum_{j=1}^{k} \frac{(n+j)!}{(n+1)!(j-1)!}$$

$$= \binom{k+n+2}{n+1} + S_{n,k} + \left[S_{n+1,k+1} - \binom{k+n+2}{k+1} - \binom{k+n+1}{k}\right].$$

Hence

$$S_{n,k} = \binom{n+k+1}{k}.$$

LEMMA 4.4.

$$\sum_{j=0}^{n} \binom{n+j}{j} \left[y^j (1-y)^{n+1} + y^{n+1}(1-y)^j \right] = 1.$$

Proof: Define $A_{n,j} = \binom{n+j}{j}$. Then, by Lemma 4.3, $\sum_{j=0}^{k} A_{n,j} = A_{n+1,k}$. Define

$$S_n(y) = \sum_{j=0}^{n} \binom{n+j}{j} \left[y^j (1-y)^{n+1} + y^{n+1}(1-y)^j \right].$$

Clearly,

$$S_0(a) = (1-a) + a = 1.$$

We shall prove that $S_n(a) = S_{n-1}(a)$, which proves the lemma. By repeatedly inserting factors $[(1-a)+a] = 1$, we find

$$\begin{aligned}
S_{n-1}(a) &= \sum_{j=0}^{n-1} A_{n-1,j} \left[(1-a)^n a^j + a^n (1-a)^j \right] \\
&= A_{n-1,0} \left[(1-a)^{n+1} + a^{n+1} \right] \\
&\quad + (A_{n-1,0} + A_{n-1,1}) \left[(1-a)^n a + a^n (1-a) \right] \\
&\quad + \sum_{j=2}^{n-1} A_{n-1,j} \left[(1-a)^n a^j + a^n (1-a)^j \right] \\
&= \cdots \\
&= \sum_{j=0}^{n-1} \left[\sum_{k=0}^{j} A_{n-1,k} \right] \left[(1-a)^{n+1} a^j + a^{n+1} (1-a)^j \right] \\
&\quad + 2 \left[\sum_{k=0}^{n-1} A_{n-1,k} \right] (1-a)^n a^n \\
&= \sum_{j=0}^{n-1} A_{n,j} \left[(1-a)^{n+1} a^j + a^{n+1} (1-a)^j \right] \\
&\quad + 2 A_{n,n-1} \left[(1-a)^{n+1} a^n + a^{n+1} (1-a)^n \right] \\
&= S_n(a) \qquad \left(\text{since } 2A_{n,n-1} = A_{n,n} \right).
\end{aligned}$$

It follows that the polynomial of order $N-1$,

$$P_N(y) = \sum_{j=0}^{N-1} \binom{N-1+j}{j} y^j, \tag{4.13}$$

solves (4.9). Since all the coefficients in this polynomial are positive, (4.10) is clearly also satisfied.

The two explicit examples of compactly supported wavelet bases we have seen so far, i.e., Example 3.1 and (3.42), correspond exactly to a polynomial of type (4.13), with $N = 1, 2$, respectively. For Example 3.1 one has $m_0(\xi) = \frac{1}{2}(1 + e^{i\xi})$, i.e., $N = 1$, and $Q(e^{i\xi}) = 1$, hence $P(y) = 1 = P_1(y)$. For the second example (3.42), we find (see (4.4)) $m_0(\xi) = [\frac{1}{2}(1 + e^{i\xi})]^2 \frac{1}{2}[(1 \mp \sqrt{3})e^{i\xi}]$, corresponding to $N = 2$ and $|Q(e^{i\xi})|^2 = 2 - \cos\xi = 1 + 2\sin^2\frac{1}{2}\xi$; hence $P(y) = 1 + 2y = P_2(y)$.

In fact, for given N, P_N is the *only* polynomial of order less than N which solves (4.9). Even more is true: for *any* polynomial P solving (4.9), the first N terms (orders 0 up till $N-1$) are exactly given by P_N. This is because (4.9) completely determines the first N coefficients $p_0, \cdots, p_{N-1}$ in $P(y) = \sum_{n=0}^{K} p_n y^n$. Since the first term in (4.9) is already of order N, only the second term plays a role in the cancellations for y^k, $k = 0, \cdots, N-1$. This leads to

$$\begin{aligned} p_0 &= 1, \\ p_k &= \sum_{n=0}^{k-1} (-1)^{k-n+1} \binom{N}{k-n} p_n, \qquad k = 1, \cdots, N-1, \end{aligned} \tag{4.14}$$

from which the p_k, $k = 1, \cdots, N-1$, can be determined recursively. Since P_N solves (4.9), it follows from (4.13) that

$$p_k = \binom{N+k-1}{k}.$$

Consequently, *any* polynomial P solving (4.9) is of the form

$$P(y) = P_N(y) + y^N R(y). \tag{4.15}$$

Substitution of (4.15) into (4.9) leads to the following equation for the polynomial R:

$$y^N(1-y)^N R(1-y) + (1-y)^N y^N R(y) = 0,$$

or

$$R(1-y) + R(y) = 0.$$

The polynomial R is therefore antisymmetric with respect to $y = \frac{1}{2}$, or

$$R(y) = \tilde{R}(\tfrac{1}{2} - y),$$

where $\tilde{R}$ is an odd polynomial.

To summarize, we have the following explicit characterization of all solutions m_0 of (4.7), corresponding to only finitely many non-zero $h(n)$.

PROPOSITION 4.5. *Any trigonometric polynomial solution m_0 of (4.7) is of the form*

$$m_0(\xi) = \left[\tfrac{1}{2}(1 + e^{i\xi})\right]^N Q(e^{i\xi}), \tag{4.16}$$

where $N \in \mathbb{N}$, $N \geq 1$, and where Q is a polynomial such that

$$|Q(e^{i\xi})|^2 = \sum_{k=0}^{N-1} \binom{N-1+k}{k} \sin^{2k}\tfrac{1}{2}\xi + \left[\sin^{2N}\tfrac{1}{2}\xi\right] R\left(\tfrac{1}{2}\cos\xi\right), \tag{4.17}$$

where R is an odd polynomial.

Remarks. 1. Since the proof of Lemma 4.2 shows explicitly how to construct all possible polynomials Q once $|Q(e^{i\xi})|^2$ is known, this proposition is indeed an explicit characterization of all the solutions m_0 of (4.7).

2. In constructing m_0, there are therefore 3 steps at which choices can be made,

(i) choosing $N \in \mathbb{N} \setminus \{0\}$,
(ii) choosing an odd polynomial R (with some restrictions),
(iii) choosing pairs of zeros out of each quadruplet of complex zeros, and one zero out of each duplet of real zeros, of $P_N(z) + z^N R(z - \frac{1}{2})$ (see the proof of Lemma 4.2).

The odd polynomial R cannot be chosen completely freely. One needs, of course, the fact that

$$P_N(y) + y^N R\left(\tfrac{1}{2} - y\right) \geq 0 \quad \text{for} \quad 0 \leq y \leq 1. \tag{4.18}$$

Moreover, condition (v) in Theorem 3.6 requires that

$$\sup_{0 \leq y \leq 1} \left[P_N(y) + y^N R\left(\tfrac{1}{2} - y\right)\right] < 2^{2(N-1)}. \tag{4.19}$$

3. For $N = 1$, (4.16), (4.17) and (4.18) reduce to

$$m_0(\xi) = \tfrac{1}{2}(1 + e^{i\xi}) Q(e^{i\xi}) \tag{4.20}$$

with

$$\left|Q(e^{i\xi}\right|^2 = 1 + \sin^2\tfrac{1}{2}\xi R\left(\tfrac{1}{2}\cos\xi\right), \tag{4.21}$$

where R is an odd polynomial such that

$$-\frac{2}{1 - 2|x|} \leq R(x) \leq \frac{2}{1 + 2|x|} \quad \text{for} \quad |x| \leq \tfrac{1}{2}.$$

These conditions can already be found in the construction of conjugate quadrature mirror filters in [24]. The condition (4.19) is impossible to satisfy, however, because $P_1(0) = 1$.

