Renaud SIRDEY

A GENTLE INTRODUCTION TO THE WAVELET THEORY

A Gentle Introduction to the Wavelet Theory

by Renaud Sirdey

rsirdey@geocities.com

Introduction

The last decade has seen the development of new types of signal representations, generically known as the wavelet transforms (named after the well-known article of Alex Grossmann and Jean Morlet [GM84]). As in Fourier analysis, the wavelet transform consists in decomposing a given function onto a set of "building blocks". However, as opposed to the Fourier transformation (in which the "building blocks" are the well-known complex exponentials), the wavelet transform uses the dilated and translated version of a "mother wavelet" which has convenient properties according to time/frequency localization. As we will see later, this allows to perform a time/frequency analysis of signals which is much more relevant than those provided by other decompositions, e.g. the windowed Fourier transform. In the last few years, the wavelet analysis has been applied successfully to a wide range of problems from pure mathematics to engineering (characterization of some functionnal spaces, study of turbulence, signal processing, . . .).

This document intends to provide a <u>short</u> introduction to the wavelet theory. The subjects which are adressed are (in chronological order): the continuous wavelet transform, the dyadic wavelet transform, the notions of multiresolution analysis and orthogonal multiresolution analysis (in which orthogonal and non-redundant decompositions arise). Fast algorithms are presented for the dyadic and the orthogonal wavelet transforms. In terms of applications we quickly present the well-known wavelet-based denoising methods and the wavelet maxima representations.

Obviously, this document does not pretend to be exhaustive. It intends to provide the beginner with an idea of what the wavelet transforms are

and to give pointers to more specialised reading. This is the first version of the document, later versions will probably be more exhaustive (I am currently planing to write a chapter speaking about the connections between multifractal and wavelet analysis but time is running against me!).

This document has been written using principaly [JS93, Dau92] and the recent book by Stéphane Mallat: "A Wavelet Tour of Signal Processing" [Mal98], which is very complete (not only on a wavelet point of view). Some other "specialized" references have been used as well (all of them can be found in the bibliography).

1 Preliminaries

Definition 1 ($L^2(\mathbb{R})$**)** $L^2(\mathbb{R})$ *denotes the space of square integrable functions, i.e.*

$$L^2(\mathbb{R}) = \left\{ f / \int_{-\infty}^{+\infty} |f(x)|^2 dx < \infty \right\}$$

provided the scalar product

$$< f, g > = \int_{-\infty}^{+\infty} f(x) g^*(x) dx$$

and the associated norm

$$\|f\|^2 = < f, f >$$

This space (in association with this scalar product) has a Hilbert space structure (see definition 9, page 64).

Definition 2 (Fourier transform) *The Fourier transform[1] of a function* $f \in L^2(\mathbb{R})$ *is defined as*

$$\hat{f}(\xi) = \int_{-\infty}^{+\infty} f(x) e^{-i\xi x} dx$$

and its inverse is given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\xi) e^{i\xi x} d\xi$$

Many of the results presented in the next sections are dependant on this definition.

---

[1]In general [Weiar], a Fourier transform pair can be defined using two arbitrary constants A and B such that $\hat{f}(\xi) = A \int_{-\infty}^{+\infty} f(x) e^{Bi\xi x} dx$ and $f(x) = \frac{B}{2\pi A} \int_{-\infty}^{+\infty} \hat{f}(\xi) e^{-Bi\xi x} d\xi$.

Theorem 1 (Poisson sommation formula) *For all* $f, g \in L^2(\mathbb{R})$*, the Poisson sommation formula gives the two equalities*

$$\sum_{l=-\infty}^{+\infty} f(x-l) = \sum_{k=-\infty}^{+\infty} \hat{f}(2k\pi) e^{2i\pi kx} \tag{1}$$

and

$$\sum_{l=-\infty}^{+\infty} < f, \tau_l g > e^{-i\xi l} = \sum_{k=-\infty}^{+\infty} \hat{f}(\xi + 2k\pi) \hat{g}^*(\xi + 2k\pi) \tag{2}$$

A proof of this result can be found in [HD97].

2   The continuous wavelet transform

This section presents the continuous wavelet transform and dicusses its basic properties, e.g. time/frequency localization, inversion, redundancy, ... The connections between the continuous transform and other "discrete" wavelet transforms (dyadic or orthogonal wavelet transforms) will be emphasized in the next sections.

2.1   Definition and properties

Definition 3 (Continuous wavelet transform) *The continuous wavelet transform of* $f \in L^2(\mathbb{R})$ *is defined as [GM84, Mal98, Sta92]*

$$\mathcal{W}_f(a, b) = < f, \psi_{a;b} >, \ \psi_{a;b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \tag{3}$$

where $a > 0^2$ *and* b *are respectively the scale and the translation parameter.*

$\psi \in L^2(\mathbb{R})$ is called a wavelet function. This transformation is linear and invariant according to shift and scale. A direct consequence of the Parseval theorem[3] is

$$\mathcal{W}_f(a, b) = \frac{1}{2\pi} < \hat{f}, \hat{\psi}_{a;b} >, \ \hat{\psi}_{a;b}(\xi) = \sqrt{a} e^{-i\xi b} \hat{\psi}(a\xi) \tag{4}$$

---

[2]Other authors ([Dau92, JS93] for example) define the transform for all $a \neq 0$. It is therefore necessary to introduce an absolute value in equation (3).

[3]$< f, g > = A < \hat{f}, \hat{g} >$, the value of A depends on the definition of the Fourier transform. Here, $A = \frac{1}{2\pi}$.

Equations (3) and (4) imply that the wavelet coefficients contain some information about f coming from both the time and the frequency domains. The wavelet transform is therefore a time/frequency representation (the proper naming is time/scale representation) as the windowed Fourier transform introduced by Gabor [Gab46, FS97], or the Wigner-Ville distribution [Vil48, HBB92, Mal98].

Unfortunately, this type of representations is subject to a limitation due to the Weyl-Heisenberg undeterminacy relation[4] (this is not directly true for the Wigner-Ville distribution but its practical use involves an averaging which leads to a loss of time-frequency resolution [Mal98]). **In the rest of this document, the wavelet function is considered to be real** (complex wavelets are studied in details in [Mal98]).

2.2   The Weyl-Heisenberg undeterminacy relation

Theorem 2 (Weyl-Heisenberg undeterminacy relation) *Given a function $f \in L^2(\mathbb{R})$ such that $\|f\|^2 = 1$, the Weyl-Heisenberg relation indicates that*

$$\underbrace{\left(\int_{-\infty}^{+\infty} (x - \bar{x})^2 |f(x)|^2 dx \right)}_{\sigma_x^2} \underbrace{\left(\int_{-\infty}^{+\infty} (\xi - \bar{\xi})^2 |\hat{f}(\xi)|^2 d\xi \right)}_{\sigma_\xi^2} \geq A \qquad (5)$$

Where

- $\bar{x} = \int_{-\infty}^{+\infty} x |f(x)|^2 dx$;

- $\bar{\xi} = \int_{-\infty}^{+\infty} \xi |f(\xi)|^2 d\xi$.

Proofs of this theorem can (notably) be found in [BH96, Mal98] and in almost every books about quantum physics. The value of A also depends on the definition of the Fourier transform, here $A = \frac{1}{4}$. Optimizing equation (5) using techniques based on the calculus of variations (see [HD97] for an introduction) shows that Gauss functions of the form $K(a)e^{-ax^2}$ satisfy the optimum. A direct consequence of the theorem is that a function cannot be simultaneously well localized in both the time and the frequency domain, and it is obviously true for the wavelet function in equation (3).

By considering the time-frequency spread of $\psi_{a;b}$, it follows that most of the information contained in $\mathcal{W}_f(a, b)$ comes from the intervals $[b + a\bar{x} - a\sigma_x, b + a\bar{x} + a\sigma_x]$ (time domain) and $[(\bar{\xi} - \sigma_\xi)/a, (\bar{\xi} + \sigma_\xi)/a]$ (frequency domain) [JS93, Mal98]. These intervals define time-frequency windows,

---

[4]Also known as the Weyl-Heisenberg uncertainty principle. Here, we use the vocabulary introduced by E. Cornell [CW98].

known as Heisenberg boxes, whose areas depend on the translation and scale parameters. From Weyl-Heisenberg relation, the area of a given box has a lower bound: $4\sigma_x\sigma_\xi \geq 2$. However, an interesting property of the wavelet transform is that the dimensions of a given window can be adapted according to the "subject" of interest (as opposite to the windowed Fourier transform). Typically, it consists in using a "good" time resolution for studying the high frequencies and a "good" frequency resolution for the low frequencies.

2.3 Inversion of the continuous wavelet transform

Theorem 3 (Calderón identity) *If the wavelet* $\psi \in L^2(\mathbb{R})$ *satisfies the admissibility condition*

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\xi)|^2}{\xi} < \infty$$

then every function $f \in L^2(\mathbb{R})$ *is such that*

$$f(x) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}_f(a,b)\psi_{a;b}(x)\frac{db\,da}{a^2}$$

Proofs of this result can (notably) be found in [Mal98, Dau92]. This identity has been rediscovered in [GM84] and was known in harmonic analysis since 1964. The admissibily condition requires that $\hat{\psi}(0) = \int_{-\infty}^{+\infty} \psi(x)\,dx = 0$ which means that the wavelet must be an oscillating function[5]. Note that a given function can also be reconstructed from its wavelet transform using another wavelet [Sta92] $\chi(x)$ if

$$\int_0^{+\infty} \frac{\hat{\chi}(\xi)\hat{\psi}^*(\xi)}{\xi}d\xi < \infty$$

the inversion formula is then given by

$$f(x) = \int_0^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}(a,b)\chi_{a;b}(x)\frac{db\,da}{a^2}$$

2.4 Reproducing kernel

Now by inserting the Calderón identity in equation (3), we end up with

$$\mathcal{W}_f(a_0,b_0) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}_f(a,b)\kappa(a,a_0,b,b_0)\frac{db\,da}{a^2} \tag{6}$$

---

[5]Or the null function which is limited in interest.

where $\kappa(a, a_0, b, b_0) = < \psi_{a;b}, \psi_{a_0;b_0} >$ is called a reproducing kernel[6] [Mal98]. The modulus of the reproducing kernel measures the correlation between the two wavelets $\psi_{a;b}$ and $\psi_{a_0;b_0}$ and illustrates the redundancy of the continuous wavelet transform. Note that any function $\Phi(a, b)$ is the wavelet transform of some function $f \in L^2(\mathbb{R})$ if and only if it satisfies equation (6).

2.5   Scaling function

When the wavelet transform is known only for $a < a_0$, f cannot be recovered from its wavelet coefficients. Basically, the Calderón identity is broken into two parts

$$f(x) = \frac{1}{C_\psi} \int_0^{a_0} \int_{-\infty}^{+\infty} \mathcal{W}_f(a, b)\psi_{a;b}(x)\frac{db\,da}{a^2} + \frac{1}{C_\psi} \int_{a_0}^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}_f(a, b)\psi_{a;b}(x)\frac{db\,da}{a^2}$$

The role of the scaling function ϕ is to provide the information presents in the second term of the previous equation so that it becomes equal to

$$\int_{-\infty}^{+\infty} \mathcal{L}_f(a_0, b)\phi_{a_0;b}(x)db, \ \mathcal{L}_f(a, b) = < f, \phi_{a;b} >$$

By using the fact that $\mathcal{W}_f(a, b) = f \otimes \bar{\psi}_a(b)$, $\mathcal{L}_f(a, b) = f \otimes \bar{\phi}_a(b)$ and that

$$\int_{-\infty}^{+\infty} \mathcal{W}_f(a, b)\psi_{a;b}(x)db \ = \ \mathcal{W}_f(a, .) \otimes \psi_a(x),$$

$$\int_{-\infty}^{+\infty} \mathcal{L}_f(a, b)\phi_{a;b}(x)db \ = \ \mathcal{L}_f(a, .) \otimes \phi_a(x)$$

we end up with

$$f \otimes \bar{\phi}_{a_0} \otimes \phi_{a_0} = f \otimes \Psi(x), \ \Psi(x) = \int_{a_0}^{+\infty} \bar{\psi}_a \otimes \psi_a(x)\frac{da}{a^2}$$

This leads, via the convolution theorem, to the following constraint on $|\hat{\phi}(\xi)|^2$

$$|\hat{\phi}(\xi)|^2 = \int_1^{+\infty} |\hat{\psi}(a\xi)|^2\frac{da}{a^2}$$

The phase of $\hat{\phi}(\xi)$ can be arbitrarily chosen [Mal98].

2.6   Examples of wavelets

This subsection is obviously not exhaustive and gives only two examples of wavelet given in [Sta92].

---

[6]If the original function is reconstructed using another wavelet, the reproducing kernel becomes $\kappa(a, a_0, b, b_0) = < \chi_{a;b}, \psi_{a_0;b_0} >$.

Morlet's wavelet

The Morlet's wavelets is a complex wavelet whose real part is given by

$$\Re\{\psi\}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cos 2\pi\nu_0 x$$

and imaginary part by

$$\Im\{\psi\}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sin 2\pi\nu_0 x$$

ν_0 is a constant term. For this wavelet, the admissibility condition is not satisfied but if ν_0 is sufficiently large it becomes "pseudo-admissible" [Sta92]. See figure 1 (a) & (b) ($\nu = 0.4$).
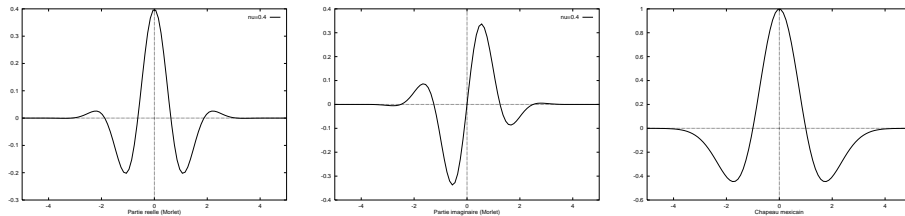
Mexican hat

The mexican hat is defined as the second derivative of a gaussian, its expression is therefore given by

$$\psi(x) = (1 - x^2) e^{-\frac{x^2}{2}}$$

The well-known property of the Fourier transform: $\frac{df^{(n)}}{dx^n}(x) \Leftrightarrow (i\xi)^n \hat{f}(\xi)$ implies directly that $\hat{\psi}(0) = \int_{-\infty}^{+\infty} \psi(x) dx = 0$ and the fact that the first-order moment of a Gauss function is finite prooves that this wavelet is admissible. See figure 1 (c).

Figure 1 Examples of wavelets.

(a) Morlet (real part).   (b) Morlet (im. part).   (c) Mexican hat.

3 Dyadic wavelet transform

A dyadic wavelet transform is obtained by discretizing the scale parameter a according to the dyadic sequence $\{2^j\}_{j\in\mathbb{Z}}$. In order to preserve the translation invariance property of the continuous wavelet transform the translation parameter is not discretized. Under particular conditions, the dyadic wavelet coefficients can be computed using a fast algorithm, known as an *algorithme à trous*.

3.1 Definition and inversion formula

Definition 4 (Dyadic wavelet transform) *The dyadic wavelet transform of* $f \in L^2(\mathbb{R})$ *is defined as*

$$\mathcal{W}_f(2^j, b) = <f, \psi_{2^j;b}>, \ \psi_{2^j;b}(x) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{x-b}{2^j}\right) \tag{7}$$

If the frequency plane is completly covered by dilated dyadic wavelets, then the dyadic wavelet transform defines a complete and stable[7] representation. The following theorem relates the dyadic wavelet transform to the frame theory [Mal98, Dau90, Dau92] and gives an inversion formula.

Theorem 4 *If there exists two constants* $A, B \in \mathbb{R}^{2+*}$ *such that*

$$\forall \xi \in \mathbb{R}, \ A \leq \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j\xi)|^2 \leq B$$

then

$$A\|f\|^2 \leq \sum_{j=-\infty}^{+\infty} \frac{1}{2^j}\|\mathcal{W}_f(2^j, b)\|^2 \leq B\|f\|^2$$

Moreover, if χ *satisfies*

$$\forall \xi \in \mathbb{R}^+, \ \sum_{j=-\infty}^{+\infty} \hat{\psi}^*(2^j\xi)\hat{\chi}(2^j\xi) = 1 \tag{8}$$

---

[7]The terms "complete" and "stable" should be understood in a frame theory context [DS52, Dau90, Mal98]. Roughly, a sequence $\{\theta_n\}_{n\in\tau}$ is said to be a frame of an Hilbert space $\mathcal{H}$ if there exist $A, B \in \mathbb{R}^{2+*}$ so that $\forall f \in \mathcal{H}$, $A\|f\|^2 \leq \sum_{n\in\tau}| < f, \theta_n >|^2 \leq B\|f\|^2$. This is a necessary and sufficient condition so that the operator $U_f[n] = <f, \theta_n>$ is invertible on its image with a bounded inverse. If $A = B$ the frame is said to be tight and if $A = B = 1$ the frame is an orthogonal basis of $\mathcal{H}$ [Dau90]. See (notably) [Mal98, Dau90, Dau92] for more details.

then

$$f(x) = \sum_{j=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}_f(2^j, b) \chi_{2^j;b}(x) \, db \tag{9}$$

A proof of this theorem can be found in [Mal98, Dau90]. χ denotes the reconstruction wavelet.

