

Wavelets: The Mathematical Background

ALBERT COHEN AND JELENA KOVAČEVIĆ, MEMBER, IEEE

Invited Paper

We present here the mathematical foundations of the wavelet transform, multiresolution analysis and discrete-time transforms, and algorithms. This article serves as background material for the rest of the special issue.

I. INTRODUCTION

When we deal with a given physical object, we encounter many of its different faces or *representations*. For example, we can represent numbers in various systems depending on the application; in everyday life, we use the decimal system, while for use in computers we employ the binary representation. Consequently, in many fields, such as numerical analysis or signal processing, a preliminary task is to find an adapted representation of the signal that may be particularly suitable for a problem at hand. For example, in images, one of the common tasks is to attempt a representation that will facilitate extraction of features.

A way to obtain a specific representation is to *decompose* a signal x into elementary building blocks x_i , of some importance, as follows: $x = \sum_i x_i$, where the x_i are simple waveforms. Moreover, one may want that these waveforms have a specific "physical" interpretation. For example, in an image the blocks might correspond to textures and edges.

How do we practically decompose a signal? We need a *fast algorithm* in order to do it, since otherwise a particular decomposition/representation might be only of theoretical importance. Once we have the building blocks, we might attempt yet another task: *approximation*, that is, we try to get as good a rendition of the original signal as possible with only a few of the building blocks. The multiresolution concept, which will be defined later, allows us to do this in a natural way: we approach the original signal by successively adding details to it, that is, by successively refining it.

One of the classic tools to achieve such different representations of a signal is the Fourier theory, for which we have a whole arsenal of tools at our disposal: from

Manuscript received June 1, 1995; revised December 15, 1995.

A. Cohen is with Ceremade Université Pierre et Marie Curie, 75005 Paris, France.

J. Kovačević is with the Communications Analysis Research Department, AT&T Bell Laboratories, Murray Hill, NJ 07974 USA.

Publisher Item Identifier S 0018-9219(96)02575-3.

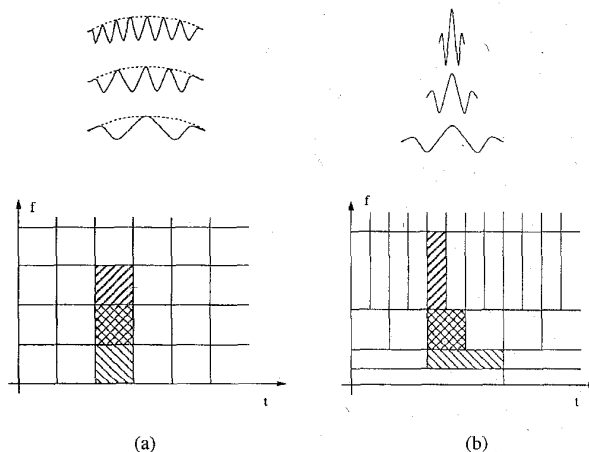


Fig. 1. Basis functions and corresponding tilings of the time-frequency plane [12]. (a) Short-time Fourier transform and (b) wavelet transform.

the purely continuous time, such as the Fourier integral, to discrete time and the fast algorithm to implement such a representation—fast Fourier transform (FFT) algorithm. Although the algorithms differ, the underlying mathematical ideas are the same for all these representations.

If we are given a pure frequency signal $e^{i\omega t}$, Fourier-based methods will isolate a peak at the frequency ω . However, already when confronted with the case of a signal built of two pure oscillations occurring in two adjacent intervals (that is, $e^{i\omega_1 t} \chi_{[a,b]}(t) + e^{i\omega_2 t} f_{[b,c]}(t)$), we run into problems: we obtain two peaks, without localization in time. This immediately points out to the need for a time-frequency representation of a signal which would give us local information in time and in frequency. In the Fourier case, it is obvious that we need a more local waveform to achieve this. The most intuitive way to overcome this obstacle is to localize the sinusoids in the Fourier representation by windowing, that is, the building blocks now become $w(t - \tau)e^{i\omega t}$, where $w(t)$ will denote a window with compact support allowing for time localization. We have thus obtained the windowed Fourier transform, or the short-time Fourier transform.

Let us take a look at what we have achieved; to that end, we will introduce the concept of the time-frequency plane, and we will loosely show where the building blocks reside in this plane. We will see that the building blocks used in different decomposition techniques “tile” the time-frequency plane in different ways. Note here that it is not possible to obtain arbitrarily fine localization in time and in frequency due to the uncertainty principle. Also note that we will always be looking for a representation or a decomposition that is linked to a fast algorithm so that the scheme is implementable. Fig. 1(a) shows an example of the time-frequency plane tiling for the short-time Fourier transform (STFT). The shaded squares in the figure correspond to waveforms which are localized in the same time interval $[T, T + \Delta T]$ and in three adjacent frequency levels.

In this representation we have fixed both the level of time and of the frequency localizations. Wavelets offer a different compromise: the frequency localization is logarithmic, that is, proportional to the frequency level. As a consequence time localization gets finer in the highest frequencies. Such a situation is given in Fig. 1(b).

Ultimately, one would like to obtain an “arbitrary” tiling of the time-frequency plane. The wavelet theory based on the multiresolution analysis (MRA) concept and its generalizations offer a natural way to achieve this. An example is given by wavelet packets, or arbitrary trees. A more detailed discussion of wavelet packets is given by Hess–Nielsen and Wickerhauser [1].

The outline of this paper is as follows: Section II gives an overview of the continuous and oversampled wavelet transform. Section III discusses how, by sampling the continuous wavelet transform, we obtain orthonormal wavelet bases. Multiresolution analysis, as a framework for studying wavelet bases, is also presented in this section. Finally, Section IV deals with discrete-time wavelet representations, filter banks, and fast algorithms.

