# EXHIBIT N

TO DECLARATION OF S. MERRILL WEISS IN SUPPORT OF PLAINTIFF ACACIA MEDIA TECHNOLOGIES CORPORATION'S MEMORANDUM OF POINTS AND AUTHORITIES IN OPPOSITION TO ROUND 3 DEFENDANTS' MOTION FOR SUMMARY JUDGMENT OF INVALIDITY UNDER 35 U.S.C. § 112 OF THE '992, '863, AND '702 PATENTS; AND SATELLITE DEFENDANTS' MOTION FOR SUMMARY JUDGMENT OF INVALIDITY OF THE '992, '863, AND '720 PATENTS

VOLUME XLI, NUMBER 7, OCTOBER 1988

# Communications on PURE AND APPLIED MATHEMATICS



A JOURNAL ISSUED BY
THE COURANT INSTITUTE OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY



WILEY
A Wiley-Interscience Publication
JOHN WILEY & SONS
New York · Chichester · Brisbane · Toronto · Singapore

CPAMA 41(7) 865–996 (1988)
ISSN 0010-3640

# 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, bave 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

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

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 $m, n \in \mathbb{Z}$,

$$h_{mn}(x) = a_0^{-m/2} h(a_0^{-m} x - n b_0). \tag{1.6}$$

Note that this corresponds to the choices

$$a = a_0^m,$$

$$b = n b_0 a_0^m,$$

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 $\mathbf{Z}^2$,

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

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

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

for all $f$ in $L^2(\mathbf{R})$, then the set $\{h_{mn};\ m, n \in \mathbf{Z}\}$ 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

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

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 $l^2(\mathbf{Z}^2)$. 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 $C^\infty$-function $h$ of rapid decay (in fact $\hat{h}$, in his example, is a compactly supported $C^\infty$-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 $k \in \mathbf{N}$, a $C^k$-function $\psi$ with compact support, such that the corresponding $\psi_{mn}$,

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

constitute an orthonormal basis. The size of the support increases linearly with the regularity. Moreover, $\psi$ 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 $\psi$. 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 $V_m \subset L^2(\mathbf{R})$, $m \in \mathbf{Z}$,

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

such that (ii)

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

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 $m \in \mathbf{Z}$, the $\phi_{mn}$ constitute an unconditional basis for $V_m$, that is, (iv)

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

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

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

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 $f \in L^2(\mathbf{R})$. The condition (2.3) ensures that the $V_m$ correspond to different scales, while the translational invariance

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

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,

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

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 $\phi$ the characteristic function of the interval $[0, 1[$,

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

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 $\phi_{0n}$ are orthonormal (which automatically implies that the $\phi_{mn}$ are orthonormal for every fixed $m$). If the $\phi_{0n}$ are not orthonormal to start with, then one defines $\tilde{\phi}$ by

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

(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 $\phi_{0n}$ 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 $\psi_{mn}$ 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 $\psi_{mn}$ of a single function $\psi$. 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 $\psi$ is nothing but the Haar function (see (1.9)), and it is therefore no surprise that the $\psi_{mn}$ 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 $\phi$ 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,

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

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 $L^2(\mathbf{R})$,

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

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 $\psi$ such that its integer translates span $W_0$, i.e.,

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

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

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

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

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 $\phi_{00}$, while $V_{-1}$ is generated by the integer translates of *two* functions, namely $\phi_{-10}$ and $\phi_{-11}$. 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 $\psi$ satisfying (2.14) would however not be enough for practical purposes. A more detailed analysis leads to the following algorithm for the *construction* of $\psi$ (see [16], [19]). We start from a function $\phi$ such that the $\phi_{0n}$ 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 $\psi_{0n}$ will constitute an orthonormal basis of $W_0$; see [16], [19]. Consequently the $\psi_{mn}$, 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 $\phi_{0n}$ 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 $\phi$, $\psi$ 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 $\phi$, $\psi$ are sufficiently well-behaved for these integrals to make sense (in all examples of practieal interest, $\phi, \psi \in L^1$). In fact, one does not even need to assume that the $\phi_{0n}$ or $\psi_{0n}$ are orthonormal to derive (2.17)–(2.18). In [15] it is shown that (2.17) has to be satisfied even if the $\psi_{mn}$ constitute only a frame (see the introduction). Note also that the transition (2.5) from $\phi$ to $\tilde{\phi}$, orthonormalizing the $\phi_{0n}$, preserves $\int dx\, \phi(x) \neq 0$.

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

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

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

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

EXAMPLE 2.2.

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

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.

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

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 $\psi$ 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 $\phi$ is

compactly supported, but the $\phi_{0n}$ 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 $\psi$. 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 $\phi$.