3.2 Reproducing kernel

As in the continuous case, the dyadic wavelet transform is a redundant representation whose redundancy is illustrated by a reproducing kernel equation. By inserting equation (7) in equation (9), we end up with

$$\mathcal{W}(2^{j_0}, b_0) = \sum_{j=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}_f(2^j, b) \kappa(j, j_0, b, b_0) \, db \tag{10}$$

where $\kappa(j, j_0, b, b_0) = < \chi_{2^j;b}, \psi_{2^{j_0};b_0} >$. Another equivalent way[8] of seeing this reproducing kernel [MZ92] consists in using the fact that $\mathcal{W}_f(2^j, b) = f \otimes \bar{\psi}^*_{2^j}(b)$ and that $f(x) = \sum_j \mathcal{W}_f(2^j, .) \otimes \chi_{2^j}(x)$. Inserting the last expression of $f(x)$ in the one of $\mathcal{W}_f(2^{j_0}, b_0)$ gives

$$\mathcal{W}_f(2^{j_0}, b_0) = \sum_{j=-\infty}^{+\infty} \mathcal{W}_f(2^j, .) \otimes \kappa'_{2^j, 2^{j_0}}(b_0), \quad \kappa'_{2^j, 2^{j_0}}(b) = \chi_{2^j} \otimes \bar{\psi}^*_{2^{j_0}}(b) \tag{11}$$

3.3 Dyadic wavelets and *algorithme à trous*

If the wavelets and scaling functions are properly designed, the dyadic wavelet transform can be computed via a fast algorithm based on filter banks [Mal98, MZ92, RD92, She92], known as an *algorithme à trous*. It requires that there exists two discrete filters h and g with $\sum_k h_k = \sqrt{2}$ so that the scaling function ϕ and the wavelet ψ respectively satisfy

$$\hat{\phi}(\xi) = \hat{h}(\xi/2)\hat{\phi}(\xi/2) \tag{12}$$

and

$$\hat{\psi}(\xi) = \hat{g}(\xi/2)\hat{\phi}(\xi/2) \tag{13}$$

where $\hat{h}(\xi) = \frac{1}{\sqrt{2}} \sum_k h_k e^{-i\xi k}$ is the Fourier transform of the distribution $\frac{1}{\sqrt{2}} \sum_k h_k \delta_k$ (same for $\hat{g}(\xi)$). If $\mathcal{L}_f(2^j, b) = < f, \phi_{2^j;b} >$ is known, we can calculate

$$\mathcal{W}_f(2^{j+1}, b) = < f, \psi_{2^{j+1};b} > \text{ and } \mathcal{L}_f(2^{j+1}, b) = < f, \phi_{2^{j+1};b} >$$

---

[8]This holds for the continuous wavelet transform as well.

by using only the discrete filters h and g. Since $\mathcal{L}_f(2^{j+1}, b) = f \otimes \bar{\phi}_{2^{j+1}}(b)$, we have[9] (from equation (12))

$$
\begin{aligned}
\mathcal{L}_f(2^{j+1}, b) &\Leftrightarrow \hat{f}(\xi)\hat{\phi}^*_{2^{j+1}}(\xi) \\
&= \hat{h}^*_{2^j}\hat{f}(\xi)\hat{\phi}^*_{2^j}(\xi) \\
&\Leftrightarrow \bar{h}_{2^j} \otimes \mathcal{L}_f(2^j, .)(b)
\end{aligned}
\tag{14}
$$

The same kind of argument gives

$$
\mathcal{W}_f(2^{j+1}, b) = \bar{g}_{2^j} \otimes \mathcal{L}_f(2^j, .)(b)
\tag{15}
$$

h_{2^j} (resp. g_{2^j}) is obtained from h (resp. g) by inserting $2^j - 1$ zeros between the samples of h (resp. g). The pair φ, χ (respectively the scaling function and wavelet) used for reconstructing the signal should as well satisfy two similar equations as (12) and (13) with filters $\tilde{h}$ and $\tilde{g}$ instead of h and g. Obviously, we must be able to recover $\mathcal{L}_f(2^j, b)$ from $\mathcal{L}_f(2^{j+1}, b)$ and $\mathcal{W}_f(2^{j+1}, b)$. This is done via the following formula

$$
\mathcal{L}_f(2^j, b) = \tilde{g}_{2^j} \otimes \mathcal{W}_f(2^{j+1}, .)(b) + \tilde{h}_{2^j} \otimes \mathcal{L}_f(2^{j+1}, .)(b)
\tag{16}
$$

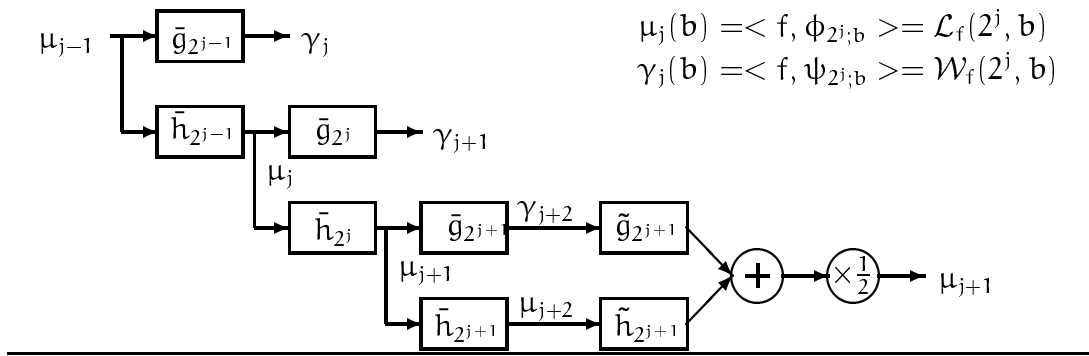Algorithm 2 illustrates the working of the algorithm. Equation (15) is equivalent to

$$
\mathcal{L}_f(2^j, b) = (\bar{h}_{2^j} \otimes \tilde{h}_{2^j} + \bar{g}_{2^j} \otimes \tilde{g}_{2^j}) \otimes \mathcal{L}_f(2^j, .)(b)
$$

Therefore, the required perfect reconstruction introduces the constraint

$$
\hat{h}^*(\xi)\hat{\tilde{h}}(\xi) + \hat{g}^*(\xi)\hat{\tilde{g}}(\xi) = 1, \ \forall \xi \in [-\pi, \pi]
$$

which is equivalent to condition (8) (proof in [Mal98]).

Figure 2 *Algorithme à trous.*

---

[9]The convolution operators in equations (14), (15) and (16) should be understood in a distribution theory context. See [HD97].

12

3.4 Practical considerations

In practical cases, i.e. discrete signals of finite duration, the convolutions in equations (14), (15) and (16) are replaced by circular convolutions. Since the scalar product of the discrete sequence with $\phi_{2^{\log_2 N};k}$ is constant [Mal98] (N is the length of the signal), the scale only goes from $2^0 = 1$ to $2^{\log_2 N}$. Most of the time, the samples of the discrete input sequence are considered as the average of a function f weighted by $\phi(x - k)$ and give the first approximation required for starting the algorithm. The complexity of the algorithm is in $\mathcal{O}(N \log_2 N)$. Figure 3 shows the dyadic wavelet transform of a signal computed by means of the quadratic spline wavelet given in [MZ92].

4 Multiresolution analysis of $L^2(\mathbb{R})$

The multiresolution analysis introduced by Stéphane Mallat in 1989 [Mal89] provides a theoretical context in which non redundant and orthogonal wavelet decompositions arise. However, the definition of a multiresolution analysis does not require any constraint of orthogonality. In this section we focus on the basic properties of a multiresolution analysis and their implications on the pair wavelet/scaling function.

4.1 Definition

Definition 5 (Multiresolution analysis) *A multiresolution analysis is a set of closed subspaces V_j of $L^2(\mathbb{R})$, which satisfies the following six properties[10] [Mal89, JS93, BH96]*

1. *$V_j \subset V_{j+1}$, $\forall j$.*

2. *$v(x) \in V_0 \Leftrightarrow v(x - k) \in V_0$, $\forall k \in \mathbb{Z}$.*

3. *$v(x) \in V_j \Leftrightarrow v(2x) \in V_{j+1}$, $\forall j \in \mathbb{Z}$.*

4. *$\lim_{j \to -\infty} V_j = \bigcap_{j=-\infty}^{+\infty} V_j = \{0\}$.*

5. *$\lim_{j \to +\infty} V_j = \bigcup_{j=-\infty}^{+\infty} V_j = L^2(\mathbb{R})$.*

6. *There exists a scaling function $\phi \in L^2(\mathbb{R})$ such that $\{\tau_k \phi\}_{k \in \mathbb{Z}}$ is a Riesz basis of V_0.*

---

[10]Some authors [Mal98, Dau92] use $V_{j+1} \subset V_j$ and $f(x) \in V_j \Leftrightarrow f(x/2) \in V_{j+1}$ for (respectively) properties 1 and 3.

Figure 3 Beginning of a dyadic wavelet transform.
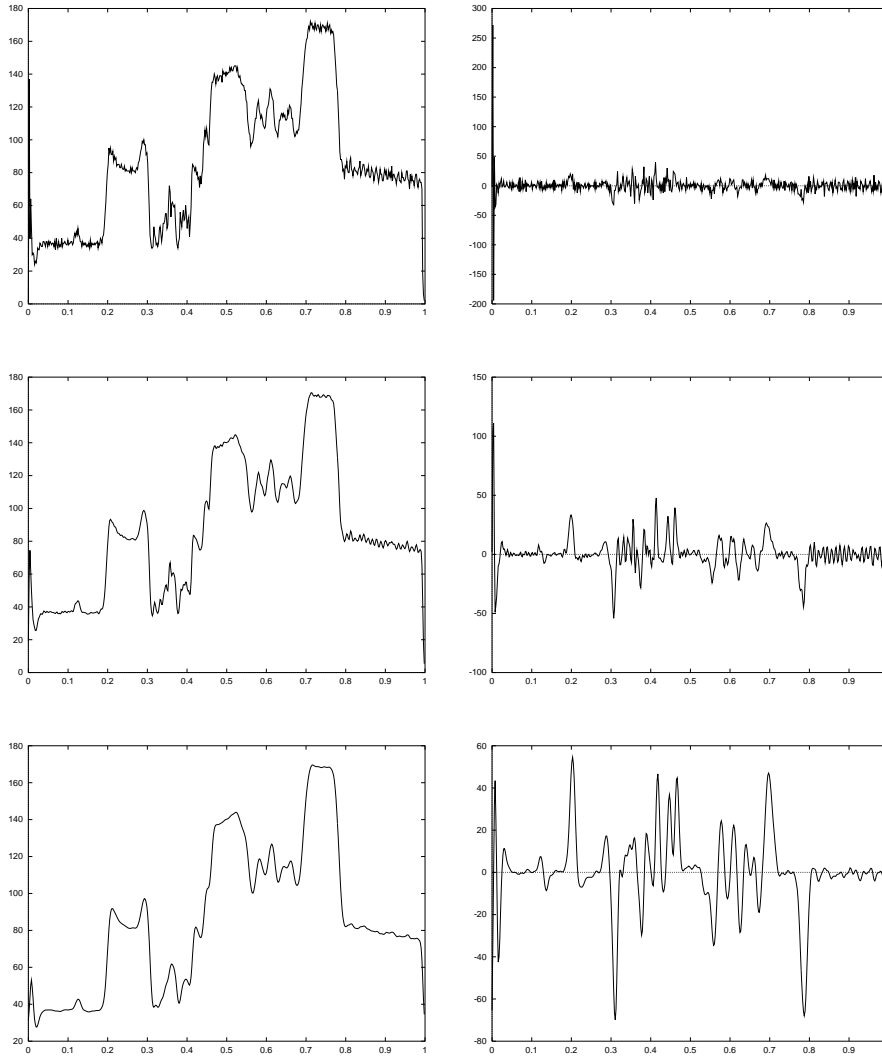
(e) Scaling coefficients.　　　　(f) Wavelet coefficients.

Property 1 (causality property) means that an approximation in V_j contains all the information for computing an approximation at a coarser resolution. Property 2 indicates that V_0 is invariant under integer translations. Property 3 says that the null fonction is the only common object to all the subspaces V_j, i.e. we lose all the details about f as j goes to $-\infty$. Property 4 means that every functions of $L^2(\mathbb{R})$ can be approximated to an arbitrary precision. The definition of a Riesz basis (property 6) is available page 64. Properties 3 and 6 directly imply that the family $\{\phi_{j;k}\}_{k\in\mathbb{Z}}$ (now $\phi_{j;k}(x)$ stands for $\sqrt{2^j}\phi(2^j x - k)$) forms a Riesz basis of V_j.

Simple examples of multiresolution analysis are: the piecewise constant approximations (related to the Haar wavelet), the shannon approximations (related to the Shannon wavelet) and the spline approximations, the two first examples defines some orthogonal multiresolutions while the third defines a non-orthogonal one, more details are avalaible in [Mal98, JS93].

4.2 Dilation equation and basic consequences

Theorem 5 (Dilation equation) *Let $\phi \in L^2(\mathbb{R})$ be the scaling function of a multiresolution analysis, then [JS93, Dau92]*

$$\exists\{h_k\}_{k\in\mathbb{Z}}/\phi(x) = \sqrt{2}\sum_{k=-\infty}^{+\infty} h_k\phi(2x-k) \qquad (17)$$

This theorem follows directly from properties 1 and 6: as a function of V_0 (property 6) and because $V_0 \subset V_1$ (property 1), ϕ can be expressed as a linear combination of the basis function of V_1.

Equation (17) is known as a **dilation equation** [Str94] or a **scaling equation** [Mal98] and plays a fundamental role in the orthogonal dyadic wavelet theory. Integrating equation (17) on both sides implies that $\sum_k h_k = \sqrt{2}$.

Introducing the dilation equation in the Fourier transform of ϕ leads to

$$\hat{\phi}(\xi) = \hat{h}(\xi/2)\hat{\phi}(\xi/2) \qquad (18)$$

where $\hat{h}(\xi) = \frac{1}{\sqrt{2}}\sum_k h_k e^{-i\xi k}$ denotes the Fourier transform (2π-periodic) of the distribution $\frac{1}{\sqrt{2}}\sum_k h_k \delta_k$. Equation (18) can be used recursively and gives (at least formally)

$$\hat{\phi}(\xi) = \prod_{k=1}^{\infty} \hat{h}(\xi/2^k)$$

15

This product can be interpreted as an infinite cascade of convolutions of the distribution $\frac{1}{\sqrt{2}}\sum_k h_k \delta_k$ by itself, its convergence properties are (notably) studied in [Dau88].

4.3 Complementary subspaces

Let W_j denotes the subspace complementing V_j in W_{j+1} i.e.

$$V_{j+1} = V_j \oplus W_j$$

where $\oplus$ denotes the direct sum operator. As a consequence of property 5, we have $\bigoplus_{j=-\infty}^{+\infty} W_j = L^2(\mathbb{R})$.

Definition 6 (Wavelet function) *In a multiresolution context, a function ψ is said to be a wavelet function if the family $\{\tau_k \psi\}_{k \in \mathbb{Z}}$ is a Riesz basis of the complementary subspace W_0 [JS93].*

As a function of V_1, the wavelets also obey a dilation equation

$$\psi(x) = \sqrt{2} \sum_{k=-\infty}^{+\infty} g_k \phi(2x - k)$$

which leads to

$$\hat{\psi}(\xi) = \hat{g}(\xi/2)\hat{\phi}(\xi/2) \tag{19}$$

and (as in the case of the scaling function) to an infinite product of the form

$$\hat{\psi}(\xi) = \hat{g}(\xi/2) \prod_{k=2}^{+\infty} \hat{h}(\xi/2^k)$$

The family of functions $\{\psi_{j;k}\}_{j,k \in \mathbb{Z}^2}$ forms a Riesz basis of $L^2(\mathbb{R})$ [JS93, Mal98]. As a consequence, every functions of $L^2(\mathbb{R})$ can be written as

$$f(x) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \mu_{j;k} \psi_{j;k}(x) \tag{20}$$

This equation can be seen as an inverse wavelet transform where the scale and the translation parameters have been discretized.