4. Using a different method, Y. Meyer constructs in [28] another polynomial solving (4.7). The solutions to (4.7) proposed in [28] are

$$|m_0(\xi)|^2 = 1 - \frac{(2N-1)!}{[(N-1)!]^2 2^{2N-1}} \int_0^\xi \sin^{2N-1} x \, dx. \tag{4.22}$$

This is clearly an even trigonometric polynomial of order $2N-1$. It turns out to be divisible by $(\frac{1}{2}(1+\cos\xi))^N = (\cos^2\frac{1}{2}\xi)^N$. Therefore, by Proposition 4.5, (4.22) is exactly equal to

$$\left(\cos^2\tfrac{1}{2}\xi\right)^N P_N\left(\sin^2\tfrac{1}{2}\xi\right).$$

4.C. A family of examples with arbitrarily high regularity. In the remainder of this section, we shall concern ourselves with a special family of functions m_0, and the corresponding wavelet bases. We follow the prescriptions of Remark 2 after Proposition 4.5. For every $N \in \mathbb{N}$, $N \geq 1$, we choose Q of minimal order, i.e., $R \equiv 0$, $|Q(e^{i\xi})|^2 = P_N(\sin^2\frac{1}{2}\xi)$. This choice satisfies both the conditions (4.18) and (4.19). From (4.13) the positivity of $P_N(y)$ for $0 \leq y \leq 1$ is immediate. Since P_N is strictly increasing for $y \geq 0$, it follows that

$$\begin{aligned} \sup_{y\in[0,1]} P_N(y) = P_N(1) &= \binom{2N-1}{N-1} = \frac{1}{2}\left[\binom{2N-1}{N-1} + \binom{2N-1}{N}\right] \\ &< \frac{1}{2}\sum_{k=0}^{2N-1}\binom{2N-1}{k} = 2^{2(N-1)}, \end{aligned} \tag{4.23}$$

where we have used Lemma 4.3 in the second equality. This fixes $|Q|^2$. In the construction (via Lemma 4.2) of Q from $|Q|^2$, we systematically retain all the zeros inside the unit circle (this corresponds to a "minimal phase" choice in filter design). For $N \in \mathbb{N}$, $N > 1$ fixed, this determines Q unambiguously, up to a phase factor $e^{iK\xi}$, $K \in \mathbb{Z}$. For the sake of definiteness we fix this phase factor so that Q contains only positive frequencies, starting from zero, i.e.,

$$Q_N(e^{i\xi}) = \sum_{n=0}^{N-1} q_N(n) e^{in\xi} \quad \text{with} \quad q_0 \neq 0. \tag{4.24}$$

These choices uniquely determine Q_N. We shall denote the corresponding m_0 by ${}_N m_0$,

$$\begin{aligned} {}_N m_0(\xi) &= \left[\tfrac{1}{2}(1+e^{i\xi})\right]^N \sum_{n=0}^{N-1} q_N(n) e^{in\xi} \\ &= 2^{-1/2} \sum_{n=0}^{2N-1} h_N(n) e^{in\xi}. \end{aligned}$$

Table 1 lists the coefficients $h_N(n)$ for the cases $N = 2, 3, \cdots, 10$. For the lowest values of N, $Q_N(\xi)$ can be determined analytically. One has, e.g.,

$$Q_2(\xi) = \tfrac{1}{2}\left[(1+\sqrt{3}) + (1-\sqrt{3})e^{i\xi}\right] \qquad \text{(see (4.4))}$$

and

$$Q_3(\xi) = \tfrac{1}{4}\Big[\left(1+\sqrt{10}+\sqrt{5+2\sqrt{10}}\right) + 2(1-\sqrt{10})e^{i\xi} + \left(1+\sqrt{10}-\sqrt{5+2\sqrt{10}}\right)e^{2i\xi}\Big].$$

For larger values of N, the coefficients in Table 1 were computed numerically.

Since the ${}_Nm_0$ satisfy all the conditions of Theorem 3.6, there exists an associated orthonormal basis of continuous wavelets with compact support for every ${}_Nm_0$. Let us denote the corresponding ϕ, ψ functions by ${}_N\phi$, ${}_N\psi$. Since $h_N(n) = 0$ for $n < 0$ and $n > 2N-1$, it follows (see the discussion at the start of Section 4) that $\operatorname{supp}({}_N\phi) = [0, 2N-1]$. The support of ${}_N\psi$,

$$({}_N\psi)(x) = \sum_{n=0}^{2N-1} (-1)^n h_N(-n+1)\,{}_N\phi(2x-n),$$

is therefore given by $[-(N-1), N]$. Note that an additional phase factor $e^{iK\xi}$, $K \in \mathbf{Z}$, in (4.24) would amount to shifting the $h_N(n)$ by K, i.e., to shifting the function ${}_N\phi$ by an integer, which does not affect the multiresolution analysis construction. The wavelet ${}_N\psi$ is unaffected by this shift.

From Theorem 3.6, we know that ${}_N\phi$ and ${}_N\psi$ are bounded, continuous functions. For large N, ${}_N\phi$ and ${}_N\psi$ are, in fact, much more regular. To see this, we shall need the following generalization of Lemma 3.2.

LEMMA 4.6. *If* $m_0(\xi) = [\frac{1}{2}(1+e^{i\xi})]^N \mathcal{F}(\xi)$, *where* $\mathcal{F}(\xi) = \sum_n f_n e^{in\xi}$ *satisfies*

$$\sum_n |f_n|\,|n|^\varepsilon < \infty \quad \textit{for some} \quad \varepsilon > 0, \tag{4.25}$$

$$\sup_\xi \left|\mathcal{F}(\xi)\mathcal{F}\left(\tfrac{1}{2}\xi\right)\cdots\mathcal{F}\left(2^{-k+1}\xi\right)\right| = B_k, \tag{4.26}$$

then

$$\left|\prod_{j=1}^{\infty} m_0(2^{-j}\xi)\right| \leqq C(1+|\xi|)^{-N+\log B_k/(k\log 2)}. \tag{4.27}$$

Proof: Define

$$\mathcal{F}_k(\xi) = \prod_{j=0}^{k} \mathcal{F}(2^{-j}\xi).$$

Table 1. The coefficients $h_N(n)$ $(n = 0, \cdots, 2N-1)$ for $N = 2, 3, \cdots, 10$.

	n	$h_N(n)$		n	$h_N(n)$
$N=2$	0	.482962913145	$N=8$	0	.054415842243
	1	.836516303738		1	.312871590914
	2	.224143868042		2	.675630736297
	3	−.129409522551		3	.585354683654
$N=3$	0	.332670552950		4	−.015829105256
	1	.806891509311		5	−.284015542962
	2	.459877502118		6	.000472484574
	3	−.135011020010		7	.128747426620
	4	−.085441273882		8	−.017369301002
	5	.035226291882		9	−.044088253931
$N=4$	0	.230377813309		10	.013981027917
	1	.714846570553		11	.008746094047
	2	.630880767930		12	−.004870352993
	3	−.027983769417		13	−.000391740373
	4	−.187034811719		14	.000675449406
	5	.030841381836		15	−.000117476784
	6	.032883011667	$N=9$	0	.038077947364
	7	−.010597401785		1	.243834674613
$N=5$	0	.160102397974		2	.604823123690
	1	.603829269797		3	.657288078051
	2	.724308528438		4	.133197385825
	3	.138428145901		5	−.293273783279
	4	−.242294887066		6	−.096840783223
	5	−.032244869585		7	.148540749338
	6	.077571493840		8	.030725681479
	7	−.006241490213		9	−.067632829061
	8	−.012580751999		10	.000250947115
	9	.003335725285		11	.022361662124
$N=6$	0	.111540743350		12	−.004723204758
	1	.494623890398		13	−.004281503682
	2	.751133908021		14	.001847646883
	3	.315250351709		15	.000230385764
	4	−.226264693965		16	−.000251963189
	5	−.129766867567		17	.000039347320
	6	.097501605587	$N=10$	0	.026670057901
	7	.027522865530		1	.188176800078
	8	−.031582039318		2	.527201188932
	9	.000553842201		3	.688459039454
	10	.004777257511		4	.281172343661
	11	−.001077301085		5	−.249846424327
$N=7$	0	.077852054085		6	−.195946274377
	1	.396539319482		7	.127369340336
	2	.729132090846		8	.093057364604
	3	.469782287405		9	−.071394147166
	4	−.143906003929		10	−.029457536822
	5	−.224036184994		11	.033212674059
	6	.071309219267		12	.003606553567
	7	.080612609151		13	−.010733175483
	8	−.038029936935		14	.001395351747
	9	−.016574541631		15	.001992405295
	10	.012550998556		16	−.000685856695
	11	.000429577973		17	−.000116466855
	12	−.001801640704		18	.000093588670
	13	.000353713800		19	−.000013264203