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 $\phi$, $\psi$ satisfying (2.1)–(2.4) and (2.14), where the $\phi_{0n}$ and the $\psi_{0n}$ 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),

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

where $\Phi_{mn}$ is defined by

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

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 complemcnt $\mathbf{W}_m$ 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 $n_1, n_2 \in \mathbb{Z}$, or equivalently, by the two-dimensional wavelets $\Psi^l_{mn}$,

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

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 $\Psi^l_{mn}$, $l = 1, 2, 3$, $m \in \mathbb{Z}$, $n \in \mathbb{Z}^2$, 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]—

> "... 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, $(c_n)_{n\in\mathbb{Z}}$, 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,

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

or

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

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

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

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

$$c^0_n = d^0_n + \sum_k w(n-2k)c^1_k.$$

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 $F$: $l^2(\mathbb{Z}) \to l^2(\mathbb{Z})$ ($F$ 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. Gr
a. The tra
$w(n) = 0$ if $|n$
b. The tra

Here we u
$F$. Note th
that the d
and this c
The to
$L = 5$ or

(2.34)

From the

(2.35)

At ev
(2.35) the
direct an

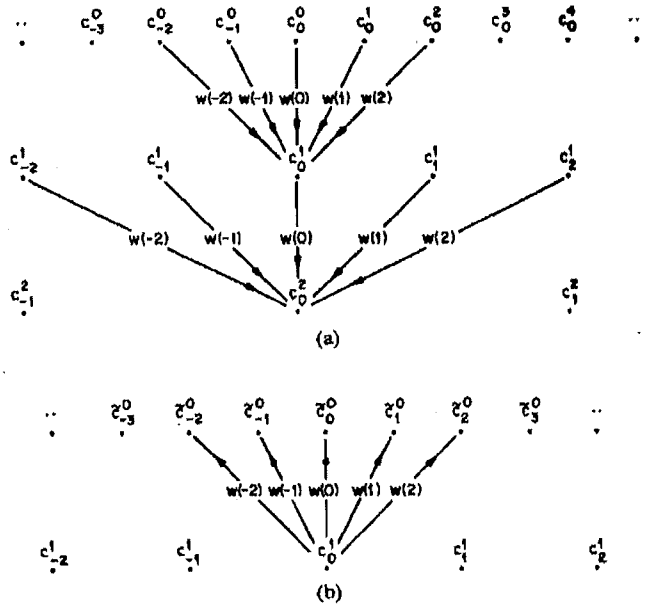


Figure 1. Graphical representation of the Laplacian pyramid scheme (redrawn from [20]).
a. The transition $c^0 \to c^1 \to c^2$. For simplicity's sake we have restricted ourselves to the case $w(n) = 0$ if $|n| > 2$, and only the computation of $c_0^1$ and $c_0^2$ are depicted.
b. The transition $c^1 \to \tilde{c}^0$.

Here we use the standard notation $F^*$ for the adjoint of the (bounded) operator $F$. Note that we implicitly assume that $c^0 \in l^2(\mathbf{Z})$, 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 $(F^*)^L l^2(\mathbf{Z})$, 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 $-\frac{1}{2} \leq x < \frac{1}{2}$, 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

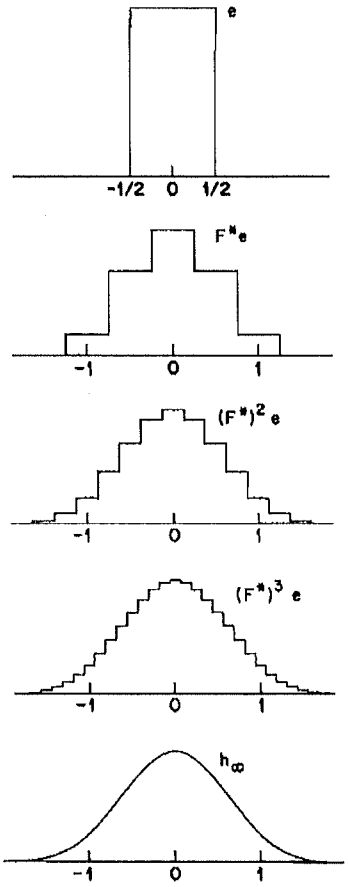


Figure 2. The successive sequences $e$, $F^*e$, $(F^*)^2e$, $(F^*)^3e$ represented by histograms, and the limit function $h_\infty$ (see text). We have taken the $w(m)$ as defined by (2.26), with $a = .375$.

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

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

and

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

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 $h_1$,

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

(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

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

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

Figure 3. ... b) $h_1$ c) $h_1$ ... function ... d) E... point as ... e) T...