5 Orthogonal multiresolution analysis

5.1 Definition and perfect reconstruction constraint

Definition 7 (Orthogonal multiresolution analysis) *An orthogonal multiresolution analysis is a multiresolution analysis such that for all $j \in \mathbb{Z}$, W_j is the orthogonal complement of V_j in V_{j+1} [JS93, Dau92].*

A sufficient condition for a multiresolution to be orthogonal is given by [JS93]

$$V_0 \perp W_0 \text{ i.e. } < \phi, \tau_k \psi >= 0, \ \forall k \in \mathbb{Z} \tag{21}$$

A consequence of this definition is the existence of an unique scaling function ϕ so that the family $\{\tau_k \phi\}_{k \in \mathbb{Z}}$ forms an orthogonal basis of V_0 [Mal89], i.e.

$$< \phi, \tau_k \phi >= \delta_{0,k}, \ \forall k \in \mathbb{Z} \tag{22}$$

Now, the families $\{\phi_{j;k}\}_{k \in \mathbb{Z}}$, $\{\psi_{j;k}\}_{k \in \mathbb{Z}}$ and $\{\psi_{j;k}\}_{j,k \in \mathbb{Z}^2}$ form orthogonal basis of (respectively) V_j, W_j and $L^2(\mathbb{R})$ (proof in [Mal98]). Hence, in this context, equation (20) can be written as

$$f(x) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} < f, \psi_{j;k} > \psi_{j;k}(x)$$

By using the Poisson formula (equation (2)), equation (22) is equivalent to

$$F(\xi) = \sum_{k=-\infty}^{+\infty} |\hat{\phi}(\xi + 2k\pi)|^2 = 1 \tag{23}$$

Since (from equation (18) and from the fact that both $F(\xi)$ and $\hat{h}(\xi)$ are 2π-periodic)

$$
\begin{aligned}
F(2\xi) &= \sum_{k=-\infty}^{+\infty} |\hat{\phi}(2\xi + 2k\pi)|^2 \\
&= \sum_{k=-\infty}^{+\infty} |\hat{h}(\xi + k\pi)|^2 |\hat{\phi}(\xi + k\pi)|^2 \\
&= \sum_{k=-\infty}^{+\infty} |\hat{h}(\xi + 2k\pi)|^2 |\hat{\phi}(\xi + 2k\pi)|^2 \\
&\quad + \sum_{k=-\infty}^{+\infty} |\hat{h}(\xi + \pi + k\pi)|^2 |\hat{\phi}(\xi + \pi + k\pi)|^2 \\
&= |\hat{h}(\xi)|^2 F(\xi) + |\hat{h}(\xi + \pi)|^2 F(\xi + \pi)
\end{aligned}
$$

we end up with the following theorem [JS93, Mal98, Dau88].

Theorem 6 (Perfect reconstruction) *Let $\phi \in L^2(\mathbb{R})$ be the scaling function of an orthogonal multiresolution, then $\hat{h}(\xi)$ satisfies*[11]

$$|\hat{h}(\xi)|^2 + |\hat{h}(\xi + \pi)|^2 = 1 \tag{24}$$

---

[11]Note that the right hand side depends on the definition of $\hat{h}(\xi)$. Here: $\hat{h}(\xi) = \frac{1}{\sqrt{2}} \sum_k h_k e^{-i\xi k}$. Other authors [Dau88, Mal98] end up with the right hand side equals to 2.

This constraint is fundamental for the design of orthogonal wavelets and connects wavelets to the conjuguate quadrature filters (tree-structured subband coders with exact reconstruction [SB86]) theory[12] [Dau88, Mal98, CDF92]. Note that equation (24) is a sufficient condition so that $\phi \in L^2(\mathbb{R})$ [CDF92].

5.2 Relation between $\hat{h}(\xi)$ and $\hat{g}(\xi)$

Now consider the sufficient condition for a multiresolution to be orthogonal (equation (21)), again from Poisson formula it is equivalent to

$$G(\xi) = \sum_{k=-\infty}^{+\infty} \hat{\phi}(\xi + 2k\pi)\hat{\psi}^*(\xi + 2k\pi) = 0$$

which leads to (from equations (18), (19) and (23))

$$
\begin{aligned}
G(2\xi) &= \sum_{k=-\infty}^{+\infty} \hat{\phi}(2\xi + 2k\pi)\hat{\psi}^*(2\xi + 2k\pi) \\
&= \sum_{k=-\infty}^{+\infty} \hat{h}(\xi + k\pi)\hat{g}^*(\xi + k\pi)|\hat{\phi}(\xi + k\pi)|^2 \\
&= \sum_{k=-\infty}^{+\infty} \hat{h}(\xi + 2k\pi)\hat{g}^*(\xi + 2k\pi)|\hat{\phi}(\xi + 2k\pi)|^2 \\
&+ \sum_{k=-\infty}^{+\infty} \hat{h}(\xi + \pi + 2k\pi)\hat{g}^*(\xi + \pi + 2k\pi)|\hat{\phi}(\xi + \pi + 2k\pi)|^2 \\
&= \hat{h}(\xi)\hat{g}^*(\xi) + \hat{h}(\xi + \pi)\hat{g}^*(\xi + \pi) = 0
\end{aligned}
$$

This implies the following relation [JS93]

$$\hat{g}(\xi) = \alpha(\xi)\hat{h}^*(\xi + \pi)$$

where $\alpha(\xi)$ is a 2π-periodic function such that $\alpha(\xi) = -\alpha(\xi + \pi)$. Now, from Parseval theorem, equation (22) can be written as

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{h}(\xi/2)|^2|\hat{\phi}(\xi/2)|^2 e^{-i\xi k}dx = \delta_{0,k}, \ \forall k \in \mathbb{Z} \qquad (25)$$

Since

$$<\psi, \tau_k\psi> = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{\psi}(\xi)|^2 e^{-i\xi k}d\xi$$

---

[12]For every orthogonal bases of compactly supported wavelets, there exists a pair of discrete filters which defines a subband coder allowing perfect reconstruction [CDF92] (the opposite is not generally true).

$$= \int_{-\infty}^{+\infty} |\hat{g}(\xi/2)|^2 |\hat{\phi}(\xi/2)|^2 e^{-i\xi k} d\xi,$$

$$= \int_{-\infty}^{+\infty} |\alpha(\xi/2)|^2 |\hat{h}(\xi/2 + \pi)|^2 |\hat{\phi}(\xi/2)|^2 e^{-i\xi k} d\xi,$$

$$= \int_{-\infty}^{+\infty} |\alpha(\xi/2)|^2 |\hat{h}(\xi/2)|^2 |\hat{\phi}(\xi/2)|^2 e^{-i\xi k} d\xi,$$

the orthogonality of the wavelet is implied by the orthogonality of the scaling function if $|\alpha(\xi)|^2 = 1$. We then impose the following constraint: if the scaling function has a compact support, i.e. $\hat{h}(\xi)$ is a trigonometric polynomial, the wavelet must have a compact support. This constraint requires that $\alpha(\xi)$ is a trigonometric polynomial as well. The only trigonometric polynomials which have these two properties are of the form

$$\alpha(\xi) = Ke^{-i(2k+1)\xi}$$

with $|K| = 1$. Choosing $K = \pm 1$ implies that if the coefficients $\{h_k\}_{k \in \mathbb{Z}}$ are real, then the coefficients $\{g_k\}_{k \in \mathbb{Z}}$ are also real. The "classical" choice [JS93, Dau88] is $\alpha(\xi) = -e^{-i\xi}$ and leads to

$$\hat{g}(\xi) = -e^{-i\xi}\hat{h}^*(\xi + \pi) \tag{26}$$

$$= -e^{-i\xi} \sum_{-\infty}^{+\infty} h_k^* e^{i(\xi + \pi)k}$$

$$= -\sum_{k=-\infty}^{+\infty} (-1)^k h_k^* e^{-i\xi(1-k)}$$

$$= \sum_{l=-\infty}^{\infty} (-1)^l h_{1-l}^* e^{-i\xi l}$$

hence

$$g_k = (-1)^k h_{1-k}^* \tag{27}$$

Other authors (notably) [Mal98] choose $\alpha(\xi) = e^{-i\xi}$ and then end up with $g_k = (-1)^{1-k} h_{1-k}^*$ instead of equation (27).

5.3 Extension: biorthogonal multiresolution analysis

The notion of biorthogonal multiresolution analysis [CDF92, JS93] generalizes the idea of multiresolution analysis by using different scaling function/wavelet pairs for respectively the decomposition and the reconstruction of the signal. The idea consists in defining two ladders of

closed subspaces[13]

$$\ldots \subset V_{-j} \subset \ldots \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots \subset V_j \subset \ldots$$

and

$$\ldots \subset \tilde{V}_{-j} \subset \ldots \subset \tilde{V}_{-1} \subset \tilde{V}_0 \subset \tilde{V}_1 \subset \ldots \subset \tilde{V}_j \subset \ldots$$

such that they respectively lead to a multiresolution analysis and a dual multiresolution analysis. Moreover, it is required that

$$\tilde{W}_j \perp V_j \text{ and } W_j \perp \tilde{V}_j$$

so that $\{\psi_{j;k}\}_{j,k \in \mathbb{Z}^2}$ and $\{\tilde{\psi}_{j;k}\}_{j,k \in \mathbb{Z}^2}$ define two dual Riesz basis of $L^2(\mathbb{R})$. By mimicing the orthogonal case, it is possible to derive conditions on the four filters h, g, $\tilde{h}$, $\tilde{g}$ such that they lead to a perfect reconstruction subband coding scheme (with different analysis and synthesis filters). For more precisions the reader is sent to (notably) [CDF92, Mal98]. Finally, every function $f \in L^2(\mathbb{R})$ can be expressed as [CDF92]

$$f(x) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} <f, \psi_{j;k}> \tilde{\psi}_{j;k}(x) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} <f, \tilde{\psi}_{j;k}> \psi_{j;k}(x)$$

which illustrates the fact that the role of the two basis can be interchanged. The interest of building biorthogonal multiresolution analysis comes from the fact that more freedom is allowed in the design of the wavelets/filters and that it becomes possible to create symetric wavelets.

6 Orthogonal wavelets and fast algorithm

6.1 Fast orthogonal wavelet transform

Now let $\mu_{j;k} = <f, \phi_{j;k}> = \mathcal{L}_f(2^{-j}, 2^{-j}k)$ and $\gamma_{j;k} = <f, \psi_{j;k}> = \mathcal{W}_f(2^{-j}, 2^{-j}k)$. Generalizing the scaling equation (17) gives

$$\begin{aligned}
\phi_{j;k}(x) &= \sum_{l=-\infty}^{+\infty} h_l \phi_{j+1;2k+l}(x) \\
&= \sum_{m=-\infty}^{+\infty} h_{m-2k} \phi_{j+1;m}
\end{aligned}$$

---

[13]As in the simple multiresolution case some authors ([CDF92] notably) are using $\ldots \supset V_j \supset V_{j+1} \supset \ldots$ and $\ldots \supset \tilde{V}_j \supset \tilde{V}_{j+1} \supset \ldots$ instead of the previous definition.

Putting this last equation in the expression of $\mu_{j;k}$ leads to

$$\mu_{j;k} = \sum_{l=-\infty}^{+\infty} h_{l-2k} < f, \phi_{j+1;l} >$$

$$= \sum_{l=-\infty}^{+\infty} \bar{h}_{2k-l}\mu_{j+1;l} = \bar{h} \otimes \mu_{j+1}[2k] \tag{28}$$

The same kind of arguments gives

$$\gamma_{j;k} = \bar{g} \otimes \mu_{j+1}[2k] \tag{29}$$

Equations (28) and (29) are discrete convolutions followed by a down-sampling operation. We come back to the $\mu_{j+1;k}$ from the $\mu_{j;k}$ and the $\gamma_{j;k}$ by inserting

$$f_{j+1}(x) = \underbrace{\sum_{l=-\infty}^{+\infty} \mu_{j;l}\phi_{j;l}(x)}_{\in V_j} + \underbrace{\sum_{l=-\infty}^{+\infty} \gamma_{j;l}\psi_{j;l}(x)}_{\in W_j}$$

in $\mu_{j+1;k} =< f, \phi_{j+1;k} >=< f_{j+1}, \phi_{j+1;k} >$. Hence (and from the orthogonality of the scaling function),

$$\mu_{j+1;k} = \sum_{l=-\infty}^{+\infty} \mu_{j;l} \sum_{m=-\infty}^{+\infty} h_{m-2l} < \phi_{j+1;m}, \phi_{j+1;k} >$$

$$+ \sum_{l=-\infty}^{+\infty} \gamma_{j;l} \sum_{m=-\infty}^{+\infty} g_{m-2l} < \phi_{j+1;m}, \phi_{j+1;k} > \tag{30}$$

$$= \sum_{l=-\infty}^{+\infty} \mu_{j;l} \sum_{m=-\infty}^{+\infty} h_{m-2l}\delta_{k,m} + \sum_{l=-\infty}^{+\infty} \gamma_{j;l} \sum_{m=-\infty}^{+\infty} g_{m-2l}\delta_{k,m}$$

$$= \sum_{l=-\infty}^{+\infty} \mu_{j;l}h_{k-2l} + \sum_{l=-\infty}^{+\infty} \gamma_{j;l}g_{k-2l} \tag{31}$$

The first and second terms of this last equation are discrete convolutions preceded by an upsampling operation (insert one zero between every sample). See algorithm 4.

Equations (28), (29) and (31) define a perfect reconstruction decimated filter banks (perfect reconstruction subband coder). If we note $\hat{\mu}_j(\xi) = \sum_k \mu_{j;k}e^{-i\xi k}$ the discrete Fourier transform of the sequence $\{\mu_{j;k}\}_{k\in\mathbb{Z}}$ (same for $\hat{\gamma}_j(\xi)$) we have: $\hat{\mu}_j(\xi) = \hat{h}^*(\xi/2)\hat{\mu}_{j+1}(\xi/2)$, $\hat{\gamma}_j(\xi) = \hat{g}^*(\xi/2)\hat{\mu}_{j+1}(\xi/2)$ and $\hat{\mu}_{j+1}(\xi) = \hat{h}(\xi)\hat{\mu}_j(2\xi) + \hat{g}(\xi)\hat{\gamma}_j(2\xi)$. Therefore,

$$\hat{\mu}_{j+1}(\xi) = (\hat{h}(\xi)\hat{h}^*(\xi) + \hat{g}(\xi)\hat{g}^*(\xi))\hat{\mu}_{j+1}(\xi)$$

21

Figure 4 Fast decimated filter bank algorithm.

$$\mu_{j;k} = < f, \phi_{j;k} > = \mathcal{L}_f(2^{-j}, 2^{-j}k)$$
$$\gamma_{j;k} = < f, \psi_{j;k} > = \mathcal{W}_f(2^{-j}, 2^{-j}k)$$
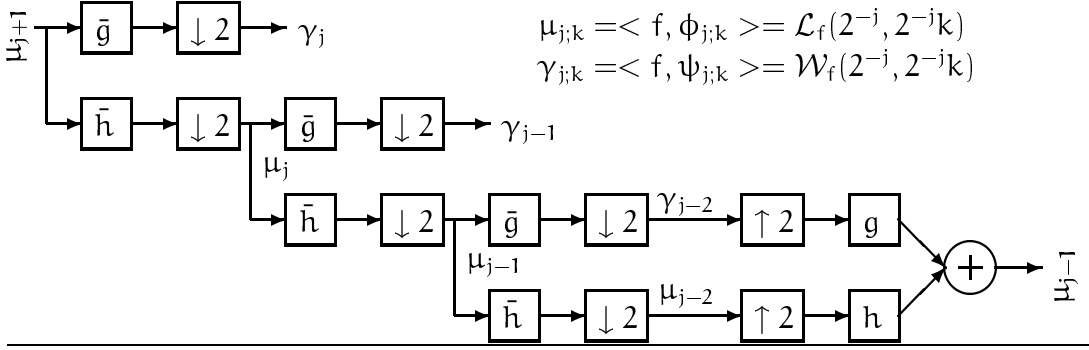
The perfect reconstruction constraint is then given by $\hat{h}(\xi)\hat{h}^*(\xi) + \hat{g}(\xi)\hat{g}^*(\xi) = 1$ or in "Z-transform" notation

$$H(e^{i\xi})H(e^{-i\xi}) + G(e^{i\xi})G(e^{-i\xi}) = 1$$

this condition is equivalent to equation (24), using the expression of $\hat{g}(\xi)$ derived in §5.2, and (roughly) to the condition derived in [SB86].