Then

$$\prod_{j=1}^{\infty} \mathscr{F}(2^{-j}\xi) = \prod_{j=0}^{\infty} \mathscr{F}_k\left[(2^k)^{-j}\tfrac{1}{2}\xi\right].$$

Repeating the proof of Lemma 3.2, with multiplication factor 2^k instead of 2 leads to

$$\left|\prod_{j=0}^{\infty} \mathscr{F}_k\left[(2^k)^{-j}\xi\right]\right| \leqq C\exp\{\log B_k \log|\xi|/\log(2^k)\}.$$

This implies (4.27).

To interpolate between the standard spaces C^k of k times continuously differentiable functions, we shall use, for $\alpha \notin \mathbb{N}$, $\alpha > 0$, the spaces defined by

$$f \in C^\alpha \Leftrightarrow \int dx\, |\hat{f}(\xi)|(1+|\xi|)^{1+\alpha} < \infty. \tag{4.28}$$

Note that, for $\alpha = k \in \mathbb{N}$, the condition (4.28) implies $f \in C^k$, but is not necessary.

We then have the following

PROPOSITION 4.7. *There exists* $\lambda > 0$ *such that, for all* $N \in \mathbb{N}$, $N \geqq 2$,

$$ {}_N\phi, {}_N\psi \in C^{\lambda N}. \tag{4.29}$$

Proof: We shall apply Lemma 4.6. Since $Q_N(e^{i\xi})$ has only a finite number of terms, (4.25) is obviously satisfied. We compute

$$B_2 = \sup|Q_N(e^{i\xi})Q_N(e^{i\xi/2})| = \sup_{\xi} |P_N(\sin^2\tfrac{1}{2}\xi)P_N(\sin^2\tfrac{1}{4}\xi)|$$

$$= \sup_{0\leqq y\leqq 1} |P_N(4y(1-y))P_N(y)|.$$

First, note that (see (4.28))

$$\sup_{0\leqq y\leqq 1} P_N(y) = P_N(1) < 2^{2(N-1)}.$$

Secondly,

$$P_N(y) = \sum_{k=0}^{N-1} \binom{N+k-1}{k} y^k$$

$$\leqq \sum_{k=0}^{N-1} 2^{N+k-1} y^k \leqq 2^{N-1} N \max\left(1, (2y)^N\right).$$

Hence, for $y \leqq \frac{1}{2}$,

$$P_N(y)P_N(4y(1-y)) \leqq N2^{N-1}2^{2(N-1)} = N2^{3(N-1)}.$$

For $y \geqq \frac{1}{4}(2+\sqrt{2})$, or $4y(1-y) \leqq \frac{1}{2}$,

$$P_N(y)P_N(4y(1-y)) \leqq 2^{2(N-1)}N2^{N-1} = N2^{3(N-1)}.$$

Finally, for $\frac{1}{2} \leqq y \leqq \frac{1}{4}(2+\sqrt{2})$,

$$P_N(y)P_N(4y(1-y)) \leqq N^2 2^{4N-2}\Big(\sup_{0\leqq y\leqq 1}\left[4y^2(1-y)\right]\Big)^N$$

$$= N^2 2^{4N-2}\left(\tfrac{16}{27}\right)^N,$$

or

$$B_2 \leqq N2^{2N-1}\left(\tfrac{16}{27}\right)^{N/2}.$$

Consequently,

$$\left|({}_N\phi)\hat{}\,(\xi)\right| = (2\pi)^{-1/2}\left|\prod_{j=1}^{\infty} m_0(2^{-j}\xi)\right|$$

$$\leqq C(1+|\xi|)^{[\log N - N\log(3\sqrt{3}/4)]/2\log 2}.$$

This exponent is smaller than -1 for $N \geqq 16$. For smaller values of N, one can use the explicit estimate

$$B_1 = \left[\binom{2N-1}{N}\right]^{1/2}$$

to prove that

$$\left|({}_N\phi)\hat{}\,(\xi)\right| \leqq C(1+|\xi|)^{-1-\kappa N}$$

for some $\kappa > 0$, for all $N \leqq 16$. Hence (4.29) holds for ${}_N\phi$, for some $\lambda > 0$, and for all $N \geqq 2$. Since ${}_N\psi$ is always a finite linear combination of translated and dilated versions of ${}_N\phi$, the same holds for ${}_N\psi$.

Remarks. 1. Since $|\mathrm{supp}({}_N\phi)| = |\mathrm{supp}({}_N\psi)| = 2N-1$, (4.29) shows that the regularity of ${}_N\phi$, ${}_N\psi$ increases linearly with their support width, as announced in the introduction. It turns out that linear increase of the support width with the regularity of ϕ, ψ is the best one can do. More precisely, if a C^k-function ϕ satisfies an equation of the type

$$\phi(x) = \sum_{n=0}^{N} c_n \phi(2x-n)$$

(without necessarily being connected to multiresolution analysis), and if supp $\phi \subset [0, N]$, then $k \leqq N - 2$. For a proof, see [30].

2. The estimate for λ obtained in this proof is, of course, not very good; the argument is too simple. Asymptotically, for large N, one finds

$$ {}_N\phi, {}_N\psi \in C^{(\mu-\varepsilon)N} $$

with

$$ \mu \sim \frac{\log\left(\frac{3}{4}\sqrt{3}\right)}{2\log 2} + O\left(\frac{\log N}{N}\right) \geqq .1887 + O\left(\frac{\log N}{N}\right). $$

The same technique, with a little more work, leads to slightly better estimates if larger values of k are used. Using $k = 4$, e.g., leads to

$$ \mu \geqq .1936 + O(N^{-1}\log N). $$

Since the map $y \mapsto 4y(1 - y)$ has a fixed point, at $y = \frac{3}{4}$, one finds

$$ B_k \geqq \left[P_N\left(\tfrac{3}{4}\right)\right]^{k/2}. $$

One can show that

$$ P_N\left(\tfrac{3}{4}\right) \sim C3^N. $$

Even for arbitrarily large k, the values of μ obtained by this method are therefore limited by

$$ \mu \leqq 1 - \frac{\log 3}{2\log 2} + O(N^{-1}\log N) \cong .2075 + O(N^{-1}\log N). $$

3. Using a more sophisticated method than the brutal estimates above, Y. Meyer [28] showed that, again asymptotically for large N,

$$ {}_N\phi, {}_N\psi \in C^{(\mu-\varepsilon)N} $$

with $\mu = \log(4/\pi)/\log 2 \cong .3485$. His proof uses (4.22) rather than P_N.

4. For small values of N, better estimates can be obtained for the regularity of the ${}_N\phi$, ${}_N\psi$ by yet a third method. This method is based on a generalization of a technique used by Riesz in the proof that "Riesz products" can lead to continuous, nowhere differentiable functions. I would like to thank Y. Meyer for introducing me to this technique, and for showing me how to use it to prove ${}_2\phi, {}_2\psi \in C^{.5-\varepsilon}$. The proof, and a generalization for $N \geqq 3$, are given in the Appendix. It works very well for small values of N, but does not, however, give good asymptotic results. For large N, it leads to logarithmic rather than linear increase of the regularity of the ${}_N\phi$, ${}_N\psi$.