II. CONTINUOUS AND OVERSAMPLED WAVELET TRANSFORM

The most natural way to obtain a time-scale representation as described in the previous section is to define a family of functions

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad a > 0, b \in \mathbb{R}$$

where ψ is a fixed function, called “mother wavelet,” that is well localized both in time and frequency, i.e., we can ask that

$$|\psi(t)| \leq c(1 + |t|)^{-1-\epsilon}, \quad |\hat{\psi}(\omega)| \leq c(1 + |\omega|)^{-1-\epsilon}$$

for some $\epsilon > 0$. Here, $\hat{\psi}(\omega) = \int_{-\infty}^{+\infty} \psi(t)e^{-i\omega t} dt$ is the Fourier transform of ψ . The factor $1/\sqrt{a}$ ensures that the functions $\psi_{a,b}$ have a constant norm in the space $L_2(\mathbb{R})$ of square integrable functions, that is, functions f such that

$$\|f\|^2 = \int_{-\infty}^{+\infty} |f(t)|^2 dt < +\infty. \quad (1)$$

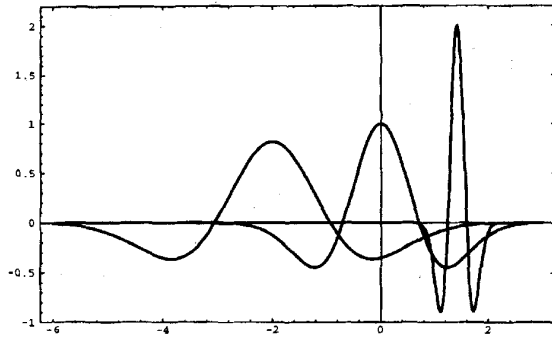


Fig. 2. A few wavelets obtained from the mother wavelet $\psi(t) = (1 - 2t^2)e^{-t^2} = \psi_{1,0}(t)$, which is the second derivative of a Gaussian. Displayed are (from left to right): $\psi_{(3/2),-2}(t)$, $\psi_{1,0}(t)$, $\psi_{(1/4),2^{1/2}}$.

We recall that this space is a Hilbert space, with the scalar product defined as

$$\langle f, g \rangle = \int f(t)\overline{g(t)} dt.$$

In this context, we define the continuous wavelet transform of a function $f \in L_2(\mathbb{R})$ as

$$Tf(a, b) = \langle f, \psi_{a,b} \rangle = \frac{1}{\sqrt{a}} \int f(t)\overline{\psi\left(\frac{t-b}{a}\right)} dt. \quad (2)$$

Here $\psi_{a,b}$ plays the same role as $e^{i\omega t}$ in the definition of the Fourier transform. The existence of an inverse transform depends on the choice of ψ . More precisely, if ψ is such that

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} < +\infty \quad (3)$$

then f can be reconstructed by

$$f(t) = C_\psi^{-1} \int_0^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} Tf(a, b)\psi_{a,b}(t) db \quad (4)$$

that is, the truncated integral

$$\int_{1/A}^A \frac{da}{a^2} \int_{-B}^B Tf(a, b)\psi_{a,b}(t) db$$

converges to $C_\psi f(t)$ in $L_2(\mathbb{R})$ as A and B go to $+\infty$. Note that condition (3) implies in particular that $\hat{\psi}(0) = 0$ so that ψ oscillates, that is, $\int \psi dt = \hat{\psi}(0) = 0$. More generally, one can impose more cancellations on ψ , in the sense that $[(d/d\omega)^k \hat{\psi}](0) = 0 = \int t^k \psi(t) dt$, for $k = 0, 1, \dots, N$.

Fig. 2 displays examples of wavelets obtained with the mother wavelet $\psi(t) = (1 - 2t^2)e^{-t^2}$, which is the second derivative of a Gaussian. In that case, one has $\int \psi(t) dt = \int t\psi(t) dt = 0$, that is, $N = 1$.

An important consequence of the cancellations of ψ is that the values of $Tf(a, b)$ are influenced by the regularity of the function f . To be more specific, for any $\alpha > 0$, we say that f is α -Hölder regular at some point t_0 if and only if there exists a polynomial $P(t)$ of degree $n < \alpha$, such that

$$|f(t) - P(t)| \leq C_1 |t - t_0|^\alpha. \quad (5)$$

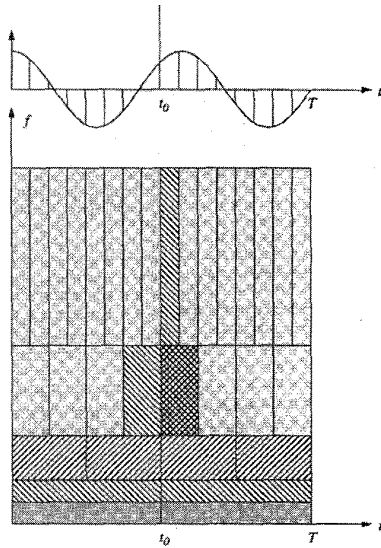


Fig. 3. Time-scale tiling for a smooth function with an isolated singularity. The vertical axis represents either increasing frequency (f) or decreasing scale ($1/a$).

Note that this property is satisfied in particular when f is m -times continuously differentiable in a neighborhood of t_0 with $m > \alpha$ (by taking $P(t) = \sum_{0 \leq k < \alpha} (d/dt)^k f(t_0) ((t - t_0)^k / k!)$). Note also that for $\alpha < 1$, one can only take $P(t) = f(t_0)$. If ψ has cancellations up to some order $N \geq \alpha - 1$ and sufficient decay at infinity so that $\int \psi(t) |t|^\alpha dt < +\infty$, we can use (5) to derive

$$\begin{aligned} |Tf(a, b)| &= |\langle f, \psi_{a,b} \rangle| = |\langle f - P, \psi_{a,b} \rangle| \\ &\leq C_1 \int_{-\infty}^{+\infty} |t - t_0|^\alpha |\psi_{a,b}(t)| dt \\ &= C_1 \sqrt{a} \int_{-\infty}^{+\infty} |ay + b - t_0|^\alpha |\psi(y)| dy \\ &\leq C_2 \sqrt{a} \int_{-\infty}^{+\infty} (|a|^\alpha |y|^\alpha + |b - t_0|^\alpha) |\psi(y)| dy. \end{aligned}$$

We see here that if we impose $|b - t_0| \leq C_3 a$, then we have $|Tf(a, b)| \leq K |a|^{\alpha + (1/2)}$. This shows that as the scale a goes to zero, the amplitude of $Tf(a, b)$ in the region $|b - t_0| \leq C a$ decays very fast if f is regular at t_0 , slower if f has some singularity at this point. This property of the wavelet transform is illustrated in Fig. 3; the darker regions of the time-scale tiling correspond to larger values of the wavelet transform.