6.2   Practical considerations

There exists different ways of modifying the fast wavelet transform algorithm for dealing with practical signals, i.e. sampled signals of finite duration. Most of the time, (as in the *algorithme à trous* case), the samples are interpreted such that they gives $\mu_{0;k}$. Note that the complexity of the algorithm is in $\mathcal{O}(\log_2 N)$ (faster than the fast Fourier transform—$\mathcal{O}(N \log_2 N)$).

6.2.1   Zero padding

The most intuitive way of dealing with the border problem consists in assuming that the function vanishes outside the sampling interval i.e. $\in L^2([0, N])$. The fast wavelet transform algorithm is therefore applied without modification. However, the signal is interpreted as if it was discontinuous at $x = 0$ and $x = N$: large coefficients are created in the neighbourhood of this points and significant errors may appear during the reconstruction process.

6.2.2   Periodic wavelets

A better solution consists in using a proper orthogonal basis of $L^2([0, N])$. For example, an orthogonal basis of $L^2([0, N])$ can be contructed by periodizing an orthogonal wavelet basis of $L^2(\mathbb{R})$. By using the periodic

22

extension of a function $f \in L^2([0, N])$ **i.e.**

$$f^{(\pi)}(x) = \sum_{k=-\infty}^{\infty} f(x + kN)$$

it can be shown[14] **[Mal98] that the family**

$$\{\psi_{j;k}^{(\pi)}\}_{j,k \in \mathbb{Z}^2}, \ \psi_{j;k}^{(\pi)}(x) = \sqrt{2^j} \sum_{l=-\infty}^{+\infty} \psi(2^j x - k + 2^j lN)$$

forms an orthogonal basis of $L^2([0, N])$ such that

$$< \psi_{j;k}^{(\pi)}, \psi_{j';k'}^{(\pi)} >_{L^2([0,N])} = \delta_{j,j'} \delta_{k,k'}$$

and

$$f(x) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} < f, \psi_{j;k}^{(\pi)} >_{L^2([0,N])} \psi_{j;k}^{(\pi)}(x)$$

This is similar to consider a wavelet decomposition on a torus instead on the the real line [JS93] and the main modification is to replace the convolution operators in equations (28), (29) and (31) by circular convolutions. However, the problem of creating large coefficients in the neighbourhood of 0 and N is not avoided since there is no garantee that $f(0) = f(N)$.

6.2.3 Boundary wavelets

Using boundary wavelets avoids to create large wavelet coefficients at the border. Basically, it consists in using modified wavelets functions, which have as many vanishing moments as the original, for processing the borders. For a proper presentation, the reader is sent to [CDJV92, CDV93, Mal98].

Note that folded wavelets are usable if the corresponding basis of $L^2(\mathbb{R})$ is constructed using symetric or antisymetric wavelets. This cannot occur in the one-dimensional orthogonal case, but can happen for biorthogonal wavelets. This solution preserves the continuity at the border but acts as if the signal had a discontinuous first-order derivative in the neighbourhood of 0 and N.

---

[14]The proof is based on using the fact that $f(x) = f^{(\pi)}(x), \ \forall x \in [0, N], f \in L^2([0, N])$ and on periodizing the decomposition $\sum_j \sum_k < f, \psi_{j;k} > \psi_{j;k}(x)$.

6.3 Examples of orthogonal wavelets

This subsection gives some important properties that a wavelet function may have (this list has been taken from [JS93]) and presents a few families of orthogonal wavelet. Pointers to articles in which biorthogonal wavelets are constructed are also given.

6.3.1 Properties of a wavelet function

Orthogonality. The orthogonality is convenient to have in many situations. First, it directly links the L^2-norm of a function to a norm on its wavelet coefficients. Second, the fast wavelet transform is a unitary transformation ($W^{-1} = W^\dagger$) which means that the condition number of the transformation ($\|W\|\|W^{-1}\|$) is equal to 1 (optimal case), i.e. stable numerical computations are possible. Moreover, if the multiresolution is orthogonal the projection operators onto the different subspaces (V_j, W_j) yield optimal approximations in the L^2 sense.

Compact support. If the wavelet has a compact support the filter h and g have a finite impulse response. Obviously, this is convenient for implementing the fast wavelet transform. However, if the wavelet does not have a compact support, a fast decay is required so that h and g can be reasonably approximated using FIR filters.

Rational coefficients. For efficient computations, it can be interesting that the coefficients of h and g are rational or dyadic rational. Binary shifts are much faster than floating point operations.

Symmetry. If the scaling function and wavelet are symmetric, the filters h and g have generalized linear phase. The absence of this property can lead to phase distortion.

Regularity. As pointed in the works of Yves Meyer [Mey90] and David Donoho [Don91] the regularity of the multiresolution analysis is crucial for many applications such as data compression, statistical estimation, . . . In the biorthogonal case, the regularity of the primary multiresolution is more important than the regularity of the dual one [JS93].

Number of vanishing moments. The number of vanishing moments is connected to the regularity of the wavelet and vice versa [Mey90].

Analytic expressions.   In some case, it can be useful to have analytic expressions for the scaling function and wavelet.

Interpolation.   If the scaling function satisfies

$$\phi(k) = \delta_k, \ k \in \mathbb{Z}$$

then the computation of the first scaling coefficients (required for starting the fast wavelet transform) is trivial and the assumption discussed in §6.2 is valid.

Obviously, a given multiresolution cannot satisfies all these properties (e.g. orthogonality, compact support and symmetry are exclusive properties in one dimension except for the Haar wavelet) and it is necessary to make a trade-off between them.

6.3.2   Some families of orthogonal wavelets

The Haar transform.   The Haar transform has been invented in 1910, long before the invention of the terms "wavelet" and "multiresolution". Some books about image processing present it as a curiosity [GW92]. The Haar transform corresponds to an orthogonal multiresolution, associated with the following scaling function and wavelet

$$\phi(x) = \chi_{[0,1]}(x), \ \psi(x) = \chi_{[0,1/2]}(x) - \chi_{[1/2,1]}(x)$$

The discrete filter h is then equal to $\{1, 1\}$. However, the Haar transform is not very used in practice because the analysing functions are too discontinuous. Note that the Haar wavelet is a particular case of a Daubechies wavelet for $N = 1$.

The Shannon wavelet.   The Shannon wavelet is constructed from the Shannon multiresolution approximations which approximates functions by their restrictions to low frequency intervals. The scaling function is then a cardinal sine and the wavelet is equal to [JS93]

$$\psi(x) = \frac{\sin 2\pi x - \sin \pi x}{\pi x}$$

This wavelet is C^∞ but it has a very slow time decay which makes it not suitable for practical purpose.

Meyer and Battle-Lemarié.   A more interesting example is given by the Meyer wavelet and scaling function [Mey90] which are C^∞ and have faster than polynomial decay (this makes them more suitable for practical purpose according to the compact support property). ϕ and ψ are respectively symmetric around 0 and $\frac{1}{2}$ and ψ as an infinite number of vanishing moments (see [Mey90] for more details). The Battle-Lemarié wavelets are created by orthogonalizing B-spline functions and have exponential decay. A Battle-Lemarié wavelet with N vanishing moments is a piecewise polynomial of degree $N - 1$ belonging to C^{N-2}. See [Bat87, Lem88, Mey90, Mal98].

Daubechies wavelets.   The first non-trivial compactly supported and orthogonal wavelet basis have been constructed by Ingrid Daubechies [Dau88]. A Daubechies scaling function/wavelet pair of order M satisfies the two following dilation equations

$$\phi(x) = \sqrt{2} \sum_{k=0}^{2M-1} h_k \phi(2x - k)$$

and

$$\psi(x) = \sqrt{2} \sum_{k=-2M+2}^{1} g_k \phi(2x - k)$$

The coefficients $\{h_k\}$ (the $\{g_k\}$ are then given by (27)) are determined by solving the following $2M$ equations

$$\frac{1}{2} \sum_{l=0}^{2M-1} h_l h_{l-2k} = \delta_k, \; k = 0, \ldots, M - 1$$

and

$$\sum_{l=0}^{2M-1} (-1)^{l+1} l^k h_l = 0, \; k = 0, \ldots, M - 1$$

The first set of equations is a reformulation of equation (24) via the Wiener-Kintchine theorem[15]. The second set of equations expresses the fact that ψ must have M vanishing moments, i.e.

$$\int_{-\infty}^{+\infty} x^k \psi(x) \, dx = 0, \; k = 0, \ldots, M - 1$$

The resolution of this system of equations is done by finding a trigonometric polynomial $\hat{h}(\xi)$ satisfying equation (24) and having a root of

---

[15]For discrete-time signals: $\mathcal{F}\{\sum_l f_l f_{l-k}\} = |\hat{f}(\xi)|^2$ where $\hat{f}(\xi) = \sum_k f_k e^{-i\xi k}$.

multiplicity M at $\xi = \pi$. This is done by means of spectral factorization techniques (see—notably— [Dau88, BS94, Mal98]). The regularity of the multiresolution analysis increases as N increases (roughly like $0.2075M$ for large M [JS93, Mey94]). However, these wavelets cannot be symmetric (except for $M = 1$ which corresponds to the Haar wavelet). Figure 5 presents some scaling functions and wavelets of the Daubechies family. These functions have been generated using a cascade algorithm [Del93, BS94] which (roughly) consists in applying the inverse wavelet transform algorithm on respectively the $\{h_k\}$ and the $\{g_k\}$. The filter coefficients for $M = 2, \ldots, 10$ are available in [Dau88] (table 1, page 980).

Other orthogonal wavelet basis have been built using this philosophy. An interesting example is the coiflets contructed by Ronald Coifman [BCR91]. For this family, the scaling function also has some vanishing moments (except the first one) and these functions leads to discrete filters having $3M - 1$ non-zero coefficients. See [BCR91]. Other families of orthogonal and biorthogonal wavelets are (notably) designed in [CDF92, VH92], [Mal98] provides an up-to-date exposition on the subject.

7 Bi(multi)dimensional wavelet transform

This section generalizes the notion of wavelet transform and orthogonal multiresolution analysis in two dimensions. We first (briefly) introduce the spaces $L^2(\mathbb{R}^n)$ and $L^p(\mathbb{R}^n)$, then defines the continuous wavelet transform on $L^2(\mathbb{R}^2)$ and build orthogonal multiresolution analysis of $L^2(\mathbb{R}^2)$ by means of separable wavelet basis. Obviously, these extensions are necessary for being able to use the wavelet analysis in an image processing context.

## 7.1 Spaces: $L^2(\mathbb{R}^2)$, $L^2(\mathbb{R}^n)$ and $L^p(\mathbb{R}^n)$

The spaces $L^2(\mathbb{R}^2)$, $L^2(\mathbb{R}^n)$ and $L^p(\mathbb{R}^n)$ are "natural extensions" of $L^2(\mathbb{R})$ (space of square integrable function of one variable, see §1). $L^2(\mathbb{R}^2)$, the space of square integrable functions of two variables, is defined as

$$L^2(\mathbb{R}^2) = \left\{ f / \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |f(x,y)|^2 dx dy < \infty \right\}, \ f, g \in L^2(\mathbb{R}^2)$$
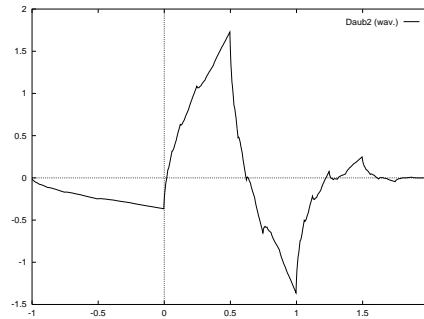
provided the following scalar product and the following norm

$$< f, g >_{L^2(\mathbb{R}^2)} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) g^*(x,y) dx dy, \ \|f\|^2_{L^2(\mathbb{R}^2)} = < f, f >_{L^2(\mathbb{R}^2)}$$
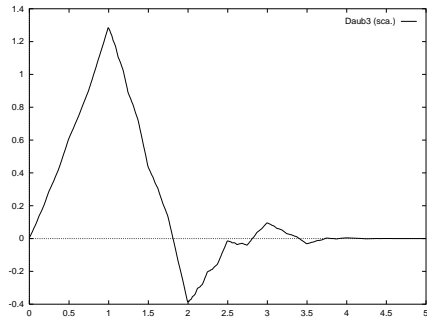
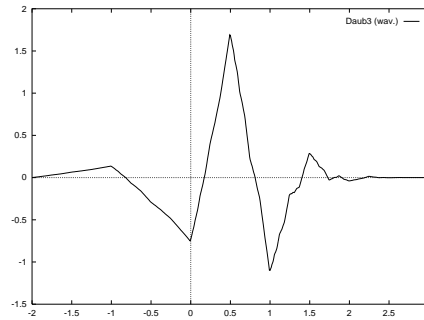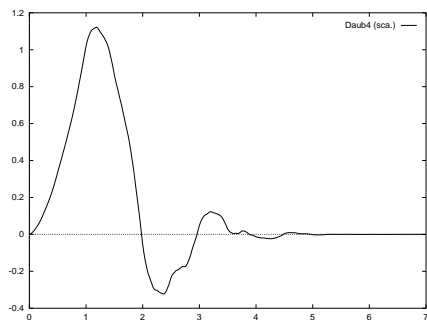**Figure 5** Some of the Daubechies scaling functions and wavelets.
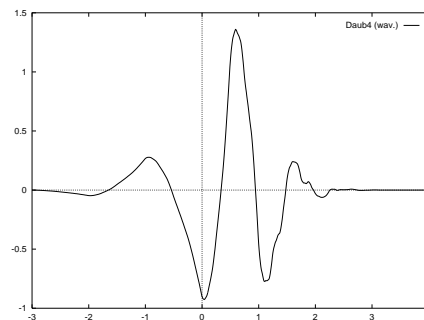

(a) ϕ_2.

(b) ψ_2.

(c) ϕ_3.

(d) ψ_3.

(e) ϕ_4.
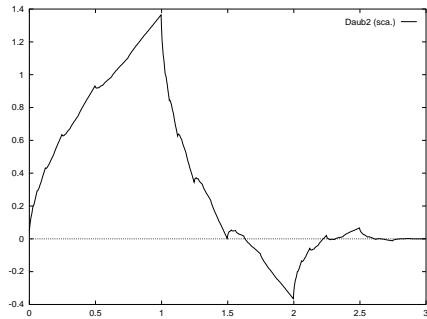
(f) ψ_4.

Obviously, the more general space $L^2(\mathbb{R}^n)$ is defined as

$$L^2(\mathbb{R}^n) = \left\{ f / \int_{\mathbb{R}^n} |f(\vec{z})|^2 d\vec{z} < \infty \right\}, \ \vec{z} \in \mathbb{R}^n \tag{32}$$

with

$$< f, g >_{L^2(\mathbb{R}^n)} = \int_{\mathbb{R}^n} f(\vec{z}) g^*(\vec{z}) d\vec{z}, \ \|f\|^2_{L^2(\mathbb{R}^n)} = < f, f >_{L^2(\mathbb{R}^n)}$$

Finally, $L^p(\mathbb{R}^n)$ is defined (for $1 \leq p \leq \infty$) by replacing 2 by p in equation (32) and the norm operator becomes (with the usual modification for $p = \infty$)

$$\|f\|_{L^p(\mathbb{R}^n)} = \left(\int_{\mathbb{R}^n} |f(\vec{z})|^p d\vec{z} \right)^{\frac{1}{p}}$$

The concepts of wavelet transform and multiresolution analysis can be easily generalized for $L^2(\mathbb{R}^n)$ (see the next subsection for $L^2(\mathbb{R}^2)$), but the generalization for $L^p(\mathbb{R}^n)$ is trickier[16], the reader is (notably) sent to [Mey90].

7.2 Continuous wavelet transform on $L^2(\mathbb{R}^2)$

The continuous wavelet transform of a function f of two variables belonging to $L^2(\mathbb{R}^2)$ is a staightforward generalization of the one-dimensional case presented in §3. Formally, given a wavelet Ψ

$$\mathcal{W}_f(a, b, b') = < f, \Psi_{a;b;b'} >_{L^2(\mathbb{R}^2)}, \ \Psi_{a;b;b'}(x, y) = \frac{1}{a} \Psi \left(\frac{x - b}{a}, \frac{y - b'}{a} \right)$$

The reader is (notably) sent to [Mal98] for more details (including the wavelet transform using wavelets with different spatial orientations).

7.3 Multiresolution analysis of $L^2(\mathbb{R}^2)$

A simple way of building an orthogonal multiresolution of $L^2(\mathbb{R}^2)$ consists in using separable wavelet basis which is done via the following theorem (notably prooved in [Mal98]).