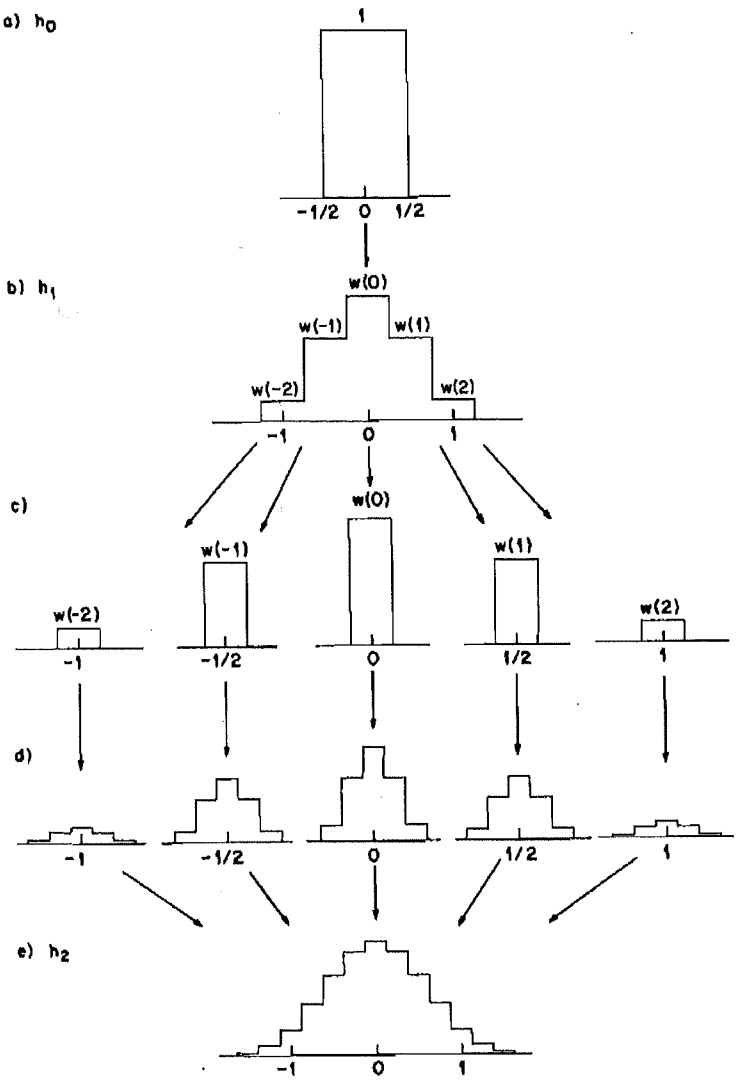


Figure 3. a) $h_0(x) = \chi_{[-1/2,1/2[}(x)$;

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

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:

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

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

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

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

$$h_j = \tilde{T}_j \cdots \tilde{T}_3 h_2 = \tilde{T}_j \cdots \tilde{T}_3 T h_1$$

$$= \tilde{T}_j \cdots \tilde{T}_3 T \tilde{T}_1 h_0 = \tilde{T}_j \cdots \tilde{T}_4 T \tilde{T}_2 \tilde{T}_1 h_0$$

$$= \cdots = T\tilde{T}_{j-1} \cdots \tilde{T}_1 h_0 = Th_{j-1},$$

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 $h_\infty$. 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 $h_\infty$ 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 $h_\infty$ (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 $h_\infty$, 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, $1 \leq p \leq \infty$.) 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 $\xi^{-1}$ in $\hat{h}_\infty$, i.e., to some regularity in $h_\infty$! Without this constraint, as can be easily checked, the graphical procedure in Figure 2 can lead to rather horrible (fractal) functions $h_\infty$ (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

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

For $.125 < a < .625$, which includes all the choices in [20], this implies that $h_\infty$ 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, $h_\infty$ is thus a fourth-order convolution of $\chi_{[0,1[}$ with itself, which results in a $C^{3-\epsilon}$ 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 $h_\infty$. Constructions similar to (2.44) will be used in Section 3, where the above "trick" for imposing regularity on $h_\infty$ 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 exploitcd, 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 $c^0 = (c_n^0)_n \in l^2(\mathbf{Z})$ 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 $\phi$ such that (2.1)–(2.4) are satisfied. We assume (if necessary, we apply (2.5) first) that the $\phi_{0n}$ are orthonormal. Let $\{\psi_{mn};\ m, n \in \mathbf{Z}\}$ 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 $\phi$ and $\psi$ are real, as they are in [8] and indeed in most practical examples.

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

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

or

$$f(x) = \sum_n c_n^0 \phi(x - n). \tag{2.47}$$

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 $W_1$,

$$f = P_1 f + Q_1 f. \tag{2.48}$$

Each of
bases $\phi_{1n}$

The sequ
while $d^1$
discussio
a functio
one has

where

This can

with

Note tha
coefficien

with

Each of these components can be expanded with respect to the orthonormal bases $\phi_{1n}$, $\psi_{1n}$, 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 $\phi_{1n}$ are orthonormal bases of $V_1$, one has

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

$$= \sum c_n^0 \langle \phi_{1k}, \phi_{0n} \rangle,$$

where

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

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

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., $\sum_n h(n) = \sqrt{2}$ (see subsection 3A for an explanation of the difference in normalization with the $w(n)$). The requirement $\sum_n h(2n) = \sum_n h(2n+1)$ 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, $\sum_n g(n) = 0$.