In practical applications, in particular those involving fast algorithms, the continuous wavelet transform can only be computed on a discrete grid of points $(a_n, b_n)_{n \in \mathbb{Z}}$. How should one choose this sampling so that it contains all the information on the function f ?

A simple example gives us a hint: consider the case where ψ is simply defined by

$$\hat{\psi}(\omega) = \begin{cases} 1 & |\omega| \in [1, a_0] \\ 0 & \text{elsewhere} \end{cases}$$

for some $a_0 > 1$. Since, in that particular case, function f_a defined by $f_a(t) = Tf(a, t)$ is band limited by $\hat{\psi}(\omega)$ to $[-(a_0/a), -(1/a)] \cup [(1/a), (a_0/a)]$, it is completely determined by its samples $f_a(na/\pi(a_0 - 1))$, $n \in \mathbb{Z}$. Moreover, the frequency axis is tiled by the intervals $[-a_0^n, -a_0^{n-1}] \cup [a_0^{n-1}, a_0^n]$, $n \in \mathbb{Z}$, so that f can be recovered from the data of $f_n = f_{a_0^n}$, $n \in \mathbb{Z}$.

This simple case shows that a natural sampling for the continuous wavelet transform is given by $Tf(a_0^n, ma_0^n b_0)$, $n, m \in \mathbb{Z}$, for some fixed $a_0 > 1, b_0 > 0$. For a more general wavelet ψ , we can now define

$$\psi_{n,m}(t) = a_0^{n/2} \psi(a_0^n t - b_0 m), \quad n, m \in \mathbb{Z}$$

where we have fixed the parameters $a_0 > 1$ and $b_0 > 0$. Given such a family, two questions are of importance:

- Does the sequence $(\langle f, \psi_{n,m} \rangle)_{n,m \in \mathbb{Z}}$ completely characterize the function f ?
- Is it possible to recover f from this sequence in a stable manner?

To answer these questions, it is necessary to introduce the concept of a *frame* that we briefly review in an abstract setting.

A sequence $(e_n)_{n \in \mathbb{Z}}$ in a Hilbert space H is called a frame, if and only if, for all $x \in H$, one has

$$A \|x\|^2 \leq \sum_{n \in \mathbb{Z}} |\langle x, e_n \rangle|^2 \leq B \|x\|^2 \quad (6)$$

where the frame bounds $B \geq A > 0$ are independent of x . To such a sequence, we associate the "frame operator" F that maps any $x \in H$ into a square-summable sequence $(\langle x, e_n \rangle)_{n \in \mathbb{Z}}$ and its dual operator F^* that maps any square-summable sequence $(x_n)_{n \in \mathbb{Z}}$ into $x = \sum_{n \in \mathbb{Z}} x_n e_n$.

The inequalities in (6) can be directly expressed in terms of the positive hermitian operator F^*F

$$A \text{Id} \leq F^*F \leq B \text{Id} \quad (7)$$

in the sense that $A \langle x, x \rangle \leq \langle F^*F x, x \rangle \leq B \langle x, x \rangle$ for all $x \in H$. This shows in particular that F^*F is invertible so that we can define a sequence $(\tilde{e}_n)_{n \in \mathbb{Z}}$ by $\tilde{e}_n = (F^*F)^{-1} e_n$.

Since we have

$$\langle x, \tilde{e}_n \rangle = \langle x, (F^*F)^{-1} e_n \rangle = \langle (F^*F)^{-1} x, e_n \rangle$$

it follows that $(\tilde{e}_n)_{n \in \mathbb{Z}}$ also constitutes a frame (it is called the *dual frame*) with bounds $A^{-1} \geq B^{-1} > 0$ and the associated frame operator \tilde{F} satisfies $\tilde{F} = F(F^*F)^{-1}$ so that we have

$$\tilde{F}^*F = (F(F^*F)^{-1})^* F = (F^*F)^{-1} (F^*F) = \text{Id}$$

and similarly $F^*\tilde{F} = \text{Id}$. This means that any $x \in H$ can be expressed as

$$x = \sum_{n \in \mathbb{Z}} \langle x, e_n \rangle \tilde{e}_n = \sum_{n \in \mathbb{Z}} \langle x, \tilde{e}_n \rangle e_n.$$

To reconstruct x , it is thus sufficient to know the dual frame $(\tilde{e}_n)_{n \in \mathbb{Z}}$, that is, to invert F^*F . This can be done by

a fast algorithm, remarking that we have

$$F^*F = \frac{A+B}{2}(\text{Id} - R)$$

where the residual operator $R = \text{Id} - 2/A + BF^*F$ satisfies

$$\|R\| \leq \frac{A-B}{A+B} < 1$$

according to (7). We thus have

$$(F^*F)^{-1} = \frac{2}{A+B}(I + R + R^2 + \dots) \quad (8)$$

meaning that we can approximate $(F^*F)^{-1}$ by truncating this series at a certain order depending on the desired accuracy. Note that the convergence of (8) is very fast when the frame bound are "close" in the sense that $A - B \ll A + B$.

We now return to the particular case of the wavelet family $(\psi_{n,m})_{n,m \in \mathbb{Z}}$. In that case, a criterion for the existence of frame bounds was derived by Daubechies [2] under the assumptions

$$0 < c_1 \leq \sum_{n \in \mathbb{Z}} |\hat{\psi}(a_0^n \omega)|^2 \leq c_2 < +\infty$$

and

$$\beta(\nu) = \sup_{\omega \in \mathbb{R}} \sum_{n \in \mathbb{Z}} |\hat{\psi}(a_0^n \omega) \hat{\psi}(a_0^n \omega + \nu)| \leq C(1 + |\nu|)^{-1-\epsilon}$$

for some $\epsilon > 0$. The frame bounds are then given by

$$A = \frac{2\pi}{b_0} \left\{ c_1 - \sum_{k \in \mathbb{Z} \setminus \{0\}} \left[\beta\left(\frac{2\pi}{b_0}k\right) \beta\left(\frac{2\pi}{b_0}k\right) \right]^{1/2} \right\}$$

$$B = \frac{2\pi}{b_0} \left\{ c_2 + \sum_{k \in \mathbb{Z} \setminus \{0\}} \left[\beta\left(\frac{2\pi}{b_0}k\right) \beta\left(\frac{2\pi}{b_0}k\right) \right]^{1/2} \right\}$$

under the condition that A is strictly positive. One can check that this is always the case when b_0 is smaller than a certain threshold b_t (recall that $c_1 > 0$). Moreover the ratio $(A - B)/(A + B)$ tends to zero when $a_0 \rightarrow 1$ and $b_0 \rightarrow 0$.