Theorem 7 (Separable multiresolution) *Let ϕ and ψ (respectively) be the scaling function and the wavelet generating an orthogonal multiresolution on $L^2(\mathbb{R})$ and define*

$$\Psi^{(1)}(x, y) = \phi(x)\psi(y), \ \Psi^{(2)}(x, y) = \psi(x)\phi(y), \ \Psi^{(3)}(x, y) = \psi(x)\psi(y)$$

---

[16]In that general case, we have to deal with some paradoxal behaviors e.g. if we directly use an orthogonal decomposition of the form $f(x) = \sum_j \sum_k < f, \psi_{j;k} > \psi_{j;k}(x)$ (ψ is supposed to have at least one vanishing moment) on $f(x) = 1$ ($f \in L^\infty(\mathbb{R})$) we end up with $< f, \psi_{j;k} > = 0, \ \forall j, k \in \mathbb{Z}^2$ and hence with $1 = 0$! See [Mey90] for more details.

For $\alpha \in \{1,3\}$

$$\Psi^{(\alpha)}_{j;k;k'} = 2^j \Psi^{(\alpha)}(2^j x - k, 2^j y - k')$$

Then the families $\{\Psi^{(\alpha)}_{j;k;k'}\}_{\alpha \in \{1,3\}, \; k,k' \in \mathbb{Z}^2}$ **and** $\{\Psi^{(\alpha)}_{j,k,k'}\}_{\alpha \in \{1,3\}, \; j;k;k' \in \mathbb{Z}^3}$ *(respectively) form orthogonal basis of* W_j^2 *and* $L^2(\mathbb{R}^2)$.

The wavelet transform of an image is then organized as shown on figure 6. For example, the coefficients $\gamma^{(1)}_{j;k;k'} = < f, \Psi^{(1)}_{j;k;k'} >$ correspond to the one dimensional scalar product of f with $\phi_{j;k}$ according to the rows and to the scalar product of f with $\psi_{j;k'}$ according to the columns of the image. As discussed in [BS94, Sta92]: $\gamma^{(1)}_{j;k;k'}$ corresponds to the horizontal low frequencies and to the vertical high frequencies (vertical details) of $\mu_{j+1;k;k'} = < f, \Phi_{j+1,k,k'} >$ ($\Phi(x,y) = \phi(x)\phi(y)$), while $\gamma^{(2)}_{j;k;k'}$ and $\gamma^{(3)}_{j,k,k'}$ respectively correspond to its horizontal high/vertical low and horizontal high/vertical high frequencies (horizontal and diagonal details). The algorithm for computing the wavelet coefficients of an image becomes a straightforward extension of the decimated filter banks algorithm used in one dimension (see §6.1). Basically, it consists in applying the one dimensional algorithm on the rows followed by the same operation on the columns (the order does not matter) for each scale, i.e. for computing the wavelet coefficients at scale j from the scaling coefficients at scale $j + 1$. Note that it is possible to create non-separable wavelet basis of $L^2(\mathbb{R}^2)$ [KV92], but in spite of their interesting properties e.g. orthogonal, compactly supported and symetric wavelets (which is not possible in the one dimensional case), they are not very used in practice. Figure 7 shows the wavelet decompositions of some images.

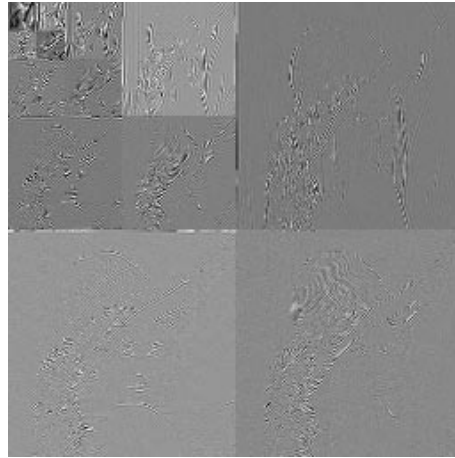**Figure 6** Organization of a two-dimensional wavelet decomposition.

Φ			$\Psi^{(2)}(\vec{z}) =$ $\psi(x)\phi(y)$ "High-Low"
		$\ldots$	
$\ldots$		$\ldots$	
$\Psi^{(1)}(\vec{z}) =$ $\phi(x)\psi(y)$ "Low-High"			$\Psi^{(3)}(\vec{z}) =$ $\psi(x)\psi(y)$ "High-High"

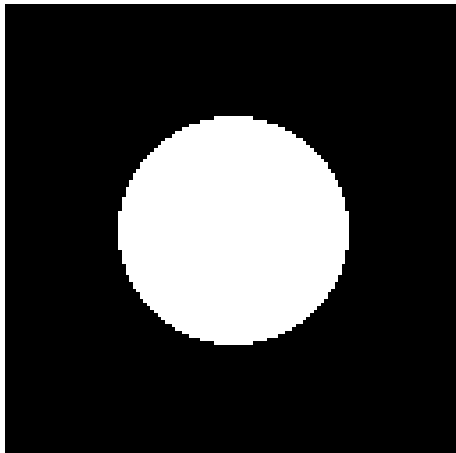$$\vec{z} = (x \; y)^\mathsf{T}$$

Figure 7 Wavelet decompositions of some images (Daubechies-8).
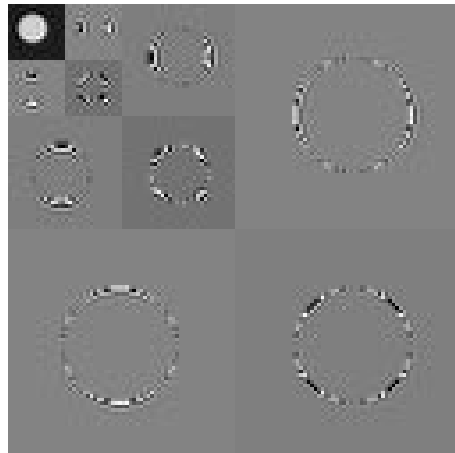

(a) "Lenna".

(b) Its wavelet decomposition.

(c) A disk.

(d) Its wavelet decomposition.

8 Denoising in the wavelet space

As a example of application we introduce the well-known wavelet-based denoising algorithms proposed a few years ago by David Donoho and Ian Johnstone [DJ91, DJKP92, DJ94, Don95]. As we shall see later, the (orthogonal) wavelet transform is optimal (in some sense which is to be defined) for characterising signals containing isolated singularities with a few high coefficients. This fact and the convenient behavior of gaussian white noise in the wavelet space can be used to design performant non-linear denoising methods.

8.1 Denoising via wavelet shrinkage

The purpose of denoising is to estimate a real function f from a set of corrupted measurements. A simple statistical model consists in considering that the samples are corrupted by an additive gaussian white noise i.e.

$$G_k = f_k + \sigma B_k, \ B_k \leadsto \mathcal{N}(0, 1) \ \textbf{iid}, \ \sigma \in \mathbb{R}^{*+}$$

In a orthogonal basis of $l^2(\{0, N-1\})$ e.g. $\{\theta_k\}_{k \in \{0, N-1\}}$, the expansion of a gaussian white noise remains a gaussian white noise [CSBF98, Mal98] (in all this section, $< ., . >$ and $\|.\|$ should be understood in a $l^2(\{0, N-1\})$ sense). Proof:

$$E[< B, \theta_k >] = \sum_{l=0}^{N-1} E[B_l]\theta_k^*[l] = 0$$

and

$$
\begin{aligned}
COV[< B, \theta_k >, < B, \theta_l >] &= E[< B, \theta_k >< B, \theta_l >^*] \\
&= \sum_{m=0}^{N-1} \sum_{n=0}^{N} \underbrace{E[B_k B_l]}_{\delta_{m,n}} \theta_k^*[m]\theta_l[n] \\
&= \sum_{m=0}^{N-1} \theta_k^*[m]\theta_l[m] =< \theta_k, \theta_l >^* = \delta_{k,l}
\end{aligned}
$$

Since a linear combination of iid gaussian random variables gives a gaussian random variable, and since the abscence of correlation $COV(X, Y) = 0$, $X \neq Y$ implies (for gaussian random variables) the independance of the two variables, we have $< B, \theta_k > \leadsto \mathcal{N}(0, 1)$ iid. In what follows, we present the two main philosophies for building some estimates of f, namely: coefficients attenuation ("implemented" via a soft thresholding) and coefficients selection (hard thresholding). The next two subsubsections summarize the ideas developed in [Mal98].

8.1.1 Coefficients attenuation

From the noisy signal G, we contruct an estimator of the form

$$\tilde{F} = \sum_{k=0}^{N-1} < G, \theta_k > \lambda[k]\theta_k$$

Here, we focus on non-linear estimators that depends on the realisation of G. We now consider the mean square error [17]

$$E[\|f - \tilde{F}\|^2] = \sum_{k=0}^{N-1} \underbrace{E[| < f, \theta_k > - < G, \theta_k > \lambda[k]|^2]}_{\varepsilon}$$

Since $< G, \theta_k >=< f, \theta_k > +\sigma < B, \theta_k >$, we have

$$\begin{aligned}
\varepsilon &= E[| < f, \theta_k > (1 - \lambda[k]) - \sigma < B, \theta_k > \lambda[k]|^2] \\
&= | < f, \theta_k > |^2(1 - \lambda[k])^2 + \sigma^2\lambda[k]^2
\end{aligned}$$

because $E[< B, \theta_k >] = 0$ and $E[| < B, \theta_k > |^2] = 1$. By solving $\frac{\partial \varepsilon}{\partial \lambda[k]} = 0$, one derives that ε is a minimum for

$$\lambda[k] = \frac{| < f, \theta_k > |^2}{| < f, \theta_k > |^2 + \sigma^2} \tag{33}$$

leading to the mean square error

$$E[\|f - \tilde{F}\|^2] = \sum_{k=0}^{N-1} \frac{| < f, \theta_k > |^2\sigma^2}{| < f, \theta_k > |^2 + \sigma^2}$$

Note that equation (33) can be seen as a "generalized" Wiener filter. If the basis functions $\{\theta\}_{k \in \{0, N-1\}}$ were the complex exponentials of Fourier analysis we would end up with

$$\lambda[k] = \frac{1}{1 + \frac{\sigma^2}{|\hat{f}[k]|^2}}$$

which is precisely the expression of the Wiener filter [GW92] for a point spread function equal to δ.

---

[17]Recall that an orthonormal basis of an abstract Hilbert space is a particular case of a Riesz basis with $A = B = 1$ [Dau92], therefore $\|f\|^2 = \sum_k | < f, e_k > |^2$ (see definition 10, page 64, as well).

8.1.2 Coefficients selection

A coefficient selection is performed by requiring that $\lambda[k]$ takes binary values, i.e. the estimator consists in selecting a subset of the noisy coefficients of G. In that case, it is obvious that the mean square error (still equal to: $E[\|f - \tilde{F}\|^2] = |<f, \theta_k>|^2(1 - \lambda[k])^2 + \sigma^2\lambda[k]^2$) is minimized by an operator of the form

$$\lambda[k] = \begin{cases} 1 & \textbf{if } |<f, \theta_k>|^2 \geq \sigma^2 \\ 0 & \textbf{otherwise} \end{cases}$$

The mean square error produced by this ideal selection procedure

$$E[\|f - \tilde{F}\|^2] = \sum_{k=0}^{N-1} \min(|<f, \theta_k>|^2, \sigma^2) \tag{34}$$

remains of the same order than the one introduced by the attenuation operator [Mal98]. Obviously, because of our lack of knowledge about $<f, \theta_k>$, the ideal coefficients attenuation and selection cannot be implemented. However, since the work of David Donoho (see notably [DJ94]), it is known that the performances of some thresholding estimators (applied on the empirical wavelet decomposition) are closed to the ones of the ideal procedures previously discussed.

8.1.3 Denoising in orthogonal wavelet basis

We have not yet spoken about denoising in orthogonal wavelet basis. Basically, the choice of the basis in which a non-linear operator is applied is crucial. The best (non-linear) approximation of a function f (with M vectors) in an orthogonal basis is given by

$$f_M = \sum_{|<f, \theta_k>| \geq \sigma} <f, \theta_k> \theta_k$$

while the approximation error is equal to

$$\|f - f_M\|^2 = \sum_{|<f, \theta_k>| < \sigma} |<f, \theta_k>|^2$$

For the ideal selection procedure previously dicussed, the mean square error (equation (34)) can therefore be written as

$$E[\|f - \tilde{F}\|^2] = \|f - f_M\|^2 + M\sigma^2$$

hence, the mean square error is small only if the approximation error and $M\sigma^2$ are both small, i.e. we want a basis in which the function

f is coded by a few large coefficients which characterize it relevantly. This, for example, eliminates the complex exponential basis for estimating a function containing some singularities: this type of signals generates non-neglectible coefficients in all the Fourier spectrum. The convenience of using orthogonal wavelet basis comes from the fact that a r-regularly (in the sense defined by Yves Meyer [Mey90]) orthogonal wavelet basis provides unconditionnal basis for a wide range of smoothness spaces [Mey90] (namely: Hölder, Sobolev, Besov spaces, . . .). For example, piecewise regular functions, i.e. functions containing isolated singularities (belonging to the Besov space(s)), are efficiently approximated with a few large coefficients [Don91, Mal98], see §12.3. The only *a priori* knowledge about the desired result is the order of a given Besov-norm and the implementation of the algorithm does not depend on its parameters (α, p, q) [Mey94].

The last problem deals with the necessity of approximating the ideal operators and estimating their parameters (e.g. σ). For example, a hard-thresholding operator

$$\tilde{F} = \sum_{k=0}^{N-1} \Lambda(< G, \theta_k >)\theta_k, \ \Lambda(x) = \left\{ \begin{array}{ll} x & \text{if } |x| \geq T \\ 0 & \text{otherwise} \end{array} \right. \tag{35}$$

with[18] $T = \sigma\sqrt{\log N}$, produces a mean square error which remains within a $2 \log N$ factor of the ideal error and is asymptotically optimal in a minimax sense [DJKP92]. The reader is notably sent to [Mal98, Don95] for some discussions on the operators (e.g. hard/soft thresholding) and the threshold choices.

9 Wavelet maxima

Multiscale edges have been introduced in order to deal with the problem of noise while performing a contour extraction task, e.g. [Ber87]. A popular strategy consists in using a detection operator which is the first or second order of a low-pass filter (e.g. gaussian filter [Can86] or the exponential filter [Der87]) in order to reduce the noise and carry

---

[18]In practice, the noise variance is not known and need to be estimated. This is done by using $\tilde{\sigma} = MED/.6745$ where MED denotes the median of the absolute values of the empirical wavelet coefficients at the finest scale (recall that $\int_{-.6745\sigma}^{+.6745\sigma} g_\sigma(x)dx = .5$ [DN89]). When f is piecewise smooth, it generates only a small number of non vanishing coefficients at the finest scale (the wavelet overlaps the singularities for only a few values of the translation parameter) and the median is not very sensitive to a few outliers.

out the edge detection. Obviously, this method have a fundamental disadvantage: the "good" localization and "good" detection criteria [Can86] (see 11.1) are dual and cannot be simultaneously arbitrary small. In order to overcome this limitation, Fredrik Berghlom [Ber87] has proposed a procedure, known as "edge focusing", which consists in computing the output of the Canny detector for different values of σ (i.e. scales) and detecting the edges using a coarse-to-fine tracking.
This philosophy has been retained for designing feature-based image representations[19], using "classical" multiscale decompositions (e.g [HM89]) or the wavelet transform [Mal91, MH92, MZ92]. These representations allow to reconstruct an approximation of the original image from its multiscale edges (the uniqueness and stability of these representations is notably adressed in [Ber91, Ber92, BB93, Mey94]). Because of the possibility of "practical" reconstruction, it has been foreseen by Stéphane Mallat [MZ92] that many image processing tasks could be implemented using edge-based algorithms.

9.1   Multiscale edges

In this section, we only focus on the wavelet maxima representation. Information about the zero-crossing representation are available (notably) in [Mal91]. The following theorem (notably prooved in [Mal98]) implies that a wavelet having one vanishing moment corresponds to the first order derivative of a smoothing operator.

Theorem 8 *A wavelet* ψ *with a fast decay has* n *vanishing moments if and only if there exists* θ *(a smoothing operator) with a fast decay such that*

$$\psi(x) = (-1)^n \frac{d^n \theta(x)}{dx^n}$$

As a consequence

$$W_f(a, b) = a^n \frac{d}{db^n}(f \otimes \bar{\theta}_a)(b)$$

where $\theta_a(x) = a^{-\frac{1}{2}}\theta(x/a)$. *Moreover,* ψ *has no more than* n *vanishing moments if and only if* $\int_{-\infty}^{+\infty} \theta(x)\,dx \neq 0$.

Hence,

$$\psi(x) = -\frac{d\theta}{dx}$$

---

[19]Known as adaptive quasi-linear representations (AQLR).