Table 2. Regularity estimates.
For $N = 2, \cdots, 10$, we give α_N so that ${}_N\phi, {}_N\psi \in C^{\alpha_N}$

N	α_N
2	$.5 - \varepsilon$
3	.915
4	1.275
5	1.596
6	1.888
7	2.158
8	2.415
9	2.661
10	2.902

To conclude this paper, we give in Figure 7 the graphs of ${}_N\phi$, ${}_N\psi$ and their Fourier transforms $({}_N\phi)\hat{}$, $({}_N\psi)\hat{}$, for $N = 3, 5, 7, 9$. (For $N = 2$, these graphs were given in Figure 5.) The graphs were plotted by means of the "graphical algorithm" explained in subsection 2B, using the coefficients $h_N(n)$ of Table 1. One clearly sees that the ${}_N\phi$, ${}_N\psi$ become more regular as N increases. Also noticeable is that $|({}_N\phi)\hat{}|$, $|({}_N\psi)\hat{}|$ become "flatter" as N increases, around 0 and $2\pi \cong 6.28$. This is a direct consequence of (4.7) and (4.8). By (4.3), $({}_Nm_0)(\xi)$ has a zero of order N at $\xi = \pi$. It follows that, by (4.7), $({}_Nm_0)(0) = 1$, and that the first $N - 1$ derivatives $({}_Nm_0)^{(j)}(\xi)$ of ${}_Nm_0$ are zero in $\xi = 0$. Since (this follows from (3.45)) $({}_N\psi)\hat{}\,(\xi) = \overline{{}_Nm_0\left(\pi + \frac{1}{2}\xi\right)}\,({}_N\phi)\hat{}\,(\frac{1}{2}\xi)$, this means that $[({}_N\psi)\hat{}\,]^{(k)}(0) = 0$ for $k = 0, \cdots, N - 1$, or $\int dx\, x^k({}_N\psi)(x) = 0$ for $k = 0, \cdots, N - 1$. The present construction leads thus also to orthonormal bases of compactly supported wavelets with an arbitrarily high number of zero moments. This property could be useful for quantum field theory (see [18]).

It is also quite striking that the "effective support" (where $|({}_N\psi)(x)| \geqq .01\|{}_N\psi\|_\infty$, say) of ${}_N\psi$ is quite a bit smaller than its total support, for N not too small. This is due to the very small value of the $h_N(n)$ for large n (see Table 1). Table 2 lists the estimates for the "regularity index" α_N (where ${}_N\phi, {}_N\psi \in C^{\alpha_N}$) for $N = 2, 3, \cdots, 10$, computed using the method explained in the Appendix.

Remark. Using a different approach (see [30]), these estimates for the regularity index α_N can be sharpened. For $N = 2$ one finds, e.g., $\alpha_2 = 2 - \ln(1 + \sqrt{3})/\ln 2 \simeq .550 \cdots$. This is the best possible exponent for $N = 2$ (see [30]).

Appendix
Sharper Regularity Estimates for ${}_N\phi, {}_N\psi$

The estimates given here are based on a different way of calculating (4.28). Using the facts that $|({}_N\phi)\hat{}\,(\xi)| = (2\pi)^{-1/2}\prod_{j=1}^{\infty}|({}_Nm_0)(2^{-j}\xi)|$ is even (because ${}_Nm_0$ is a trigonometric polynomial with real coefficients) and that $|({}_Nm_0)(\xi)| \leqq 1$

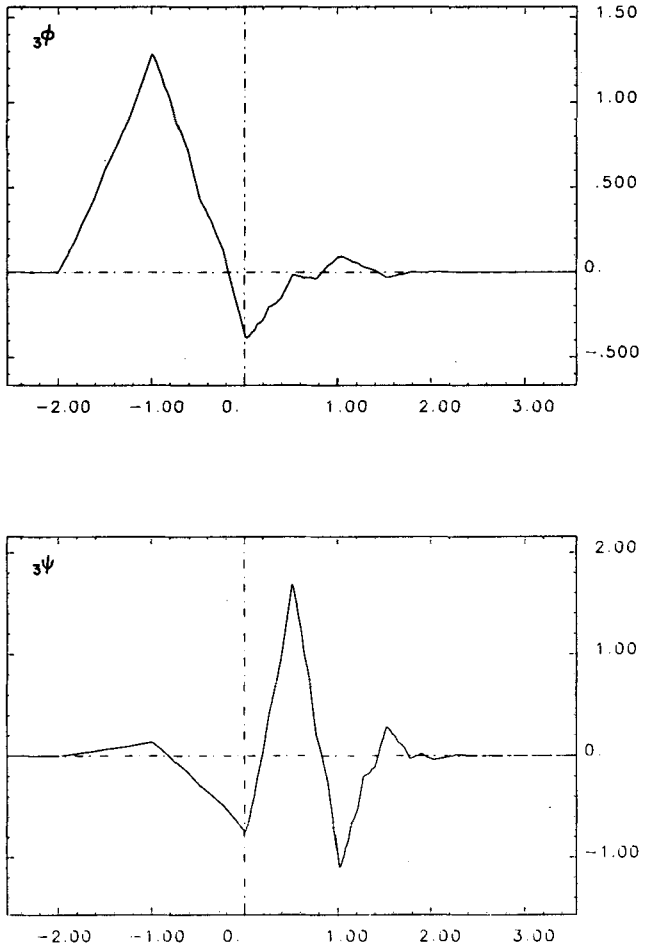
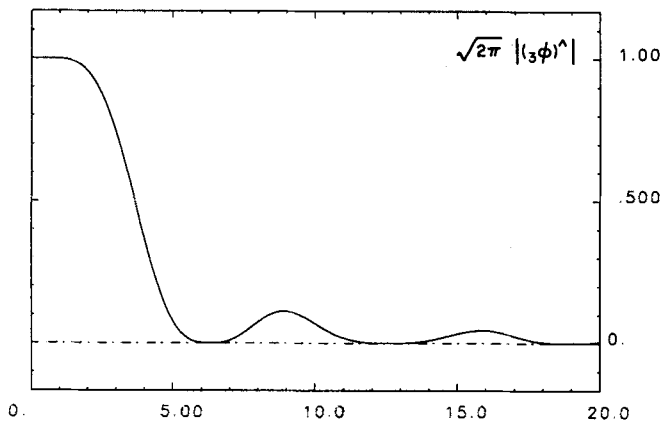


Figure 7. The functions ${}_N\phi$, ${}_N\psi$ and the modulus of their Fourier transforms $|({}_N\phi)\hat{}|$, $|({}_N\psi)\hat{}|$, for increasing values of N (see text). We have each time shifted ${}_N\phi$ by $N-1$, so that $\text{supp}({}_N\phi) = \text{supp}({}_N\psi) = [-(N-1), N]$. One clearly sees that the ${}_N\phi$, ${}_N\psi$ become more regular as N increases. The function ${}_N\phi$ has been plotted using the "graphical construction algorithm" explained in subsection 2B, with the weighting coefficients ${}_Nh(n)$ given in Table 1. Only 7 iterations were needed. The plot of ${}_N\psi$ then follows from $({}_N\psi)(x) = \sqrt{2}\Sigma_n(-1)^n h_N(-n+1)({}_N\phi)(2x-n)$.