Note that the dual frame $\tilde{\psi}_{n,m} = (F^*F)^{-1}\psi_{n,m}$ will not be generated by the translates and dilates of a mother function $\tilde{\psi}$, in general. However, in the case where $A = B$ we simply have $\tilde{\psi}_{n,m} = A^{-1}\psi_{n,m}$. By renormalizing the $\psi_{n,m}$, we obtain $A = B = 1$ and thus for any $f \in L_2(\mathbb{R})$, we have

$$f = \sum_{m,n \in \mathbb{Z}} \langle f, \psi_{n,m} \rangle \psi_{n,m}. \quad (9)$$

In that case $\{\psi_{n,m}\}_{n,m \in \mathbb{Z}}$ is called a *tight frame*. Although the expansion (9) is similar to the decomposition of f in an orthonormal basis, it is redundant in general. In particular, we have remarked that $(A - B)/(A + B)$ can be made arbitrarily close to zero, and thus A and B arbitrarily close to one after renormalization, by choosing a_0 and b_0 close to one and zero, respectively. This precisely means that we oversample the continuous wavelet transform on a very dense grid.

Is there any specific advantage to such an oversampling, in comparison to the decomposition in an orthonormal basis? The following simple example illustrates an important advantage of redundancy: consider in \mathbb{R}^2 the basis $e_1 = (1, 0)$, $e_2 = (0, 1)$ and the tight frame consisting of the vectors $f_1 = e_1/\sqrt{2}$, $f_2 = e_2/\sqrt{2}$, $f_3 = (e_1 + e_2)/2$, and $f_4 = (e_1 - e_2)/2$. Any vector x can thus be written

$$x = \langle x, e_1 \rangle e_1 + \langle x, e_2 \rangle e_2 \quad (10)$$

or

$$x = \langle x, f_1 \rangle f_1 + \langle x, f_2 \rangle f_2 + \langle x, f_3 \rangle f_3 + \langle x, f_4 \rangle f_4. \quad (11)$$

Assume now that x is known up to a noise described as a random variable $N(\omega)$ with a Gaussian, centered, and radial distribution. The resulting error on the samples $\langle x, e_i \rangle$ (or $\langle x, f_i \rangle$) is a centered Gaussian scalar variable of variance V (or $V/2$). If these samples are evaluated independently, the mean squared error of the reconstruction by (10) is given by

$$\epsilon = E(\|\eta_1 e_1 + \eta_2 e_2\|^2) = 2V$$

whereas, in the case of (11), we obtain

$$\epsilon = E(\|\eta_1 f_1 + \eta_2 f_1 + \eta_3 f_1 + \eta_4 f_1\|^2) = V.$$

This example reveals one of the main interest of oversampling with a frame: the reconstruction mean square error due to noise can be reduced by a factor which is precisely the oversampling rate. We now turn to the construction of nonoversampled wavelet families, that is, wavelet bases.

III. MULTIREOLUTION ANALYSIS AND WAVELET BASES

By adding more restrictions on the sampling parameters a_0 and b_0 , as well as on the choice of the wavelet ψ , it is possible to remove the redundancy in the reconstruction formula (4), so that it may be regarded as the expansion of f in a basis. Since functions usually live in infinite dimensional spaces, statements such as the L_2 -stability of a basis $\{e_n\}_{n \in \mathbb{N}}$, that is, the existence of two constants $C \geq c > 0$ such that

$$c \sum_n |a_n|^2 \leq \int \left| \sum_n a_n e_n(t) \right|^2 dt \leq C \sum_n |a_n|^2 \quad (12)$$

independently of the choice of the coefficients a_n , have important consequences on the numerical applications of these expansions.

However, note that (12) is immediately satisfied in the case of an orthonormal basis (one has then an equality with $c = C = 1$). Such bases can be obtained for the particular choice $a_0 = 2$ and $b_0 = 1$, that is

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad j, k \in \mathbb{Z}. \quad (13)$$

The oldest example of such bases is certainly the Haar system, where the wavelet ψ is a piecewise constant function such that $\psi(t) = 1$ on $[0, 1/2[$, -1 on $[1/2, 1[$ and zero elsewhere. The proof that this particular system constitutes an orthonormal basis is interesting in itself, since it contains several important features of the general theory.

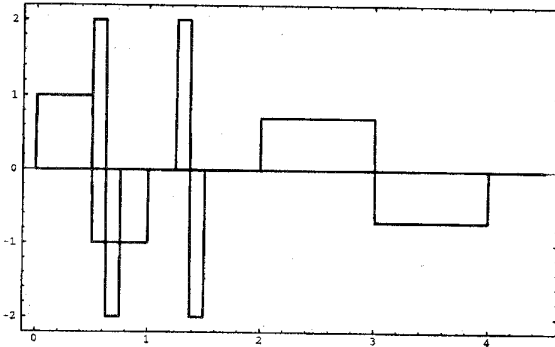


Fig. 4. A few of the Haar basis functions. From left to right: "Mother wavelet" $\psi_{0,0}(t) = \psi(t)$, $\psi_{2,2}(t)$, its shifted version $\psi_{2,5}(t)$ and $\psi_{-1,1}(t)$.

First, one checks that these functions are orthonormal with respect to the scalar product $\langle f, g \rangle = \int f(t)g(t) dt$. Indeed, they are normalized by the factor $2^{j/2}$; for a fixed j the functions $\psi_{j,k}$ have nonoverlapping support and are thus trivially orthogonal; for $j' > j$, the wavelet $\psi_{j,k}$ has a constant value along the support of $\psi_{j',k'}$ so that the integral of their product is zero. Fig. 4 displays some of these different possible situations.