36

The dyadic wavelet transform (the use of a dyadic wavelet transform is motivated by its translation invariance and its redundancy) can therefore be interpreted as

$$W_f(2^j, b) = 2^j \frac{d}{db}(f \otimes \bar{\theta}_{2^j})(b)$$

As j increases, $\mathcal{W}_f(2^j, b)$ is smoother. For example, if θ is a Gauss function, one ends up with a multiscale Canny operator. Under this condition the dyadic wavelet transform provides a multiscale gradient from which the points of sharp variation can be extracted.

9.1.1   Quadratic spline wavelet

For being able to use the *algorithme à trous* (see §3.3), it is required that

$$\hat{\phi}(\xi) = \hat{h}(\xi/2)\hat{\phi}(\xi/2) \text{ and } \hat{\psi}(\xi) = \hat{g}(\xi/2)\hat{\phi}(\xi/2)$$

where $\hat{h}(\xi)$ and $\hat{g}(\xi)$ are the Fourier transform of two discrete filters (the same constraints should be satisfied by $\hat{\tilde{\phi}}(\xi)$ and $\hat{\tilde{\psi}}(\xi)$). Since ψ should be the first order of a smoothing operator, $\hat{\psi}(\xi)$ must have a zero of order 1 at $\xi = 0$. Because $\hat{\phi}(0) = 0$ the constraint is moved onto $\hat{g}(\xi)$. Moreover, $\hat{h}(\xi)$ is chosen such that $\psi(x)$ is antisymetrical, is as regular as possible and has a small compact support. Stéphane Mallat [MZ92] has proposed the following family of filters

$$\hat{h}(\xi) = \hat{\tilde{h}}(\xi) = e^{\frac{i\xi}{2}}\cos(\xi/2)^{2n+1}, \ \hat{g}(\xi) = 4ie^{\frac{i\xi}{2}}\sin(\xi/2) \text{ and } \hat{\tilde{g}}(\xi) = \frac{1 - |\hat{h}(\xi)|^2}{\hat{g}(\xi)}$$

This leads to the filter coefficients, available in [MZ92] (table 1, page 728), and to the following scaling function, wavelet and smoothing operator

$$\hat{\phi}(\xi) = \text{sinc}(\xi/2)^{2n+1}, \ \hat{\psi}(\xi) = i\xi\text{sinc}(\xi/4)^{2n+2} \text{ and } \hat{\theta}(\xi) = \text{sinc}(\xi/4)^{2n+2}$$

note that $\text{sinc} x = \frac{\sin x}{x}$. Choosing $2n + 1 = 3$ leads to a scaling function and a smoothing operator which are respectively a cubic and a quadratic spline. Figure 8 shows the modulus of their Fourier transforms.

9.1.2   *Algorithme à trous* in two dimensions

In two dimensions, the dyadic wavelet transform is (most of the time) defined by using two spatially oriented separable wavelets [Mal98]

$$\Psi^{(1)}(x, y) = \psi(x)\phi(y) = -\frac{\partial}{\partial x}\Theta^{(1)}, \ \Psi^{(2)}(x, y) = \phi(x)\psi(y) = -\frac{\partial}{\partial y}\Theta^{(2)} \quad (36)$$

37

Figure 8 Quadratic spline wavelet.

(a) $|\hat{\phi}(\xi)|$ (b) $|\hat{\psi}(\xi)|$ (c) $|\hat{\theta}(\xi)|$
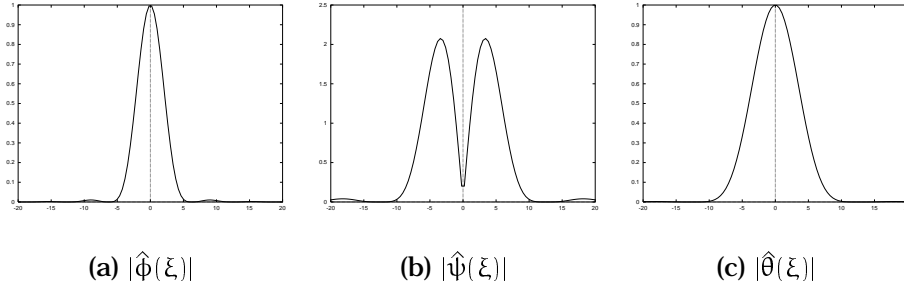
and a separable scaling function $\Phi(x, y) = \phi(x)\phi(y)$. The resulting algorithm becomes (roughly) a straightforward extension of the one dimensional case (§3.3) which consists in iteratively applying separable filters on the scaling coefficients for obtaining the scaling and wavelet coefficients at the next scale. As usual, for practical images, the first coefficients are given by the grey-scale values of the original image (see 3.4). Figure 9 shows the dyadic wavelet transform of the "lenna" image, computed using the scaling function/wavelet pair presented in the previous subsection. More details are available in [MZ92, Mal98].

9.1.3 Contours extraction

From the definition of $\Psi^{(1)}$ and $\Psi^{(2)}$, it directly follows that the wavelet coefficients are proportionnal to the gradient of the image smoothed by Θ_{2^j} ($\Theta^{(1)} \approx \Theta^{(2)}$ [MZ92]) i.e.

$$\left(\begin{array}{c} \mathcal{W}_f^{(1)}(2^j, b, b') \\ \mathcal{W}_f^{(2)}(2^j, b, b') \end{array} \right) = \left(\begin{array}{c} 2^j \frac{\partial}{\partial b}(f \otimes \bar{\Theta}_{2^j})(b, b') \\ 2^j \frac{\partial}{\partial b'}(f \otimes \bar{\Theta}_{2^j})(b, b') \end{array} \right) = 2^j \vec{\nabla} f \otimes \bar{\Theta}_{2^j}(b, b')$$

This information can therefore be used (as in a classical edge detector) for extracting the multiscale edges. The modulus of the gradient is proportionnal to the modulus of the wavelet coefficients
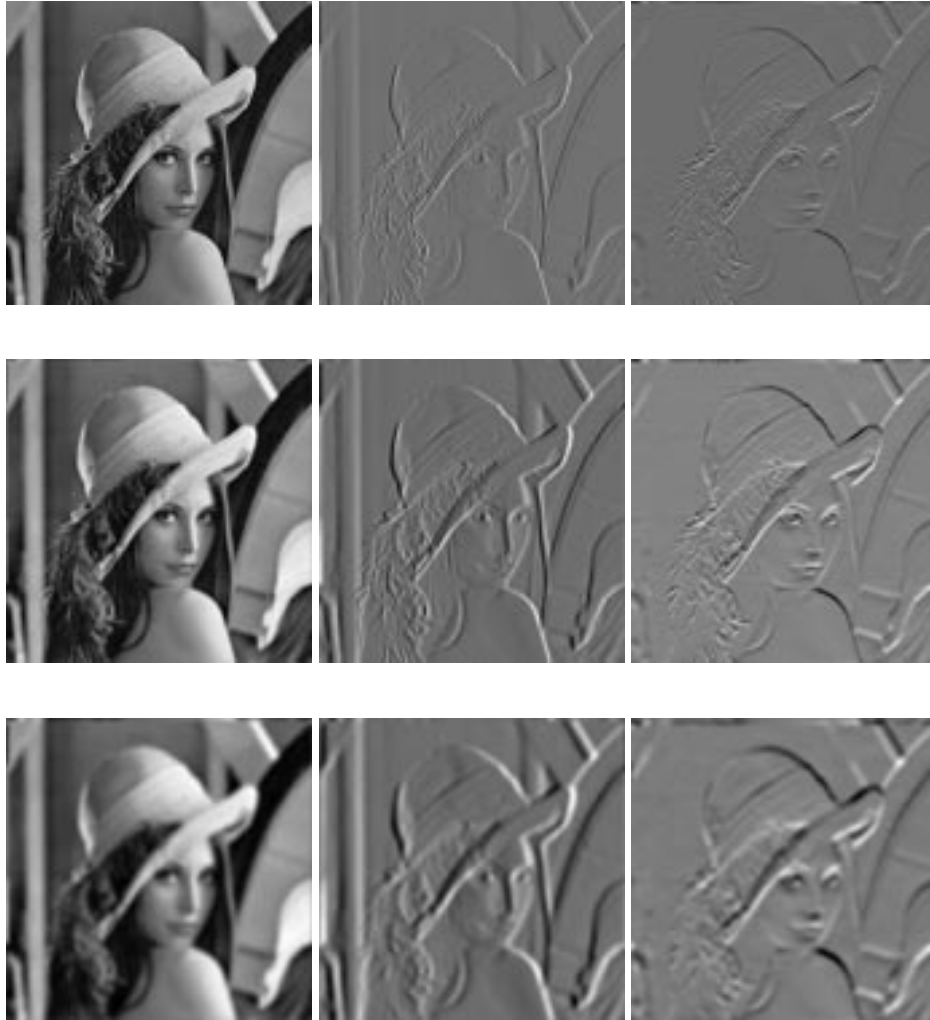
$$|\vec{\nabla} f \otimes \bar{\Theta}_{2^j}(b, b')| \propto \sqrt{|\mathcal{W}_f^{(1)}(2^j, b, b')|^2 + |\mathcal{W}_f^{(2)}(2^j, b, b')|^2}$$

and its orientation is given by

$$\alpha(2^j, b, b') = \tan^{-1} \frac{\mathcal{W}_f^{(2)}(2^j, b, b')}{\mathcal{W}_f^{(1)}(2^j, b, b')}$$

The detection consists in two main steps: it is first necessary to extract the local maxima of the gradient norm in the gradient direction and

Figure 9 Beginning of a two-dimensional dyadic wavelet transform.

(g) Scaling coeff.   (h) Wave. co. ($\psi^{(1)}$).   (i) Wave. co. ($\psi^{(2)}$).

secondly, to suppress the non-significant local maxima via a thresh-olding operation. This last operation in often implemented using an hysteresis thresholding (more details are available in [Der, CP95] and in §11 as well).
The set of values,

$$\Omega = \{\{(b_k \ b_k')^\mathsf{T}, (\mathcal{W}_f^{(1)}(2^j, b_k, b_k') \ \mathcal{W}_f^{(2)}(2^j, b_k, b_k'))^\mathsf{T}\}_{k=1,\cdots,N_j}\}_{j\in\mathbb{Z}}$$

where $(b_k \ b_k')^\mathsf{T}$ denotes the coordinates of a local maximum, is called the wavelet maxima representation of the image. For a digital $N \times N$ image, the wavelet maxima representation obviously becomes

$$\Omega = \{\{(i_k \ i_k')^\mathsf{T}, (\mathcal{W}_f^{(1)}(2^j, i_k, i_k') \ \mathcal{W}_f^{(2)}(2^j, i_k, i_k'))^\mathsf{T}\}_{k=1,\cdots,N_j}\}_{j=0,\cdots,\log_2 N}$$

where $(i_k \ i_k')^\mathsf{T}$ is an integer-valued vector. See figure 10.

9.2 Reconstruction from local maxima

As pointed in (notably) [Ber91, Mey94], the wavelet maxima represen-tation does not characterize uniquely a given function f. However, two functions having the same local maxima differ mainly and only slightly on their high-frequency content, which makes the reconstruction suit-able for practical purposes [Mal98, Mey94]. Here, we shortly present the alternate projection algorithm, introduced in Stéphane Mallat's articles. However, other alternative algorithms have been proposed in [Car, CV95]. We restrict ourselves to the one dimensional case, since the two dimensional algorithm is a straightforward extension and is fully presented in [MZ92].

9.2.1 The alternate projection algorithm

Basically, our goal is to find a sequence of functions $\{g_j\}_{j\in\mathbb{Z}}$ such that: it is the wavelet transform of a function of $L^2(\mathbb{R})$, it has the same wavelet maxima as $\mathcal{W}_f$ and **not more**. As pointed in [MZ92], this last condition (not more) is not convex and cannot be implemented easily, thus it is relaxed and replaced by requiring the following Sobolev norm to be minimum

$$\|\{g_j\}_{j\in\mathbb{Z}}\|_K^2 = \sum_j \|g_j\|_{L^2(\mathbb{R})}^2 + 2^{2j} \left\|\frac{dg_j}{dx}\right\|_{L^2(\mathbb{R})}^2$$

The constraint on $\|dg_j/dx\|_{L^2(\mathbb{R})}^2$ allows to control the appearance of spurious maxima and the multiplication by 2^j express the fact that g_j should be smoother as j increases. Let K be the set of all the se-quences $\{g_j\}_{j\in\mathbb{Z}}$ such that $\|\{g_j\}_{j\in\mathbb{Z}}\|_K$ is finite. Let V denote the space of

Figure 10 Multiscale edges extracted from the "lenna" image.
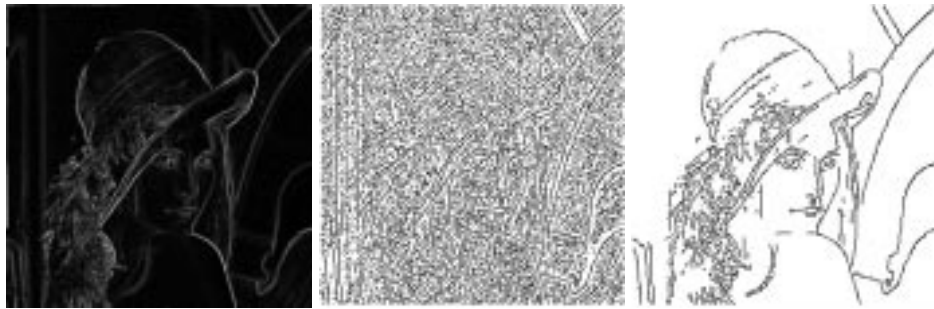
(g) Gradient norm.   (h) Local maxima.   (i) Hysteresis thresh.

all the sequences $\{g_j\}_{j \in \mathbb{Z}}$ such that their are the wavelet transform of a function belonging to $L^2(\mathbb{R})$ i.e. such that the sequence $\{g_j\}_{j \in \mathbb{Z}}$ satisfies the reproducing kernel equation (equation (10) or (11)). Let Γ be the set of sequences $\{g_k\}_{j \in \mathbb{Z}}$ such that for all j and for all maxima position x_k we have

$$W_f(2^j, x_k) = g_j(x_k)$$

The alternate projection algorithm converges to a sequence $\{g_j^\star\}_{j \in \mathbb{Z}}$, lying in $\Lambda = V \cap \Gamma$, by alternatively project the current sequence onto V and Γ. Starting from the initial guess $\{g_j\}_{j \in \mathbb{Z}}^{(0)} \in V$ (in general $g_j(x) = 0, \ \forall j$), the algorithm is simply expressed as

$$\{g_j\}_{j \in \mathbb{Z}}^{(k+1)} = P_V(P_\Gamma(\{g_j\}_{j \in \mathbb{Z}}^{(k)}))$$

where P_V and P_Γ are respectively the orthogonal projectors that project a set of functions of K onto (respectively) V and Γ.

P_V is simply equal to $W \circ W^{-1}$ where W denotes the wavelet transform operator, i.e. P_V is implemented by taking the inverse wavelet transform of $\{g_j\}_{j \in \mathbb{Z}} \in K$ followed by a wavelet transform. The P_Γ operator is trickier, it transforms a sequence in $\{g_j\}_{j \in \mathbb{Z}} \in K$ into a sequence $\{h_j\}_{j \in \mathbb{Z}} \in \Gamma$ such that its K-norm is minimum. After solving a simple problem of calculus of variation [Mal98] (again!), one finds that $h_j(x) = \epsilon_j(x) + g_j(x)$ where

$$\epsilon_j(x) = \alpha e^{2-jx} + \beta e^{-2^{-j}x}, \ x \in [x_k, x_{k+1}]$$

x_i and x_{i+1} are the absissa of two consecutive local maxima and $\alpha, \ \beta$ should be chosen such that

$$\begin{cases} \epsilon_j(x_k) & = \ \mathcal{W}_f(2^j, x_k) - g_j(x_k) \\ \epsilon_j(x_{k+1}) & = \ \mathcal{W}_f(2^j, x_{k+1}) - g_j(x_{k+1}) \end{cases}$$

An implementation of P_Γ is available page 58. For more details (stability of the reconstruction, rate of convergence, . . .) the reader is sent to the articles already cited in this subsection.

9.2.2   Practical considerations

Even if the representation is not unique, the algorithm is suitable for practical purpose. Figure 11, presents some experimental results on the "lenna" image. On a visual point of vue there is no difference between images (a) and (b) (image (b) has been reconstructed using all the wavelet maxima of the original one). If we consider only the significant maxima, we loose some textural information, but the reconstructed images still approximate correctly the original image ((c) and (d) have

been computed by considering smaller sets of local maxima). This phenomenon is stated in [Mal98], as well. Our experiments also suggest that 15 to 30 iterations are required for building a reasonnable approximation of an image. Obviously, due to its iterative nature, the reconstruction algorithm is computationaly intensive and is not suitable (in its direct form) for real-time processing.