(see (4.7)), we find

$$\int d\xi |({}_N\phi)\hat{}\,(\xi)|(1+|\xi|)^{1+\alpha} \leqq (2\pi)^{-1/2} 2^{\alpha+1}(1+a)^{\alpha+1}$$

$$\cdot \left\{ a + \sum_{m=0}^{\infty} 2^{(m+1)(\alpha+1)} \int_{2^m a}^{2^{m+1}a} d\xi \prod_{j=0}^{m} |({}_Nm_0)(2^{-j}\xi)| \right\}, \tag{A.1}$$



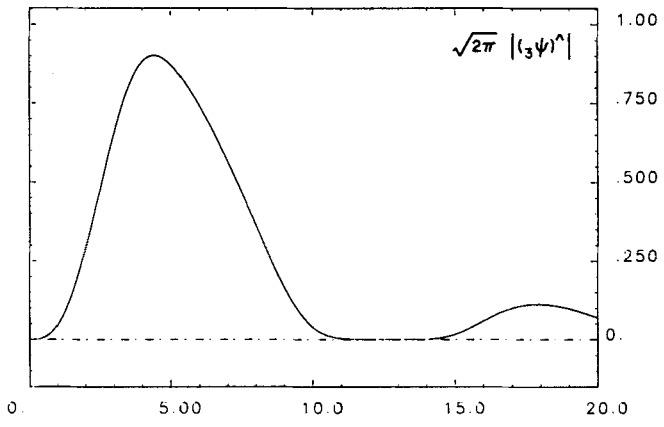


Figure 7. For the plots of $|({}_N\phi)|^\wedge$ the infinite product (3.46) was computed (truncated at $j = 10$), $(2\pi)^{1/2}|({}_N\phi)^\wedge(\xi)| = \prod_{j=1}^{\infty}|{}_Nm_0(2^{-j}\xi)|$, with ${}_Nm_0(\xi) = 2^{-1/2}\Sigma_n h_N(n)e^{in\xi}$, where the $h_N(n)$ are given in Table 1. The plot of $|({}_N\psi)^\wedge(\xi)|$ then follows from

$$|({}_N\psi)^\wedge(\xi)| = 2^{-1/2}|\Sigma_n(-1)^n h(-n+1)e^{in\xi/2}|\,|({}_N\phi)^\wedge(\tfrac{1}{2}\xi)|.$$

where $a > 0$ is arbitrary for the moment. Using $({}_Nm_0)(\xi) = [\frac{1}{2}(1 + e^{i\xi})]^N Q_N(e^{i\xi})$, we find

$$\prod_{j=0}^{m}\left|({}_Nm_0)(2^{-j}\xi)\right| = \left[\frac{|\sin\xi|}{2^{m+1}\left|\sin(2^{-m-1}\xi)\right|}\right]^N \prod_{j=0}^{m}\left|Q_N\left(e^{-2^{-j}\xi}\right)\right|.$$

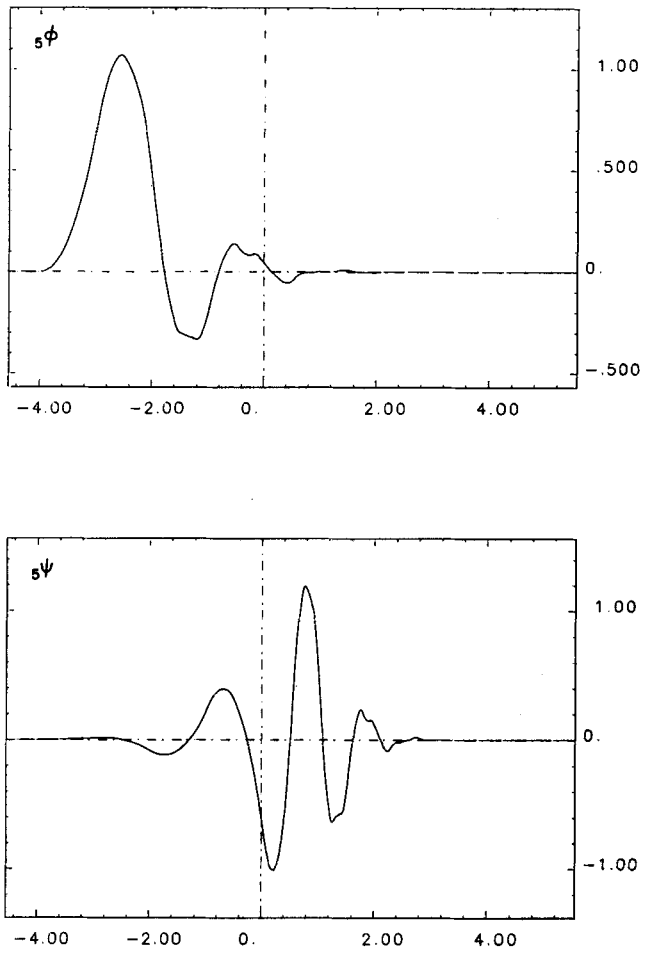


Figure 7. Continued

If we choose $a = \frac{2}{3}\pi$, then $|\sin(2^{-m-1}\xi)| \geqq \sqrt{\frac{3}{2}}$ for $2^m a \leqq |\xi| \leqq 2^{m+1}a$. Hence

$$\int d\xi \left|({}_N\phi)^{\wedge}(\xi)\right| (1 + |\xi|)^{1+\alpha}$$

$$\leqq C_1 + C_2 \sum_{m=0}^{\infty} 2^{(\alpha+1)m} 2^{-N(m+1)} \int_{2^m 2\pi/3}^{2^{m+1}2\pi/3} \prod_{j=0}^{m} \left| Q_N\left(e^{i2^{-j}\xi}\right)\right|$$

$$\leqq C_1 + C_2 \sum_{m=0}^{\infty} 2^{(\alpha+1)m} 2^{-N(m+1)} (2^{m+1}\pi)^{1/2} \left[\int_0^{2^{m+1}\pi} d\xi \prod_{j=0}^{m} \left| Q_N\left(e^{i2^{-j}\xi}\right)\right|^2\right]^{1/2}$$

$$\leqq C_1 + C_3 \sum_{m=0}^{\infty} 2^{m(\alpha-N+1)} \left[\int_0^{2\pi} d\xi \prod_{j=0}^{m} P_N\left(\sin^2\left(2^j \tfrac{1}{2}\xi\right)\right)\right]^{1/2},$$

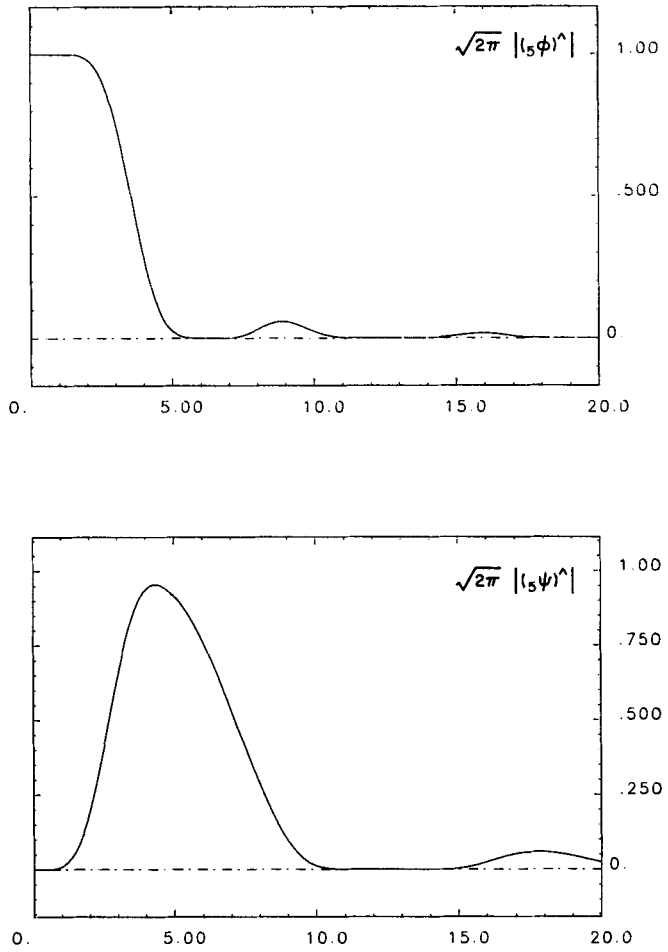


Figure 7. Continued

where we have used $|Q_N(e^{i\xi})|^2 = P_N(\sin^2\frac{1}{2}\xi)$ (see subsection 4C). It follows that (A.1) is convergent, hence ${}_N\phi,\ {}_N\psi \in C^\alpha$, if

$$\text{(A.2)}\quad \limsup_{m\to\infty}(2m\log 2)^{-1}\log\left[\int_0^{2\pi} d\xi \prod_0^{2\pi} P_N\left(\sin^2\left(2^{j-1}\xi\right)\right)\right] \leqq N-1-\alpha.$$