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 $l^2(\mathbf{Z})$ to itself,

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

$$(Ga)_k = \sum_n g(n-2k)a_n. \tag{2.49}$$

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

$$c_k^2 = \langle \phi_{2k}, P_2 f \rangle = \langle \phi_{2k}, P_1 f \rangle$$

$$= \sum_n c_n^1 \langle \phi_{2k}, \phi_{1n} \rangle.$$

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

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

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 $V^{j-1}$, $\{\phi_{j-1\,k};\ k \in \mathbf{Z}\} \to \{\phi_{jk}, \psi_{jk};\ k \in \mathbf{Z}\}$. 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 $\psi_{jk}, \phi_{jk}$ (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 $l^2(\mathbf{Z}^2)$,

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

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 $\Psi^j$, $j = 1, 2, 3$,—see (2.20)–(2.22)) $d^{1,1}$, $d^{1,2}$ and $d^{1,3}$,

$$c^1 = H_c H_r c^0,$$
$$d^{1,1} = G_c H_r c^0,$$
$$d^{1,2} = H_c G_r c^0,$$
$$d^{1,3} = G_c G_r c^0.$$

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

$$\sum_n |h(n)| < \infty, \qquad \sum_n |g(n)| < \infty. \tag{3.1}$$

This implies that the operators $H$, $G$, defined by

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

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 $\phi$. 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

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

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

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

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

for some small $\varepsilon$. 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

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

$$\sum_n h(n) = C. \tag{3.5}$$

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

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

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

But

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

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

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

$$\sum_n g(n) = 0. \tag{3.4}$$

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 $2\pi$-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)

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

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

$$\gamma(\xi) = \overline{\beta(\xi)}, \qquad \delta(\xi) = -\overline{\alpha(\xi)}. \tag{3.13}$$

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 $2\pi$-periodic functions $\alpha$ and $\beta$ 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 $2 \times 2$ matrix-valued $2\pi$-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 $\xi$.

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 $2\pi$-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 $\xi$. 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 $\nu$ 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'
$d^j = GH^j$

The iterat
2B) to stu
one non-z
(with hist
which ex
function.
To sh
histogram

Figure 4.
regularity

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

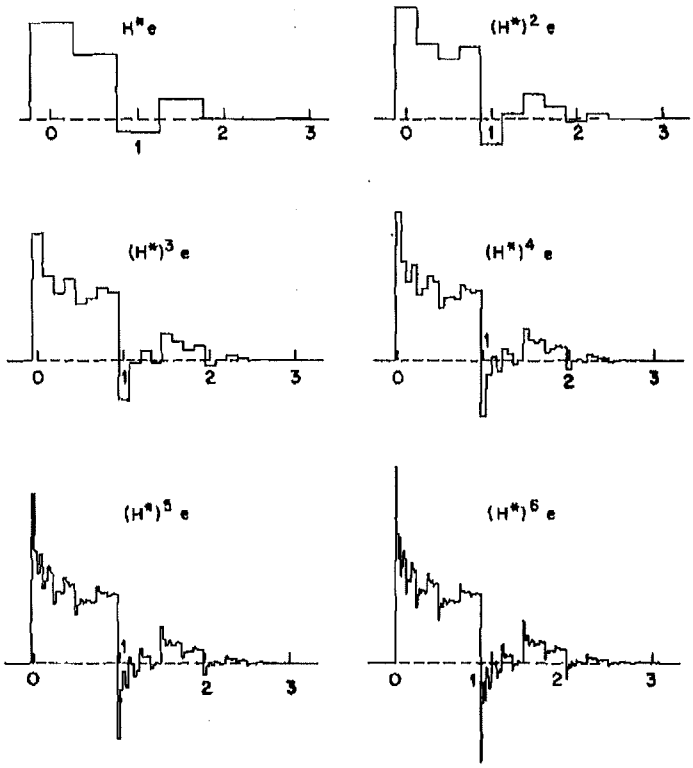


Figure 4. The histogram representations of $(H^*)^j e$, $j = 1, \cdots, 6$, for $h(n)$ which do not satisfy a regularity condition (see text).

(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 $\eta_l$ 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 $\eta_l$ 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 $|m_0(\xi) - 1| \leq \sqrt{2} \Sigma_n |h(n)|\,|\sin \tfrac{1}{2} n\xi|$. For any $0 < \delta \leq 1$ there exists $C_\delta$ 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

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