In order to check that these functions constitute a basis, it remains to show that any function f in $L_2(\mathbb{R})$ can be expanded as a combination of $\psi_{j,k}$ of the type

$$f = \sum_{j,k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}. \quad (14)$$

At this point, it is useful to introduce a piecewise constant approximation $A_j f$ of f given by the average $a_{j,k}$ of f on each dyadic interval $I_{j,k} = [2^{-j}k, 2^{-j}(k+1))$, $k \in \mathbb{Z}$, that is

$$a_{j,k} = A_j f|_{I_{j,k}} = 2^j \int_{I_{j,k}} f(t) dt.$$

We remark that the average $a_{j,k}$ is also the average of the averages on the two half intervals, that is $(a_{j+1,2k} + a_{j+1,2k+1})/2$. The reduction to $I_{j,k}$ of the difference between $A_{j+1}f$ and $A_j f$ is therefore a multiple of $\psi_{j,k}$; from this, it is clear that the details between two successive levels of approximation can be expanded in terms of wavelets, that is

$$A_{j+1}f - A_j f = \sum_k d_{j,k} \psi_{j,k}$$

as illustrated in Fig. 5. It is also easy to check, from the definition of $A_j f$, that we have $d_{j,k} = \langle f, \psi_{j,k} \rangle$.

For any j_1 , we can thus obtain $A_{j_1} f$ by taking a coarser approximation $A_{j_0} f$, $j_0 < j_1$, and adding a combination of wavelets at intermediate scales, that is

$$A_{j_1} f = A_{j_0} f + \sum_{j=j_0}^{j_1-1} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}. \quad (15)$$

The expansion (14) follows from (15) by letting j_1 or j_0 go to $+\infty$ or $-\infty$, respectively. However it is important

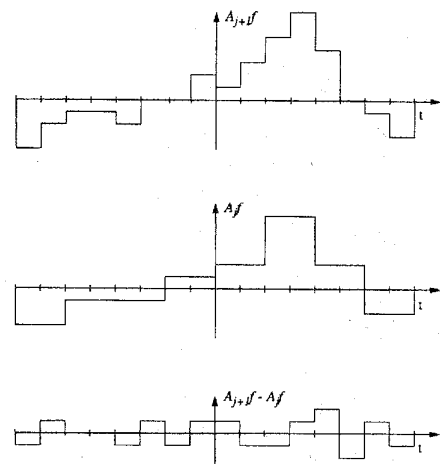


Fig. 5. Pictorial proof that the Haar system is an orthonormal basis. $A_j f$ is a piecewise constant approximation of f on the interval $[2^{-j}k, 2^{-j}(k+1)]$, and similarly for $A_{j+1} f$, except on intervals half the size. The difference between two approximations, $A_{j+1} f - A_j f$, is a linear combination of Haar basis functions.

to remark that in most practical applications, one is only interested in a decomposition of the type (15) that is, within a finite range of scales: on one side, it is of no use to go to very coarse scale when the function to be analyzed has a limited support; on the other side, one mostly deals with sampled data which can express the approximation of a function (for example, in some applications of numerical analysis of partial differential equations) or purely discrete information (for example, in digital signal processing), so that j_1 is also finite.

The functions of the Haar system suffer from a major disadvantage for many applications: as they are discontinuous and cannot provide a good approximation for smooth functions, it is often preferable to use schemes in which the derivatives of $A_j f$ approximate those of f . One of the main goals of the theory of wavelet bases is to construct systems that have the same multiscale structure as the Haar basis, but that are generated from more regular functions, so that they allow better approximation. A direct method (although somewhat *ad hoc*) is due to Meyer [3] and consists in designing directly the wavelet ψ in such a way that orthonormality and completeness are achieved for the system (13). Another strategy, which seems more relevant for numerical applications, consists in generalizing the multiresolution approximation operators A_j by introducing a new concept that we now describe.

We call *multiresolution analysis* a sequence of approximation subspaces $\{V_j\}_{j \in \mathbb{Z}}$ of $L_2(\mathbb{R})$ such that the following requirements are satisfied:

- 1) The V_j are generated by a *scaling function* $\varphi \in L_2(\mathbb{R})$, in the sense that, for each fixed j , the family

$$\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k) \quad k \in \mathbb{Z}$$

spans the space V_j and satisfies the L_2 stability condition (12) (that is, the $\varphi_{j,k}$, $k \in \mathbb{Z}$, constitute a Riesz basis for V_j).

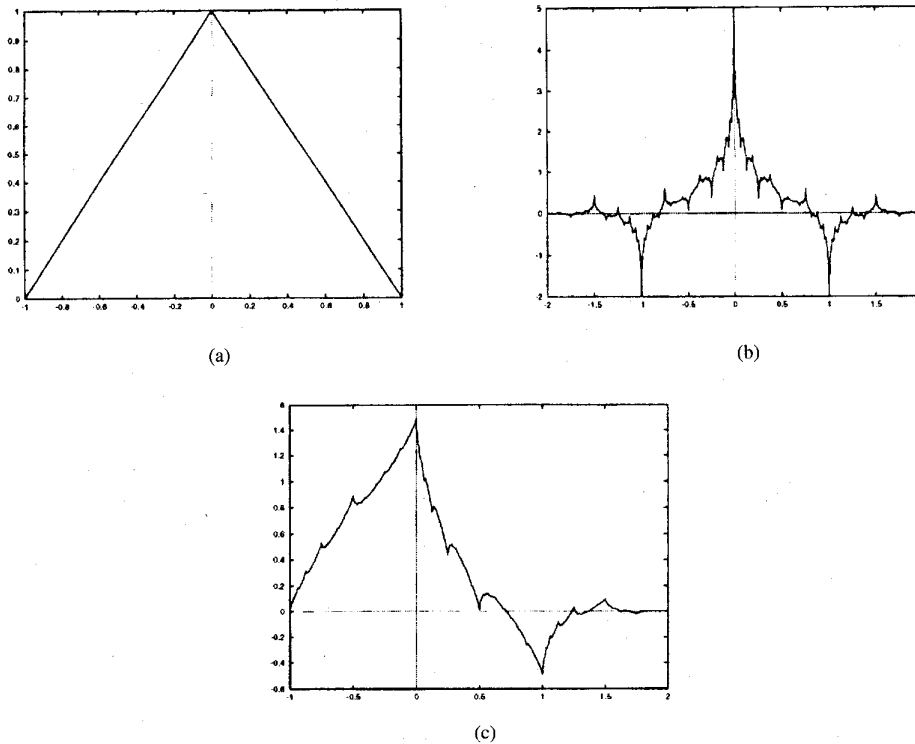


Fig. 6. Three examples of scaling functions. (a) Scaling function that generates approximation spaces of piecewise linear continuous functions. (b) Dual scaling function of the scaling function in (a). (c) Example of an orthonormal scaling function [10].