We know P_N explicitly (see (4.13)),

$$\text{(A.3)}\qquad P_N\left(\sin^2\tfrac{1}{2}\xi\right) = \sum_{k=0}^{N-1}\binom{N+k-1}{k}\left(\sin^2\tfrac{1}{2}\xi\right)^k.$$

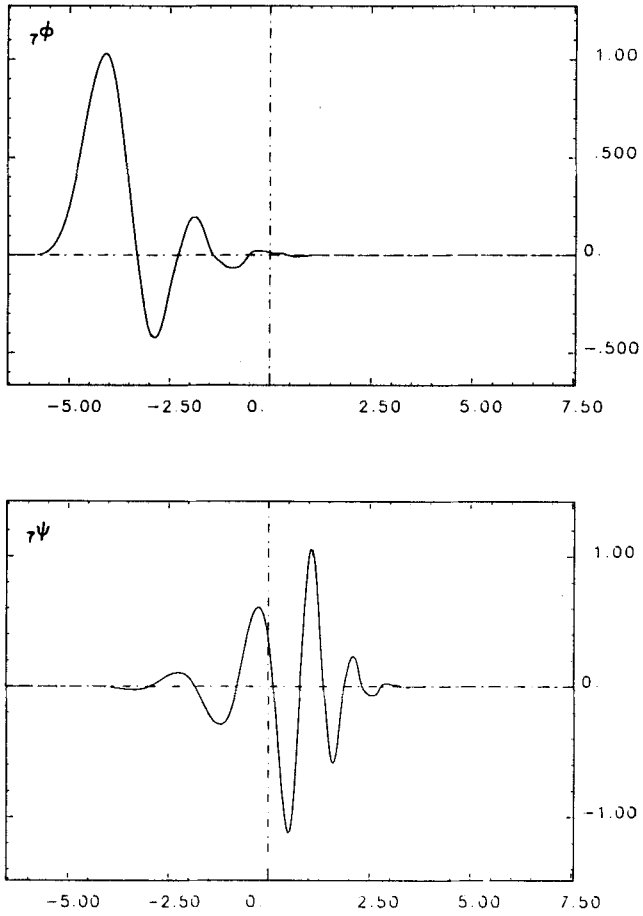


Figure 7. Continued

This can be rewritten as

$$P_N\left(\sin^2\tfrac{1}{2}\xi\right) = \sum_{l=-(N-1)}^{N-1} a_{N,l}e^{il\xi}, \tag{A.4}$$

where the $a_{N,l}$ are symmetric, $a_{N,l} = a_{N,-l}$, and can be calculated explicitly from (A.3). The product $\prod_{j=0}^{m} P_N(\sin^2(2^{j-1}\xi))$ is therefore a symmetric trigonometric polynomial of order $2^m(N-1)$,

$$\prod_{j=0}^{m} P_N\left(\sin^2\left(2^{j-1}\xi\right)\right) = \sum_{l=-(N-1)2^m}^{(N-1)2^m} J_{N,m;k}e^{ik\xi}. \tag{A.5}$$

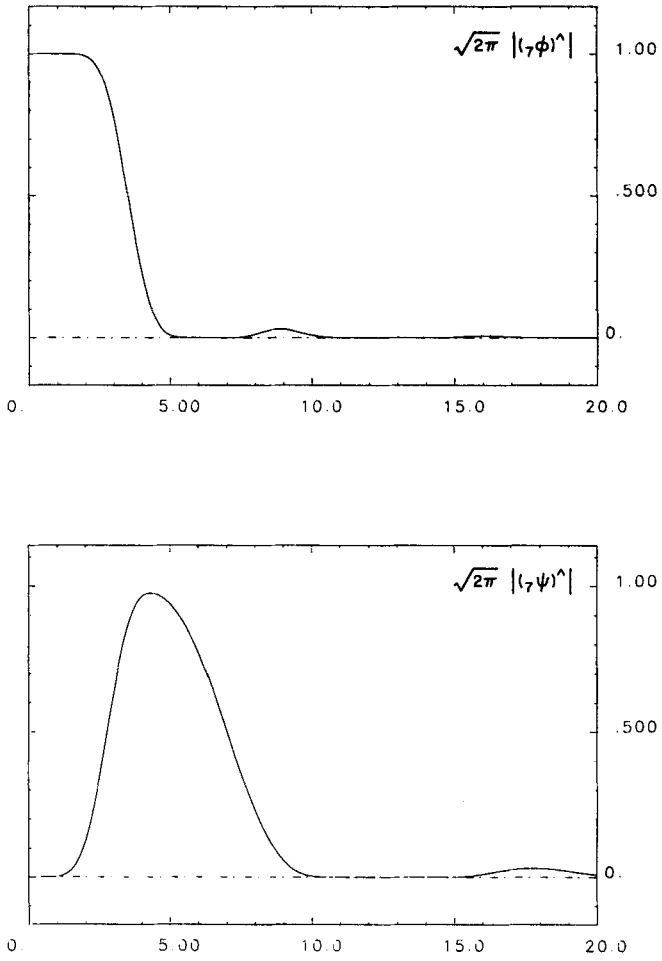


Figure 7. Continued

One easily checks that

$$J_{N,m;2k} = \sum_l a_{N,2l} J_{N,m-1;k-l},$$

(A.6)

$$J_{N,m;2k+1} = \sum_l a_{N,2l+1} J_{N,m-1;k-l},$$

with $J_{N,0;k} = a_{N,k}$, and where we implicitly make the assumption $a_{N,k} = 0$ for $|k| \geqq N$. The recursion (A.6) can be represented graphically, in a construction

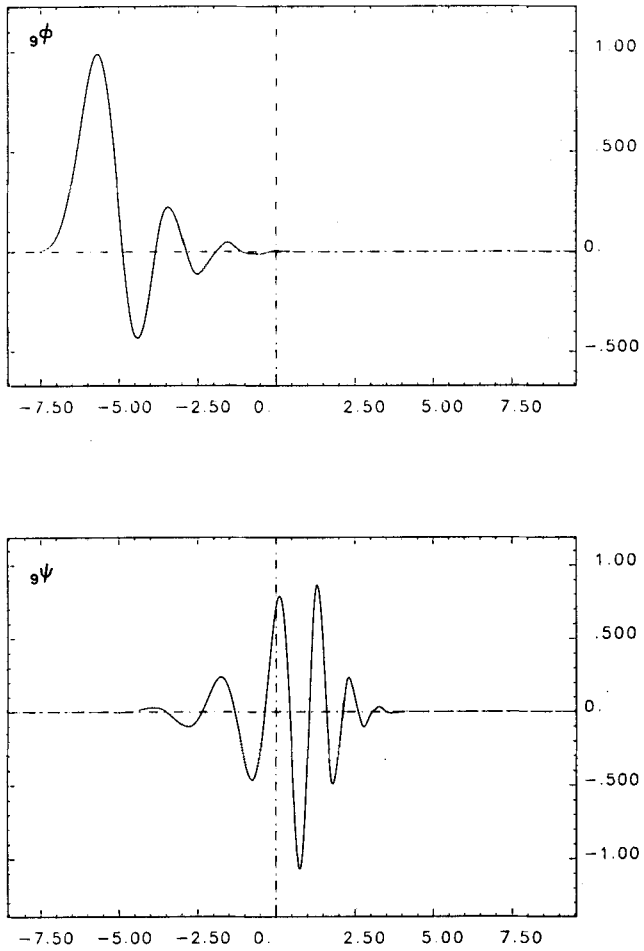


Figure 7. Continued

analogous to Figure 1. At level 0 we start with $J_{N,0}$; each successive $J_{N,n}$ is calculated from $J_{N,n-1}$ by a tree algorithm (see Figure 8). To evaluate the left-hand side of (A.2) we need to compute

$$\int_0^{2\pi} d\xi \prod_{j=0}^{m} P_N\left(\sin^2\left(2^{j-1}\xi\right)\right) = J_{N,m;0}. \tag{A.7}$$