9.3   Estimation of K, α and σ

An interesting property of the wavelet maxima representation is that it allows to estimate some of the parameters that characterize an isolated singularity[20]. The following theorem, prooved in [Mey90], relates the decay of the wavelet coefficients to the Hölder regularity (see definition page 64) of the original function.

Theorem 9 *A function* f *is uniformly Hölder-α over the interval* $[a, b]$ *if and only if there exists* $K > 0$ *such that*

$$\forall c \in [a, b], \ |\mathcal{W}_f(2^j, c)| \le K2^{j\alpha}$$

This theorem also holds for tempered distributions, e.g. δ. As a basic consequence: if a function is uniformly Hölder-α ($\alpha < 0$) the amplitude of the wavelet coefficients decrease as j increase, while if $\alpha > 0$ the coefficients increase with the scale parameter. Now, if we reintroduce the gaussian model discussed in §**??**, i.e. the function f contains an isolated Hölder-α singularity at $\nu \in [a, b]$ smoothed by a gaussian operator, we have

$$\mathcal{W}_{f \otimes g_\sigma}(2^j, c) = 2^j \frac{d}{dc}(f \otimes g_\sigma \otimes \bar{\theta}_{2^j})(c)$$

Assuming that $g_\sigma \otimes \bar{\theta}_{2^j} \approx \bar{\theta}_\Sigma$, $\Sigma = \sqrt{\sigma^2 + 2^{2j}}$ leads to

$$\mathcal{W}_{f \otimes g_\sigma}(2^j, c) \approx \frac{2^j}{\Sigma} \mathcal{W}_f(\Sigma, c)$$

Therefore, if f is Hölder-α over $[a, b]$, we end up with

$$\exists K > 0, \ \forall c \in [a, b], \ |\mathcal{W}_{f \otimes g_\sigma}(2^j, c)| \le K2^j \Sigma^{\alpha - 1}$$

Which can be rewritten as

$$\log_2 |\mathcal{W}_{f \otimes g_\sigma}(2^j, c)| \le \log_2 K + j + \frac{\alpha - 1}{2} \log_2(\sigma^2 + 2^{2j}) \tag{37}$$

---

[20]The non-isolated singularities case is notably addressed in [MH92, HM93].

Figure 11 Reconstruction via the alternate projection algorithm.

(a) Original image.

(b) Reconstruction from all the local maxima (30 it.).

(c) Reconstruction from a subset of the local maxima (30 it.).

(d) Reconstruction from a smaller subset of the local maxima (30 it.).

Given the wavelet maxima trace $\{c_j, |\mathcal{W}_{f \otimes g_\sigma}(2^j, c_j)|\}_{1 \le j \le J}$ of the singularity at $\nu \in [a, b]$, Stéphane Mallat (in [MH92, MZ92]) proposes to estimate K, α and σ by finding their values such that equation (37) is as close as possible of an equality. This is done by optimizing

$$\Delta = \sum_{j=1}^{J} \left(\log_2 |\mathcal{W}_{f \otimes g_\sigma}(2^j, c_j)| - \log_2 K - j - \frac{\alpha - 1}{2} \log_2(\sigma^2 + 2^{2j}) \right)^2$$

using a gradient descent method, e.g.

$$\begin{pmatrix} K^{(k+1)} \\ \alpha^{(k+1)} \\ \sigma^{(k+1)} \end{pmatrix} = \begin{pmatrix} K^{(k)} \\ \alpha^{(k)} \\ \sigma^{(k)} \end{pmatrix} + \rho \begin{pmatrix} \partial\Delta/\partial K \\ \partial\Delta/\partial\alpha \\ \partial\Delta/\partial\sigma \end{pmatrix}$$

where $K^{(0)}$, $\alpha^{(0)}$ and $\sigma^{(0)}$ are chosen arbitrarly. These parameters express different characteristics of the singularity: K is related to its amplitude, α to its type and σ to its degree of smoothness. However, the raw wavelet maxima representation is not sufficient for estimating these values, since we need to access the wavelet maxima trace of the singularity, i.e. we need to link the wavelet maxima across scales. This representation, known as the wavelet maxima tree, is briefly presented in the next subsection.

9.4   The wavelet maxima tree

In order to link the wavelet maxima across scales one can use an *ad hoc* algorithm, as the one proposed in [MZ92]. Basically, it consists in linking two successive wavelet maxima if they are close to each other and if their corresponding values are of the same sign and of the same order. However, smarter algorithms can be found in the litterature such as the one in [Lu93]. From the sets of wavelet maxima at scales j and $j + 1$

$$\Omega_j = \left\{ c_k^{(j)}, \mathcal{W}_f\left(2^j, c_k^{(j)}\right) \right\}_{k=1,\dots,N_j}, \quad \Omega_{j+1} = \left\{ c_k^{(j+1)}, \mathcal{W}_f\left(2^j, c_k^{(j+1)}\right) \right\}_{k=1,\dots,N_{j+1}}$$

Jian Lu proposes a measure of interaction based on the reproducing kernel corresponding to the wavelet (recall equations (10) and (11))

$$\Gamma\left(c_k^{(j+1)}, c_l^{(j)}\right) = \mathcal{W}_f\left(2^{j+1}, c_k^{(j+1)}\right) \kappa'_{2^{j+1}, 2^j}\left(c_k^{(j+1)} - c_l^{(j)}\right)$$

The wavelet maxima tree is then constructed reccursively using a coarse-to-fine strategy. For a given maxima at scale j, the maxima at scale $j + 1$ which maximizes $\Gamma\left(c_k^{(j+1)}, c_l^{(j)}\right)$ is marked as its parent node. The

45

main child of a parent node at scale $j+1$ is the one which maximizes $\Gamma\left(c_l^{(j)}, c_k^{(j+1)}\right)$ among its children. A main branch connects a maxima to the tip end of the tree and provides the approximation of the trace of a wavelet maxima required for estimating the parameters discussed in the previous subsection. Note that this algorithm works only if the smoothing operator θ is a good approximation of a Gauss function, this implies that the behavior of the wavelet maxima is convenient, e.g. it satisfies a causality property (any feature at a coarser scale must have its origin at a finer scale), see [Ber87, Lu93]. This algorithm also requires to estimates the reproducing kernel of the wavelet transform. A more detailled presentation is available in [Lu93].

Contents

References

[Bat87]   G. Battle. A block spin construction of ondelettes. <u>Communications on Mathematical Physics</u>, (110), 1987.

[BB93]   Z. Berman and J.S. Baras. Properties of the multiscale maxima and zero crossings representations. <u>IEEE Transactions on Signal Processing</u>, 41(12), 1993.

[BCR91]   G. Beyklin, R. Coifman, and V. Rokhlin. Fast wavelet transform and numerical algorithm i. <u>Communications on Pure and Applied Mathematics</u>, (44), 1991.

[Ber87]   F. Bergholm. Edge focusing. <u>IEEE Transactions on Pattern Analysis and Machine Intelligence</u>, 9(6), 1987.

[Ber91]   Z. Berman. The uniqueness question of discrete wavelet maxima representation. Technical Report TR 91-48r1, University of Maryland, 1991.

[Ber92]   Z. Berman. <u>Generalizations and Properties of the Multiscale Maxima and Zero-Crossings Representations</u>. PhD thesis, University of Maryland, 1992.

[BH96]   B. Burke Hubbard. <u>Ondes et ondelettes, la saga d'un outil mathématique</u>. Éditions Pour La Science (diffusion Belin), 1996.

[BS94]   M. Bourges-Sévenier. Réalisation d'une bibliothèque c de fonctions ondelettes. Technical Report 2362, INRIA, 1994.

[Can86]   J. Canny. A computational approach to edge detection. <u>IEEE Transactions on Pattern Analysis and Machine Intelligence</u>, 8(6), 1986.

[Car]   R.A. Carmora. Extrema reconstruction and spline smooth-
           ing, variations on an algorithm of mallat & zhong. Princeton
           University.

[CDF92]   A. Cohen, I. Daubechies, and J-C. Feauveau. Biorthogonal
           bases of compactly supported wavelets. <u>Communications on
           Pure and Applied Mathematics</u>, (45), 1992.

[CDJV92]   A. Cohen, I. Daubechies, B. Jawerth, and P. Vial. Multireso-
           lution analysis, wavelet and fast algorithms on the interval.
           <u>Comptes rendus de l'Académie des Sciences de Paris</u>, 316,
           1992.

[CDV93]   A. Cohen, I. Daubechies, and P. Vial. Wavelets bases
           on the interval and fast algorithms. <u>Journal of Applied
           Mathematics and Computational Harmonic Analysis</u>, 1,
           1993.

[CP95]   Jean-Pierre Cocquerez and Sylvie Philipp, editors. <u>Analyse
           d'image : filtrage et segmentation</u>. Enseignement de la
           physique. Masson, 1995.

[CSBF98]   P. Carré, P. Simard, D. Boichu, and C. Fernandez.
           Débruitage par transformée en ondelettes non-décimée : ex-
           tension aux acquisitions multiples. Preprint, 1998.

[CV95]   Z. Cvetković and M. Vetterli. Discrete-time wavelet extrema
           representation: Design and consistent reconstruction. <u>IEEE
           Transactions on Signal Processing</u>, 43(3), 1995.

[CW98]   E. Cornell and K. Wieman. La condensation de bose-
           einstsein. <u>Pour la science (Édition française de Scientific
           American</u>, (247), 1998.

[Dau88]   I. Daubechies. Orthogonal basis of compactly sup-
           ported wavelets. <u>Communications on Pure and Applied
           Mathematics</u>, (41), 1988.

[Dau90]   I. Daubechies. The wavelet transform, time-frequency
           localisation and signal analysis. <u>IEEE Transactions on
           Information Theory</u>, 36(5), 1990.

[Dau92]   I. Daubechies. <u>Ten Lectures on Wavelets</u>. SIAM, 1992.

[Del93]   B. Delyon. Ondelettes orthogonales et biorthogonales. Tech-
           nical Report 1985, INRIA, 1993.

[Der]   Rachid Deriche. Techniques d'extraction de contours.

[Der87]   R. Deriche. Using canny's criteria to derive a reccursively implemented optimal edge detector. International Journal of Computer Vision, 1(1), 1987.

[DJ91]   D.L. Donoho and I.M. Johnstone. Minimax estimation via wavelet shrinkage. IMS special invited lecture at the Annual Meeting of the Institute of Mathematical Statistics of Atlanta, 1991.

[DJ94]   D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage, 1994.

[DJKP92]   D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptotia? Presented at the Oberwolfach Meeting "Mathematische Stochastik", 1992.

[DL92]   R.A. Devore and B.J. Lucier. Wavelets. Acta Numerica, 1, 1992.

[DN89]   J. Daintith and R.D. Nelson. The Penguin Dictionary of Mathematics. Penguin Books, 1989.

[Don91]   D.L. Donoho. Unconditional bases are optimal bases for data compression and statistical estimation. Paper presented as "Wavelets + Decision Theory = Optimal Smoothing" at the "Wavelets and Applications Workshop" (Luminy, France) and the workshop on "Trends in the Analysis of Curve Data" (University of Heidelberg), 1991.

[Don95]   D.L. Donoho. De-noising by soft thresholding. IEEE Transactions on Information Theory, 41(3), 1995.

[DP88]   R.A. Devore and V.A. Popov. Interpolation of besov spaces. Transactions of the American Mathematical Society, 305(1), 1988.

[DS52]   R.J. Duffin and A.C. Schaeffer. A class of nonharmonic fourier series. Transactions of the American Mathematical Society, (72), 1952.

[FJ85]   M. Frazie and B. Jawerth. Decomposition of besov spaces. Indiana University Mathematical Journal, 34(4), 1985.

[FS97]   H.G. Feichtinger and T. Strohmer. Gabor Analysis and Algorithms, Theory and Applications. Birkhauser, 1997.

[Gab46]   D. Gabor. Theory of communication. <u>Journal of the IEE</u> <u>(London)</u>, 93(3), 1946.

[GM84]   A. Grossmann and J. Morlet. Decomposition of hardy func-tions into square integrable wavelets of constant shape. <u>Journal of Mathematical Analysis</u>, 15(4), 1984.

[GW92]   R.C. Gonzales and R.E. Woods. <u>Digital Image Processing</u>. Addison-Wesley, 1992.

[HBB92]   F. Hlawatsch and G.F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. <u>IEEE</u> <u>Signal Processing Magazine</u>, April 1992.

[HD97]   T. Ha Duaong. Techniques mathématiques de l'ingénieur. Cours de l'Université de Technologie de Compiègne (cycle ingénieur), 1997.

[HM89]   R. Hummel and R. Moniot. Reconstruction from zero cross-ings in scale space. <u>IEEE Transactions on Acoustic, Speech</u> <u>and Signal Processing</u>, 37(12), 1989.

[HM93]   W-L. Hwang and S. Mallat. Characterization of self-similar multifractals wit wavelets maxima. Technical Report 641, Courant Institute of Mathematical Sciences, 93.

[JS93]   B. Jawerth and W. Sweldens. Overview of wavelet based multiresolution analysis, 1993.

[KF61]   A.N. Kolmogorov and S.V. Fomin. <u>Elements of the Theory</u> <u>of Functions and Functional Analysis</u>, volume 2. Graylock Press, 1961. Translated from the First Russian Edition by H. Kamel and H. Koum.

[KV92]   J. Kovačević and M. Vetterli. Nonseparable multidimen-sional perfect reconstruction filter banks and wavelet bases for r^n. <u>IEEE Transactions on Information Theory</u>, 38(2), 92.

[Lem88]   P.G. Lemarié. Ondelettes à localisation exponentielle. <u>Journal des mathématiques pures et appliquées</u>, (67), 1988.

[Lu93]   J. Lu. <u>Signal Recovery and Noise Reduction with Wavelets</u>. PhD thesis, Dartmouth College, 1993.

[Mal89]   S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. <u>IEEE Transactions on Pattern</u> <u>Analysis and Machine Intelligence</u>, 11(7), 1989.

51

[Mal91]   S. Mallat. Zero crossings of a wavelet transform. <u>IEEE Transactions on Information Theory</u>, 17(4), 1991.

[Mal98]   S. Mallat. <u>A Wavelet Tour of Signal Processing</u>. Academic Press, 1998.

[Mey90]   Y. Meyer. <u>Ondelettes et opérateurs</u>, volume 1. Hermann, 1990.

[Mey94]   Y. Meyer. <u>Les ondelettes : algorithmes et applications</u>. Armand Colin, 1994.

[MH92]   S. Mallat and W.L. Hwang. Singularity detection and processing with wavelets. <u>IEEE Transactions on Information Theory</u>, 38(2), 1992.

[MZ92]   S. Mallat and S. Zhong. Characterization of signals from multiscale edges. <u>IEEE Transactions on Pattern Analysis and Machine Intelligence</u>, 14(7), 1992.

[PB]   V. Perrier and C. Basdevant. Besov norms in terms of the continuous wavelet transform, application to structure functions.

[RD92]   O. Rioul and P. Duhamel. Fast algorithms for discrete and continuous wavelet transform. <u>IEEE Transactions on Information Theory</u>, 38(2), 1992.

[RSN55]   F. Riesz and B. Sz.-Nagy. <u>Functional Analysis</u>. Frederic Ungar Publishing, 1955. Translated from the Second French Edition by L. Boron.

[SB86]   M.J.T. Smith and T.P. Barnwell. Exact reconstruction techniques for tree-structured subband coder. <u>IEEE Transactions on Acoustics, Speech and Signal Processing</u>, 34(3), 1986.

[She92]   M.J. Shensa. The discrete wavelet transform: Wedding the à trous and mallat algorithm. <u>IEEE Transactions on Signal Processing</u>, 40(10), 1992.

[Sta92]   J-L. Starck. <u>Analyse en ondelette et imagerie à haute résolution angulaire</u>. PhD thesis, Université de Nice-Sophia Antipolis, 1992.

[Str94]   G. Strang. Wavelets. <u>American Scientist</u>, (82), 1994.

[Tri78]   H. Triebel.   Interpolation Theory, Function Spaces, Differential Operators. North Holland Publishing, 1978.

[VH92]   M. Vetterli and C. Herley. Wavelets and filter banks: Theory and design. IEEE Transactions on Signal Processing, 40(9), 1992.

[Vil48]   J. Ville. Théorie et applications de la notion de signal analytique. Cables et transmissions, 2A(1), 1948.

[Weiar]   E.W. Weisstein.   The CRC Concise Encyclopedia of Mathematics. CRC Press, to appear.

[Zyg68]   A. Zygmund. Trigonometric Series. Cambridge University Press, 1968.