- 2) The spaces are embedded, that is, $V_j \subset V_{j+1}$.
- 3) The orthogonal projectors P_j onto V_j satisfy $\lim_{j \rightarrow +\infty} P_j f = f$ and $\lim_{j \rightarrow -\infty} P_j f = 0$ for all $f \in L_2(\mathbb{R})$.

From this definition, it follows clearly that $f \in V_j$ is equivalent to $f(2^j \cdot) \in V_{j+1}$ and that V_j is invariant under translation of 2^{-j} . From 2), we also see that φ is the solution of a two-scale equation

$$\varphi(t) = 2 \sum_{n \in \mathbb{Z}} h[n] \varphi(2t - n). \quad (16)$$

These spaces can be used to build approximation operators that generalize the previous A_j . In the case where the integer translates of φ are orthonormal, we can take the orthogonal projection

$$A_j f = \sum_{k \in \mathbb{Z}} \langle f, \varphi_{j,k} \rangle \varphi_{j,k}.$$

Note that this is the case for the piecewise constant approximation, with $\varphi = 1$ on $[0, 1]$, and zero elsewhere.

In the case where they only constitute a Riesz basis, one can either orthonormalize them and define A_j in the same way, or construct a dual scaling function $\tilde{\varphi}$ such that $\langle \tilde{\varphi}, \varphi(\cdot - k) \rangle = \delta_{0,k}$ and define an “oblique projector”

$$A_j f = \sum_{k \in \mathbb{Z}} \langle f, \tilde{\varphi}_{j,k} \rangle \varphi_{j,k}.$$

In this more general framework, it is important to require that the dual scaling function also satisfies a scaling equation

$$\tilde{\varphi}(t) = 2 \sum_{n \in \mathbb{Z}} \tilde{h}[n] \tilde{\varphi}(2t - n). \quad (17)$$

Fig. 6 shows three examples of scaling functions: φ_1 generates approximation spaces of piecewise linear continuous functions, but its integer shifts are not orthonormal. φ_2 is a dual scaling function for φ_1 , that is, $\langle \varphi_1(\cdot - k), \varphi_2(\cdot - \ell) \rangle = \delta_{k,\ell}$. (Note that there are other dual scaling functions for the same φ_1). Finally φ_3 is an example of orthonormal scaling function, that is, $\langle \varphi_3(\cdot - k), \varphi_3(\cdot - \ell) \rangle = \delta_{k,\ell}$.

It is clear that φ_1 can be defined explicitly as

$$\varphi_1(t) = \max\{0, 1 - |t|\}.$$

In contrast, φ_2 and φ_3 have no explicit expression: they are defined as solutions of the scaling equation (16) with specific choices for the coefficients $h[n]$. This approach is very relevant for practical applications because only these coefficients are used in the decomposition and reconstruction algorithms that we describe in the next section.

Moreover, several properties of the scaling function can be prescribed by imposing constraints on the coefficients:

- φ will be real valued and compactly supported if and only if the nonzero $h[n]$'s are real and finitely many, respectively. Moreover, the length of the sequence $h[n]$ equals the size of the support of φ .

- A necessary condition for the orthonormality of φ is

$$2 \sum_{n \in \mathbb{Z}} h[n] \overline{h[n+2k]} = \delta_{k,0}$$

and a necessary condition for the duality of φ and $\tilde{\varphi}$ is

$$2 \sum_{n \in \mathbb{Z}} \tilde{h}[n] \overline{h[n+2k]} = \delta_{k,0}.$$

Moreover, these conditions become sufficient under some additional technical assumptions [2], [4].

- If $h[n]$ have sufficient decay (for example, $|h[n]| < c2^{-\alpha n}$), a necessary condition for φ to be N -times differentiable is that the 2π -periodic function $m_0(\omega) = \sum_{n \in \mathbb{Z}} h[n] e^{in\omega}$ has a zero of order $N+1$ at $\omega = \pi$. Thus we can factorize $m_0(\omega)$ as

$$m_0(\omega) = \left(\frac{1 + e^{i\omega}}{2} \right)^{N+1} p(\omega) \quad (18)$$

where $p(\omega)$ is a 2π -periodic smooth function.

Conversely, an important problem is to recover the regularity properties of φ from the properties of $h[n]$. Note that when $p(\omega) = 1$, the choice $m_0(\omega) = (1 + e^{i\omega}/2)^{N+1}$ corresponds to the basic spline scaling function $\varphi_{N+1} = (*)^{N+1} \chi_{[0,1]}$ which is in C^{N-1} and satisfies $|\varphi_{N+1}^{(N-1)}(t_1) - \varphi_{N+1}^{(N-1)}(t_2)| \leq c_{N+1}|t_1 - t_2|$. For more general cases, one can estimate the exact regularity of φ through a careful study of the residual factor $p(\omega)$. Note also that the scaling function imposes the normalization $\sum_n h[n] = m_0(0) = p(0) = 1$.

Once scaling functions have been constructed, the derivation of wavelets is straightforward. One simply defines

$$g[n] = (-1)^n \tilde{h}[1-n], \tilde{g}[n] = (-1)^n h[1-n]$$

and

$$\psi(t) = 2 \sum_{n \in \mathbb{Z}} g[n] \varphi(2t-n), \quad \tilde{\psi}(t) = \sum_{n \in \mathbb{Z}} \tilde{g}[n] \tilde{\varphi}(2t-n). \quad (19)$$

In the orthonormal case, one simply has $\varphi = \tilde{\varphi}$, $h[n] = \tilde{h}[n]$, $g[n] = \tilde{g}[n]$ and $\psi = \tilde{\psi}$. With these definitions (19), a direct computation shows that

$$A_{j+1}f - A_jf = \sum_{k \in \mathbb{Z}} \langle f, \tilde{\psi}_{j,k} \rangle \psi_{j,k}$$

so that we obtain multiscale decomposition that are similar to (15) and (14). Moreover, one has

$$\langle \psi_{j,k}, \tilde{\psi}_{j',k'} \rangle = \delta_{j,j'} \delta_{k,k'}$$

and the whole system $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$ constitutes a Riesz basis of $L_2(\mathbb{R})$ [4].