One can check directly from the recursion (A.6), or one can verify on the graphical representation (see Figure 8b) that only the $J_{N,m';l}$, $0 \leqq m' < m$, with $|l| \leqq N - 2$ play a role in the computation of $J_{N,m;0}$. Define $d_N = 2N - 3$. Then

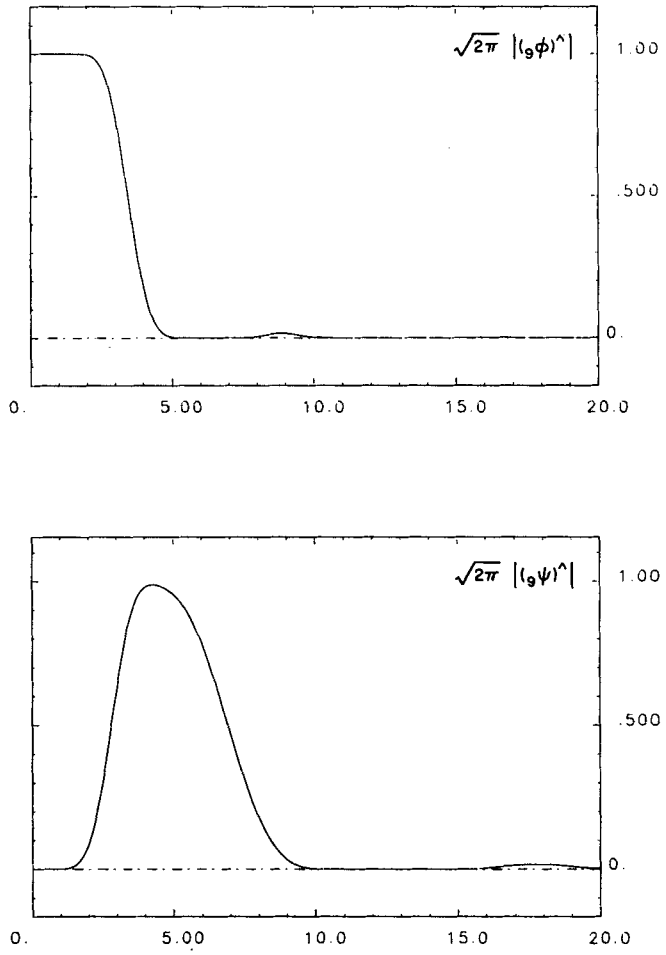


Figure 7. Continued

the set of relevant $J_{N,m';l}, \cdots, |l| \leqq N-2, \cdots$, define a vector $j_{N,m'}$ in $\mathbb{R}^{d_N}$,

$$(j_{N,m'})_k = J_{N,m';k}. \tag{A.8}$$

Note that d_N is always odd, $d_N = 2m_N + 1$, and that we index vectors $v \in \mathbb{R}^{d_N}$ by $j = -m_N, -m_n + 1, \cdots, 0, \cdots, m_N$ (see (A.8).) The recursion (A.6) defines a matrix T_N such that, for all m,

$$j_{N,m+1} = T_N j_{N,m}. \tag{A.9}$$

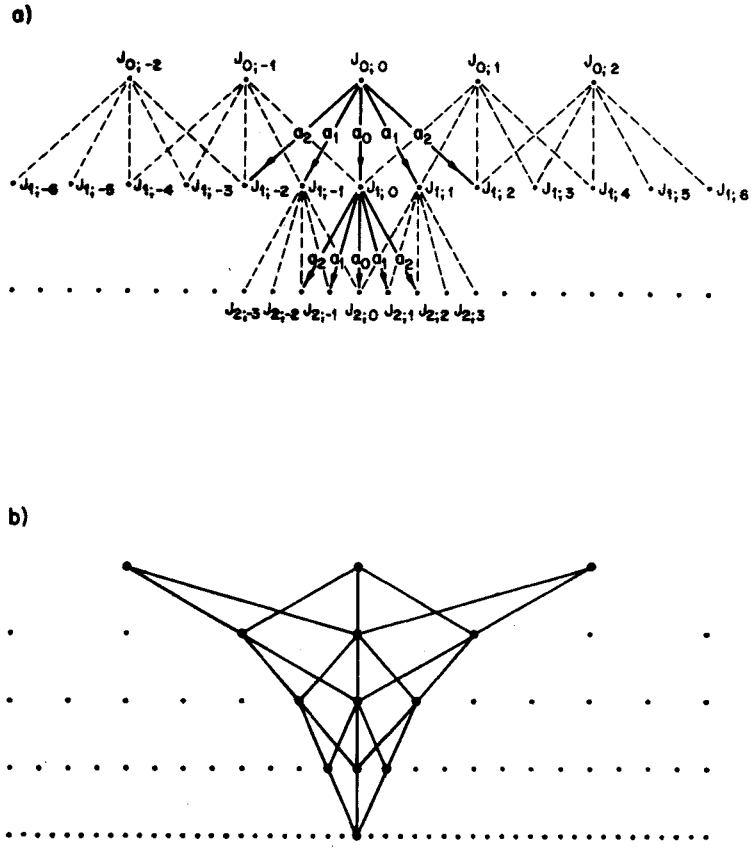


Figure 8. a. The tree algorithm for the construction of the $J_{N,m}$. For the sake of simplicity, we have taken $N = 3$. The index N is dropped on the figure.

b. Although the number of non-zero $J_{3,m;k}$ more than doubles (see a)) at every step, only 3 points, at *any* level, ultimately contribute to $J_{3,m;0}$. These are the points which can be reached from 0 by the tree, starting from the bottom.

This matrix has the following form, for N even:

(A.10)

$$T = \begin{pmatrix} a_{N-2} & a_{N-4} & a_{N-6} & \cdots & a_2 & a_0 & a_2 & \cdots & a_{N-2} & 0 & 0 & \cdots & 0 \\ a_{N-1} & a_{N-3} & a_{N-5} & \cdots & a_3 & a_1 & a_1 & \cdots & a_{N-3} & a_{N-1} & 0 & \cdots & 0 \\ 0 & a_{N-2} & a_{N-4} & \cdots & a_4 & a_2 & a_0 & \cdots & a_{N-4} & a_{N-2} & 0 & \cdots & 0 \\ 0 & a_{N-1} & a_{N-3} & \cdots & a_5 & a_3 & a_1 & \cdots & a_{N-5} & a_{N-3} & a_{N-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & a_{N-3} & a_{N-5} & a_{N-7} & \cdots & a_{N-1} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & a_{N-2} & a_{N-4} & a_{N-6} & \cdots & a_{N-2} \end{pmatrix}.$$

A completely analogous matrix is obtained for N odd. From (A.8)–(A.9) we have

$$J_{N,m;0} = \left(T_N^m j_{N,0}\right)_0.$$

Hence

$$\begin{aligned}
&\limsup_{m\to\infty} m^{-1}\log\left(J_{N,m;0}\right) \\
&\quad \leqq \limsup_{m\to\infty} m^{-1}\log\left[\|(T_N)^m\| \cdot \|j_{N,0}\|\right] \\
&\quad \leqq \limsup_{m\to\infty} \log\left[\|(T_N)^m\|^{1/m}\right] = \log\left(\rho(T_N)\right),
\end{aligned}$$

where $\rho(T_N)$ is the spectral radius of T_N. In view of (A.7) it then follows that (A.1) is convergent, i.e., ${}_N\phi \in C^\alpha$, if $\alpha < N - 1 - \frac{1}{2}\log_2[\rho(T_N)]$. It suffices therefore to compute $\rho(T_N)$, which can be done numerically, provided N is not too large. Note that the problem can be reduced considerably by using the fact that T_N commutes with the involution I,

$$I_{ij} = \delta_{i,-j}$$

(where, as before, $i, j = -m_N, \cdots, 0, \cdots, m_N$). This effectively reduces the problem of a $d_N \times d_N$ matrix to a $(m_N + 1) \times (m_N + 1)$ matrix.

If $N = 2$, then $d_N = 1$, and the matrix T_1 is given by a single number, $T_1 = a_{1;0} = 2$. Therefore one finds ${}_1\phi \in C^\alpha$ if $\alpha < \frac{1}{2}$. The cause of this simplification can be understood by looking at Figure 8b. For $N = 2$, the "tree" reduces to a single vertical line: only one possible path leads from $J_{N,0;0}$ to $J_{N,n;0}$ if $N = 2$. This is equivalent to saying that in the product $\prod_{j=0}^{N} P_N(\sin^2(2^{j-1}\xi))$ only one possible combination of terms has frequency zero. This is the idea which was borrowed from Riesz's lemma (see Remark 4 following Proposition 4.7).

Acknowledgment. Part of this work was done while I was visiting the Courant Institute (New York University) and the Mathematics Department of Yale University, in February and March of 1987, respectively. I would like to thank both these institutions for their hospitality and support. While working on this subject, I have had the pleasure of discussing these topics with many people, whom I would like to thank here. I am particularly grateful to Yves Meyer for explaining to me the multiresolution analysis concept when it was still very young and new, and for showing me how to use a technique of Riesz to obtain the estimates in the appendix, to Bob Hummel for introducing me to the Laplacian pyramid scheme, thus providing me with the background which led to the present construction, to Stéphane Mallat for letting me use his not yet published material on his discrete algorithm, and to Percy Deift, who listened to several of the early versions, and challenged me with many pertinent questions and made several

useful suggestions. I also would like to thank R. Coifman, P. Jones and T. Steeger for their enthusiastic support and for suggesting various references.

The author is "Bevoegdverklaard Navorser" at the Belgium National Science Foundation (on leave). Part of the work for this paper was done at the Department of Theoretical Physics, Vrije Universiteit, Brussel from which she also is on leave. She would like to thank both institutions for their support.

Bibliography

[1] Grossman, A., and Morlet, J., *Decomposition of Hardy functions into square integrable wavelets of constant shape*, SIAM J. Math. Anal. 15, 1984, pp. 723–736.
Goupillaud, P., Grossmann, A., and Morlet, J., *Cyclo-octave and related transforms in seismic signal analysis*, Geoexploration 23, 1984, 85.

[2] The "continuous wavelet transform" is implicitly used in Calderón, A. P., *Intermediate spaces and interpolation, the complex method*, Studia Math. 24, 1964, pp. 113–190; see also Calderón, A. P., and Torchinsky, A., *Parabolic maximal functions associated to a distribution*, I, Adv. Math. 16, 1974, pp. 1–64.

[3] Wilson, K. G., and Kogut, J. B., Physics Reports 12C, 1974, p. 77.
Glimm, J., and Jaffe, A., *Quantum Physics: a Functional Integral Point of View*, Springer, New York, 1981.

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

[5a] Tchamitchian, P., *Calcul symbolique sur les opérateurs de Calderón-Zygmund et bases inconditionnelles de* $L^2(\mathbf{R}^n)$, C. R. Acad. Sc. Paris. 303, série 1, 1986, pp. 215–218.

[5b] Tchamitchian, P., *Biorthogonalité et théorie des opérateurs*, to be published in Rev. Mat. Iberoamericana.

[6] Battle, G., and Federbush, P., *Ondelettes and phase cell cluster expansions; a vindication*, Comm. Math. Phys., 1987.
Federbush, P., *Quantum field theory in ninety minutes*, Bull. Am. Math. Soc. 17, 1987, pp. 93–103.

[7] Kronland-Martinet, R., Morlet, J., and Grossmann, A., *Analysis of sound patterns through wavelet transforms*, to be published in International Journal on Pattern Analysis and Artificial Intelligence.

[8] Mallat, S., *A theory for multiresolution signal decomposition: the wavelet transform*, Preprint GRASP Lab, Dept. of Computer and Information Science, Univ. of Pennsylvania, to be published.

[9] Grossmann, A., *Wavelet transforms and edge detection*, to be published in *Stochastic Processes in Physics and Engineering* (Ph. Blanchard, L. Streit and M. Hasewinkel, eds.),
Grossmann, A., Holschneider, M., Kronland-Martinet, R., and Morlet, J., *Detection of abrupt changes in sound signals with the help of wavelet transforms*, Preprint, Centre de Physique Théorique, CNRS, Marseille (France).

[10] Grossmann, A., Morlet, J., and Paul, T., *Transforms associated to square integrable group representations*, I. J. Math. Phys. 26, 1985, pp. 2473–2479; —II: *Examples*, Ann. Inst. H. Poincaré 45, 1986, pp. 293–309.

[11] Aslaksen, E. W., and Klauder, J. R., *Unitary representations of the affine group*, J. Math. Phys. 9, 1968, pp. 206–211; *Continuous representation theory using the affine group*, J. Math. Phys. 10, 1969, pp. 2267–2275.

[12] Paul, T., *Functions analytic on the half-plane as quantum mechanical states*, J. Math. Phys. 25, 1985, pp. 3252–3263.
Paul, T., *Affine coherent states and the radial Schrödinger equation*. I. *Radial harmonic oscillator and hydrogen atom*, to be published.

[13] Grossmann, A., and Kronland, R., Private demonstration.

[14] Daubechies, I., Grossmann, A., and Meyer, Y., *Painless non-orthogonal expansions*, J. Math. Phys. 27, 1986, pp. 1271–1283.
[15] Daubechies, I., *The wavelet transform, time-frequency localization and signal analysis*, Preprint AT & T Bell Laboratories; to be published.
[16] Meyer, Y., *Ondelettes et functions splines*, Séminaire EDP, Ecole Polytechnique, Paris, France, December, 1986.
[17] Lemarié, P. G., *Ondelettes à localisation exponentielle*, to be published, in Journ. de Math. Pures et Appl.
[18] Battle, G., *A block spin construction of ondelettes, Part I: Lemarié functions*, Comm. Math. Phys. 1987.
[19] Mallat, S., *Multiresolution approximation and wavelets*. Preprint GRASP Lab., Dept. of Computer and Information Science, Univ. of Pennsylvania, to be published.
[20a] Burt, P., and Adelson, E., *The Laplacian pyramid as a compact image code*, IEEE Trans. Comm. 31, 1983, pp. 482–540.
[20b] Burt, P., and Adelson, E., *A multiresolution spline with application to image mosaics*, ACM Trans. on Graphics, 2, 1983, pp. 217–236.
[21] Jaffard, S., Lemarié, P. G., Mallat, S., and Meyer, Y., *Multiscale analysis*, unpublished memorandum.
[22] Coifman, R., and Meyer, Y., *The discrete wavelet transform*, preprint. Dept. of Math., Yale University.
[23] Lemarié, P. G., and Meyer, Y., *Ondelettes et bases hilbertiennes*, Rev. Mat. Iberoamericana 2, 1986, pp. 1–18.
[24] Smith, M. J., and Barnwell, T. P., *Exact reconstruction techniques for tree-structured subband coders*, IEEE Trans. on ASSP 34, 1986, 434–441.
[25] Auscher, P., Thèse de Doctorat, Université de Paris-Dauphine 1988.
[26] Rioul, O., private communication.
[27] Polya, G., and Szegö, G., *Aufgaben and Lehrsätze aus der Analysis*, Vol. II, Springer, Berlin, 1971.
[28] Meyer, Y., *Wavelets with compact support*, Zygmund Lectures (University of Chicago) May 1987, and private communication.
[29] Deslauriers, G., and Dubuc, S., *Interpolation dyadique*, in *Fractals; Dimensions non Entières et Applications*, ed. G. Cherbit, Masson (Paris), 1987, pp. 44–45. See also other references listed there.
[30] Daubechies, I., and Lagarias, J., *Two-scale difference equations*. I. *Global regularity of solutions*, and —. II. *Infinite matrix products, local regularity and fractals*, Preprints AT & T Bell Laboratories; to be published.

Received December, 1987