We end this section by mentioning the most frequently used generalizations of the previous concept to the analysis of multivariate functions:

- The full tensor-product basis can be defined as

$$\Psi_{k_1, \dots, k_n}^{j_1, \dots, j_n}(t_1, \dots, t_n) = \psi_{j_1, k_1}(t_1) \cdots \psi_{j_n, k_n}(t_n), \\ j_1, \dots, j_n, k_1, \dots, k_n \in \mathbb{Z}.$$

This basis is highly nonisotropic since its elements may contain very different scales in different directions.

- The tensor-product multiresolution analysis V_j generated by the scaling functions

$$\Phi_{j, k_1, \dots, k_n}(t_1, \dots, t_n) = \varphi_{j, k_1}(t_1) \cdots \varphi_{j, k_n}(t_n) \\ k_1, \dots, k_n \in \mathbb{Z}$$

leads to a different wavelet basis

$$\Psi_{j, k_1, \dots, k_n}^\epsilon(t_1, \dots, t_n) = \psi_{j, k_1}^{\epsilon_1}(t_1) \cdots \psi_{j, k_n}^{\epsilon_n}(t_n) \\ \epsilon \in \{0, 1\}^n - (0, \dots, 0) \\ k_1, \dots, k_n \in \mathbb{Z}$$

where we have set $\psi^0 = \varphi$ and $\psi^1 = \psi$. In that context, each function Ψ^ϵ is used to characterize the details in certain directions.

- Finally one can generalize the concept of multiresolution analysis by replacing the dilation factor two by a dilation matrix D with integer entries and eigenvalues $\lambda_1, \dots, \lambda_n$ that satisfy $|\lambda_i| > 1$. In that case, it is easy to see that one needs $|\det D| - 1$ wavelets to characterize the details. However, one has to generalize in a nontrivial way the analysis of orthonormality, biorthogonality and smoothness that was done in the 1-D case [5], [6].

IV. DISCRETE-TIME WAVELET REPRESENTATIONS AND FAST ALGORITHMS

We have seen in the previous section how, by adding more restrictions on the sampling parameters a_0 and b_0 as well as on the choice of the wavelet ψ , we obtained a nonredundant representation through an orthonormal or biorthogonal basis. However, the question of computation still remains open. Mallat proposed an efficient discrete-time algorithm for the computation of the wavelet series given in Section III [7]. We assume that the multiresolution analysis axioms hold and we start with a function $f(t)$ in V_j

$$f(t) = \sum_{n=-\infty}^{+\infty} f^{(j)}[n] \varphi(2^j t - n)$$

where $f^{(j)}[n] = 2^j \langle f(t), \tilde{\varphi}(2^j t - n) \rangle$, $n \in \mathbb{Z}$. Using (16) and (19) we obtain that the projections on V_{j-1} and W_{j-1} , that is, $f^{(j-1)}[n]$ and $d^{(j-1)}[n]$, are given by

$$f^{(j-1)}[n] = \sum_k \tilde{h}[k-2n] f^{(j)}[k]$$

$$d^{(j-1)}[n] = \sum_k \tilde{g}[k-2n] f^{(j)}[k].$$

In signal processing terms, the coefficients of the projections onto V_{j-1} and W_{j-1} , are obtained by filtering by $\tilde{h}[-n]$ and $\tilde{g}[-n]$ and downsampling by two. Repeating the process on $f^{(j-1)}[n]$, we obtain projections onto V_{j-2} and W_{j-2} , and so on. The reconstruction algorithm can be derived from (16) and (19)

$$f^{(j+1)}[n] = 2 \sum_k h[n-2k] f^{(j)}[k] + 2 \sum_k g[n-2k] d^{(j)}[k]. \quad (20)$$

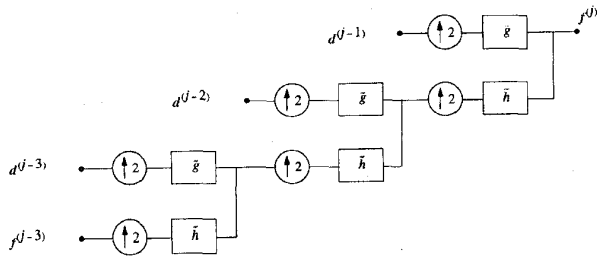


Fig. 7. The reconstruction part of Mallat's algorithm and its filter bank implementation.

Here, the sequences $f^{(j)}$ and $d^{(j)}$ have been refined by inserting zeros between each two samples and filtering with $h[n]$ and $g[n]$. The whole process can be depicted as in Fig. 7. (Note that in Section III we used notation $d_{j,k}$. Here we use $d^{(j)}[k]$ to be consistent with the discrete-time notation.)

In the signal processing literature, the two branches with filtering followed by downsampling by two together with upsampling by two and filtering, are termed a *filter bank*. Thus Fig. 7 gives a filter-bank implementation for the Mallat's algorithm. This was one of the key connections between the wavelet theory and multiresolution filter bank schemes studied in the digital signal processing literature [8]. Another link was provided by recognizing that the Laplacian pyramid scheme of Burt and Adelson [9] could also be seen as a vehicle for the multiresolution analysis, since it involved a hierarchy of averages as well as their differences. The final piece was found when Daubechies [10] and Mallat [11] showed how to go in the other direction as well; that is, the possibility of constructing wavelet bases starting from the discrete filters. Specifically, if we iterate the reconstruction algorithm on an initial Kronecker delta sequence $f^{(0)}[k] = \delta_{0,k}$ at the scale $j = 0$, we obtain at each step a sequence that is sampled on a twice finer grid. Such a refinement is also used in computer aided geometric design as a "subdivision scheme." The question now is under which conditions this process converges to a limit function, and when it does, whether this limit function is regular or not. A filter having the above properties is termed *regular*. Daubechies [2], [10] gives a sufficient condition for regularity; if the lowpass filter has a certain number of zeros at the aliasing frequency π , that is, we have the factorization (18), and if the remainder factor $p(\omega)$ is well behaved, then the process will converge to a regular scaling function $\varphi(t)$. The wavelet $\psi(t)$ is then easily obtained using (19). There exist various methods for designing regular filters that would lead to bases with a certain regularity index [4], [10], [12] as well as regularity testing procedures [10], [13]. Most of them use available tools for imposing a sufficiently high number of zeros at aliasing frequencies [8].

We can now formalize the above procedure as the *discrete-time wavelet series* (sometimes also called *discrete-time wavelet transform*) [14], [15]. We will define the discrete-time wavelet series over J octaves and the dyadic

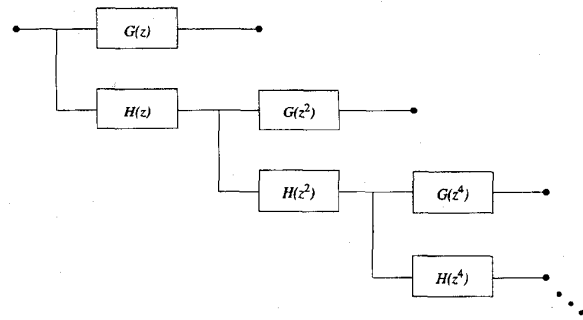


Fig. 8. Implementation of the "algorithme à trous."

sampling grid as

$$x[n] = \sum_{j=1}^J \sum_{k \in \mathbb{Z}} y^{(j)}[k] \tilde{g}^{(j)}[n - 2^j k] + \sum_{k \in \mathbb{Z}} X^{(J)}[k] \tilde{h}^{(J)}[n - 2^J k] \quad (21)$$

where

$$X^{(j)}[2k + 1] = \langle g^{(j)}[2^j k - l], x[l] \rangle, \quad j = 1, \dots, J \quad (22)$$

$$X^{(j)}[2k] = \langle h^{(j)}[2^j k - l], x[l] \rangle \quad (23)$$

and $\tilde{h}[n] = h[-n]$, $\tilde{g}[n] = g[-n]$. Here, $h^{(j)}[n]$ is the equivalent filter in the lowpass branch after j steps of iteration. This tells us that the family of "discrete wavelets" $\{\tilde{g}^{(j)}[2^j k - \cdot], \tilde{h}^{(j)}[2^j k - \cdot], j = 1, \dots, J, \text{ and } k \in \mathbb{Z}, \text{ is an orthonormal basis for } l_2(\mathbb{Z})$ [14], [15].

Note that until now, the schemes we presented were all time variant. What if we need a time-invariant system? A way to obtain it is to compute all the integer shifts, that is, avoid downsampling. We do obtain a time-invariant scheme, however, we pay a price; our system is oversampled now. Fig. 8 shows the resulting scheme. At scale i , the z -transform of the equivalent filter is

$$G_i(z) = G(z^{2^{i-1}}) \prod_{k=0}^{i-2} H(z^{2^k}).$$

This algorithm was named "algorithme à trous" (algorithm with holes) [16] and owes its name to the fact that one can take advantage of the filters being upsampled, that is, having zeros in between their samples.

An interesting generalization of these filter bank trees is that of *wavelet packets*, which can be described as arbitrary tree-structured filter banks, producing more general tilings of the time-frequency plane [1], [8].

REFERENCES

- [1] N. Hess-Nielsen and M. V. Wickerhauser, "Wavelets and time-frequency analysis," *Proc. IEEE*, this issue, pp. xxx-xxx.
- [2] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.

- [3] Y. Meyer, *Ondelettes et Opérateurs*, vol. 1. Paris: Hermann, 1990.
- [4] A. Cohen, I. Daubechies, and J. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. on Pure and Appl. Math.*, vol. 45, pp. 485–560, 1992.
- [5] A. Cohen and I. Daubechies, "Non-separable bidimensional wavelet bases," *Rev. Mat. Iberoamericana*, vol. 9, no. 1, pp. 51–138, 1993.
- [6] J. Kovačević and M. Vetterli, "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for \mathcal{R}^n ," *IEEE Trans. Inform. Theory, special issue on Wavelet Transforms and Multiresolution Signal Analysis*, vol. 38, pp. 533–555, Mar. 1992.
- [7] S. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 37, pp. 2091–2110, Dec. 1989.
- [8] K. Ramchandran and M. Vetterli, "Wavelets, subband coding and best bases," *Proc. IEEE*, this issue, pp. xxx–xxx.
- [9] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, pp. 532–540, Apr. 1983.
- [10] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure and Appl. Math.*, vol. 41, pp. 909–996, Nov. 1988.
- [11] S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathcal{R})$," *Trans. Amer. Math. Soc.*, vol. 315, pp. 69–87, Sept. 1989.
- [12] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Process.*, vol. 40, pp. 2207–2232, Sept. 1992.
- [13] O. Rioul, "Simple regularity criteria for subdivision schemes," *SIAM J. Math. Anal.*, vol. 23, pp. 1544–1576, Nov. 1992.
- [14] ———, "A discrete-time multiresolution theory," *IEEE Trans. Signal Process.*, vol. 41, pp. 2591–2606, Aug. 1993.
- [15] M. Vetterli and J. Kovačević, "Signal processing," in *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [16] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with

the help of the wavelet transform," in *Wavelets, Time-Frequency Methods and Phase Space*. Berlin: Springer, IPTI, 1989, pp. 289–297.



Albert Cohen was born in Paris, France, in 1965. He graduated from the Ecole Polytechnique, Paris, in 1987 and received the Ph.D. degree in applied mathematics from the Université Paris-IX Dauphine in 1990.

During 1990–1991 he was a Postdoctoral Researcher in the Mathematics Research Center of AT&T Bell Laboratories, Murray Hill, NJ. From 1992 to 1995 he had a teaching position at the Ecole Nationale Supérieure des Techniques Avancées, Paris, while doing research work at Ceremade (Université Paris-IX Dauphine). In 1995 he became a Professor of the Numerical Analysis Department at the Université Pierre et Marie Curie, Paris. His research interests include approximation theory, numerical analysis, signal and image processing, inverse problems, and statistics. He is the coauthor (with R. Ryan) of *Wavelets and Multiscale Signal Processing* (Chapman and Hall, 1995). He is an editor of the *Journal of Fourier Analysis and Applications*, *Applied and Computational Harmonic Analysis*, and *Constructive Approximation*.

Dr. Cohen received the V. Popov prize in 1995 for his contributions to approximation theory. He has been a co-organizer of the 1993 and 1996 Curves and Surfaces Conference of Approximation Theory, Chamonix, France.

Jelena Kovačević, for a photograph and biography, see p. 508 of this issue.