10   Wavelet transform algorithms

This section presents the implementation, using the C-language, of a few wavelet transform algorithms. Programs 10.1 and 10.2 propose some implementations of (respectively) the *algorithme à trous* and its inverse (see §3.3 and algorithm 2, page 12). Programs 10.3 and 10.4 give implementations of the fast wavelet transform and its inverse discussed in §6.1 (see algorithm 4, page 22, as well). Program 10.5 gives an implementation of the P_Γ operator, discussed in §9.2.1, coming from `cs.nyu.edu/pub/wave`.

11   Contours extraction

This section describes the pipeline of contours extraction, based on the gradient approach, i.e. only one differentiation. In what follows, we briefly introduce the Canny [Can86] and Deriche [Der87] operators, we then discuss the different steps (extraction of the local maxima and hysteresis thresholding) necessary to obtain the contours from the output of an operator.

11.1   Optimal edge detectors

11.1.1   A simple edge model

The basic idea behind optimal operators is based on a continuous edge model of the form
$$I(x) = A u_{-1}(x) + B(x)$$

Program 10.1 Implementation of the *algorithme à trous*.

```
void atrous(float *next_a,
    float *next_d,float *a,int s,int n,
    float *h,int h_begin,int h_end,
    float *g,int g_begin,int g_end)
{
    int i,j;
    for(i=0;i<n;i++)
    {
        next_a[i]=.0;
        for(j=h_begin;j<=h_end;j++)
        {
            int ii=i+(j<<s);
            ii=(ii<0?(n+ii)%n:(ii>=n?(ii%n):ii));
            next_a[i]+=a[ii]*h[j-h_begin];
        }
        next_d[i]=.0;
        for(j=g_begin;j<=g_end;j++)
        {
            int ii=i+(j<<s);
            ii=(ii<0?(n+ii)%n:(ii>=n?(ii%n):ii));
            next_d[i]+=a[ii]*g[j-g_begin];
        }
    }
}
```

Program 10.2 Implementation of the inverse *algorithme à trous*.

```
void iatrous1d(float *next_a,
    float *next_d,float *a,int s,int n,
    float *h,int h_begin,int h_end,
    float *g,int g_begin,int g_end)
{
    int i,j;
    for(i=0;i<n;i++)
    {
        next_a[i]=.0;
        for(j=h_begin;j<=h_end;j++)
        {
            int ii=i+(j<<s);
            ii=(ii<0?(n+ii)%n:(ii>=n?(ii%n):ii));
            next_a[i]+=a[ii]*h[j-h_begin];
        }
        next_d[i]=.0;
        for(j=g_begin;j<=g_end;j++)
        {
            int ii=i+(j<<s);
            ii=(ii<0?(n+ii)%n:(ii>=n?(ii%n):ii));
            next_d[i]+=a[ii]*g[j-g_begin];
        }
    }
}
```

Program 10.3 Implementation of the fast wavelet transform.

```c
int fwt1d(float *sig,int n,
    float *h,int h_begin,int h_end,
    float *g,int g_begin,int g_end)
{
    float *tmp;
    int i,j,min_l=MAX(min_length(h_begin,h_end),
                      min_length(g_begin,g_end));
    if(tmp=(float*)malloc(n*sizeof(float)),!tmp)
    {
        fprintf(stderr,"\nmalloc() error\n");
        exit(1);
    }
    while(n!=min_l)
    {
        for(i=0;i<n;i+=2)
        {
            tmp[i>>1]=.0;
            tmp[(i>>1)+(n>>1)]=.0;
            for(j=h_begin;j<=h_end;j++)
            {
                register int k=i+j;
                tmp[i>>1]+=h[j-h_begin]*(k<0?sig[n+k]:
                    (k>=n?sig[k%n]:sig[k]));
            }
            for(j=g_begin;j<=g_end;j++)
            {
                register int k=i+j;
                tmp[(i>>1)+(n>>1)]+=g[j-
g_begin]*(k<0?sig[n+k]:
                    (k>=n?sig[k%n]:sig[k]));
            }
        }
        memcpy(sig,tmp,n*sizeof(float));
        n>>=1;
    }
    free(tmp);
    return(min_l);
}
```

Program 10.4 Implementation of the fast inverse wavelet transform.

```
void ifwt1d(float *sig,int n,
    float *h,int h_begin,int h_end,
    float *g,int g_begin,int g_end)
{
    float *tmp;
    int i,j,l=MAX(min_length(h_begin,h_end),
                  min_length(g_begin,g_end));
    if(tmp=(float*)malloc(n*sizeof(float)),!tmp)
    {
        fprintf(stderr,"\nmalloc() error\n");
        exit(1);
    }
    while(l!=n)
    {
        for(i=0;i<l<<1;i++)
        {
            tmp[i]=.0;
            for(j=-h_end;j<=-h_begin;j++)
                if( !((i+j)&0x1) )
                {
                    register int k=(i+j)/2;
                    tmp[i]+=h[-j-h_begin]*(k<0?sig[l+k]
                        :(k>=l?sig[k%l]:sig[k]));
                }
            for(j=-g_end;j<=-g_begin;j++)
                if( !((i+j)&0x1) )
                {
                    register int k=(i+j)/2;
                    tmp[i]+=g[-j-g_begin]*(k<0?sig[l+k+l]:
                        (k>=l?sig[k%l+l]:sig[k+l]));
                }
        }
        l<<=1;
        memcpy(sig,tmp,l*sizeof(float));
    }
    free(tmp);
}
```

Program 10.5 Implementation of P_Γ.

```
interp(float u0, float un, int n,
    float *u, float r1)
{
    double r_1, r_2, r2,
        rn_1, r2n_2, r2n,
        r2n_2i, rn_i, ri, r2i, a0, an;
    int i;
    rn_1=r1;
    for(i=1;i < n-1;++i)
        rn_1*=r1;
    r2n_2=rn_1*rn_1;
    r_1=1./r1;
    r_2=r_1*r_1;
    r2=r1*r1;
    r2n=r2n_2*r2;
    a0=u0/(1.-r2n);
    an=un/(1.-r2n);
    u[0]=u0;
    r2n_2i=r2n_2;
    r2i=r2;
    ri=r1;
    rn_i=rn_1;
    for(i=1;i<n;i++)
    {
        u[i]=a0*ri*(1-r2n_2i)+an*rn_i*(1-r2i) ;
        r2n_2i*=r_2;
        r2i*=r2;
        ri*=r1;
        rn_i*=r_1;
    }
    u[n]=un;
}
```

where $u_{-1}(x)$ denotes the unit-step function and $B(x)$ is a centered gaussian white noise of variance equal to σ^2. We then consider the convolution $\Theta(x)$ with an edge detector $f(x)$

$$\Theta(x) = \int_{-\infty}^{+\infty} I(y)f(x-y)\,dy$$

11.1.2 The Canny criteria

According to this model, John Canny [Can86] has proposed to optimize the three following requirements, in order to find the form of the detector f.

Low probability of error (failing to mark of falsely marking real edge points). This criterion consists in finding an asymetric operator which maximises the signal-to-noise ratio, i.e.

$$\Sigma = \frac{\int_{-\infty}^{0} f(x)\,dx}{\sigma\sqrt{\int_{-\infty}^{+\infty} f^2(x)\,dx}}$$

Good localization. points marked as edges should be as closed as possible to the true edge. This criterion is defined as being the inverse of the standard deviation of the position of the true edge, i.e.

$$\Lambda = \frac{A|f'(0)|}{\sqrt{\int_{-\infty}^{+\infty} f'^2(x)\,dx}}$$

Only one response to a single edge. consists in a constraint on the average distance between two maxima (x_{max}), i.e.

$$x_{max} = \sqrt{\frac{\int_{-\infty}^{+\infty} f'^2(x)\,dx}{\int_{-\infty}^{+\infty} f''^2(x)\,dx}}$$

John Canny has then proposed a FIR operator which optimizes the product $\Sigma\Lambda$ under the constraint that the third criterion is fixed to a constant value k. In practice, the Canny operator is approximated by the first derivative of a Gauss function which leads to $\Sigma\Lambda = .92$ ($k = .51$).

59

11.1.3 The Deriche operator

Rachid Deriche [Der87] has derived an IIR operator that optimizes $\Sigma\Lambda$ and leads to $\Sigma\Lambda = 2$ ($k = .44$). The operator has the form[21]

$$f(x) = Sxe^{-\alpha|x|}$$

In one dimension, this operator can be implemented using two stable second order reccursive filters (an implementation using the C-language is given in program 11.1). An interesting property of the operator comes from the α parameter which allows to adapt it to the content of the image. Roughly, for a noisy image, α has to be small (.25 to .5) which means that Σ (detection) is favoured to the detriment of Λ (localization), one the other hand, for a "clean" image, α must be relatively large (≈ 1). In two dimensions, the output of the operator is computed via two sets of two reccursive filters applied separately on the rows and the columns of the image (this operation must be performed twice—with different parameters—for obtaining the partial derivatives according to x and y). More details (derivation, implementation, . . .) are notably available in [Der, CP95]. Figure 12 shows the output of the Deriche operator on the "singe" image.

11.2 Local maxima and hysteresis thresholding

11.2.1 Extraction of the local maxima

Given some estimations of the partial derivatives according to x and y, one can compute the norm and the direction of the gradient, i.e.

$$|\vec{\nabla}I| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}$$

and

$$\phi = \tan^{-1}\frac{\partial I/\partial y}{\partial I/\partial x} \tag{38}$$

and use this information in order to extract the local maxima of the gradient norm in the gradient direction. This is necessary for obtaining thin coutours, i.e. contours whose thickness is equal to one pixel. However, the coordinates given by the gradient direction do not coincide (in general) with integer pixel coordinates: a bilinear interpolation scheme should be applied is order to get a value at this location. A given point is then marked as a local maxima if its value is greater than those of its two neighbours in the gradient direction. See figure 13.

---

[21]It is the limit of $f(x) = \frac{S}{\omega}e^{-\alpha|x|}\sin\omega x$ when $\omega \to 0$. This case corresponds to the largest value of $\Sigma\Lambda$.

Program 11.1 Implementation of the one-dimensional Deriche operator.

```
#define SQR(x) ((x)*(x))
#define ABS(x) (x<0?-1:(x>0?1:0))
void deriche1d(float *x,int n,float a)
{
    int i;
    float *y1,*y2,b,c,k;
    if(y1=(float*)malloc(n*sizeof(float)),
       y2=(float*)malloc(n*sizeof(float)),
       !y1 || !y2)
    {
        fprintf(stderr,"\nmalloc() error");
        exit(1);
    }
    b=(float)exp((double)(-a));
    c=(float)exp((double)(-2*a));
    k=SQR(1-b)/(1.+2.*a*b-c);
    for(i=0;i<n;i++)
    {
        int ii=n-i-1;
        y1[i]=(i-1<0?.0:x[i-1])+
              2*b*(i-1<0?.0:y1[i-1])-
              c*(i-2<0?.0:y1[i-2]);
        y2[ii]=(ii+1>=n?.0:x[ii+1])+
               2*b*(ii+1>=n?.0:y2[ii+1])-
               c*(ii+2>=n?.0:y2[ii+2]);
    }
    for(i=0;i<n;i++)
        x[i]=ABS(k*b*(y1[i]-y2[i]));
    free(y1);
    free(y2);
}
```

Figure 12 Output of the Deriche operator.
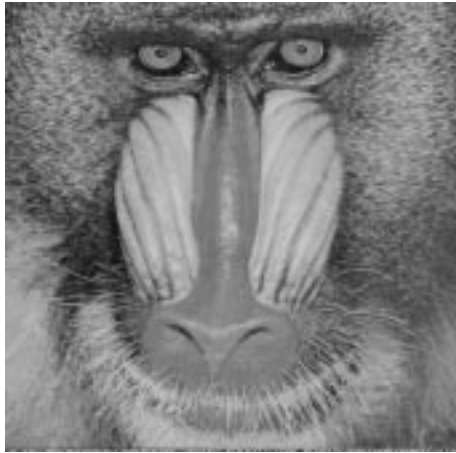
(a) Original image.

(b) $\frac{\partial}{\partial x} I \otimes f$.

(c) $\frac{\partial}{\partial y} I \otimes f$.

(d) Gradient norm.

11.2.2 Hysteresis thresholding

The last step consists in eliminating the non-significant local maxima, i.e. those which correspond to small values of the gradient norm. Most of the time, this operation is performed via a thresholding procedure. A smart way of doing it is called a hysteresis thresholding: given a high and a low level, the hysteresis thresholding keeps the local maxima with corresponding gradient norm greater than the high level or greater than the low one and connected (in a 8-connexity sense) to a local maxima whose norm is greater than the high level. The 8-connexity is defined as

Definition 8 (8-connexity) *A given point* A *is said to be connected to a given point* B *if one of its eight neighbours is the point* B *or if one of its eight neighbours is connected to the point* B.

This thresholding technique gives better results than a simple one-level thresholding operator, because it is able to extract some contour points below the noise level. Figure 13 shows the contours obtained after an hysteresis thresholding.

Figure 13 Contours extraction from figure 12.
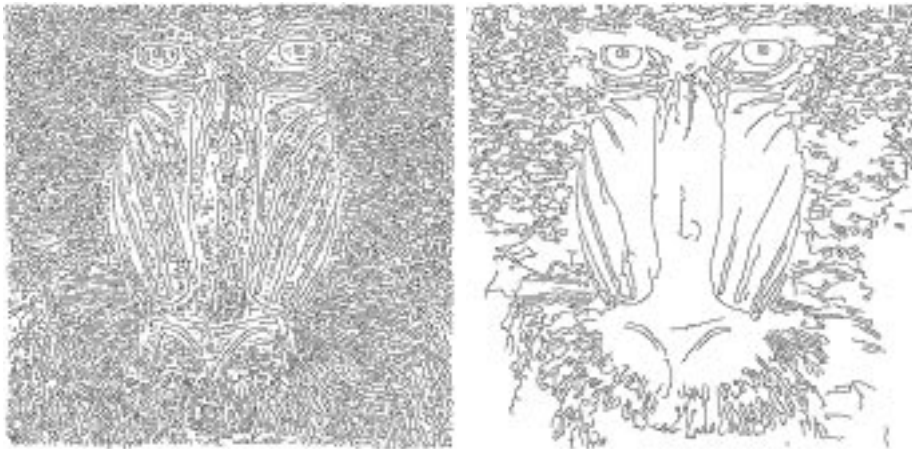
(a) Local maxima.　　　　　(b) Hysteresis tresh.

12 Mathematical complement

12.1 Hilbert space and Riesz basis

Definition 9 (Abstract Hilbert space) *A set of abstract elements which possesses the three following properties is said to be a Hilbert space $\mathcal{H}$,*

1. *$\mathcal{H}$ is a linear space i.e. the operations of addition and scalar multiplication are defined for its elements, in particular there exists an element 0 such that $0 = 0.f$ for all element of $\mathcal{H}$.*

2. *$\mathcal{H}$ is a metric space whose metric is derived from a scalar product, denoted by $< f, g >$, so that $< af, g >= a < f, g >$ for all scalar a, $< f + g, h >=< f, h > + < g, h >$, $< f, g >=< g, f >^*$, $< f, f >> 0$ for $f \neq 0$ and $< f, f >= 0$ for $f = 0$. The norm of an element f is then defined by $\|f\| =< f, f >^{\frac{1}{2}}$.*

3. *$\mathcal{H}$ is a complete space i.e. if a sequence of elements $\{f_n\}$ satisfies*

$$\lim_{n,m \to \infty} \|f_n - f_m\| = 0$$

then there exists an element $\tilde{f}$ such that $\lim_{n,m \to \infty} \|f_n - \tilde{f}\| = 0$.

The properties of these spaces are notably studied in [RSN55, KF61].

Definition 10 (Riesz basis) *A family of element $\{e_k\}$ of an abstract Hilbert space $\mathcal{H}$ is said to be a Riesz basis of $\mathcal{H}$ if the following properties are (simultaneously) satisfied for all x in $\mathcal{H}$*

1. *$\exists \{\lambda_k\}/x = \sum_k \lambda_k e_k$.*

2. *$\exists A, B \in \mathbb{R}^{2+}/\frac{1}{A}\|x\|^2 \leq \sum_k |\lambda_k|^2 \leq \frac{1}{B}\|x\|^2$.*

12.2 Hölder (Lipschitz) regularity

12.2.1 Definition

This subsection aims at defining the concept of Hölder exponant. We use the presentation of [Mal98]. The idea is to start from the taylor series of f at v (f is assumed to be N times differentiable on $[v - h, v + h]$)

$$p_v(x) = \sum_{k=0}^{N-1} \frac{f^{(k)}(v)}{k!}(x - v)^k + R_N$$

As shown in [Weiar]

$$R_N = \varepsilon(x) = f(x) - p_\nu(x) = \frac{(x-\nu)^N}{N!} f^{(N)}(x)$$

and hence

$$\forall x \in [\nu - h, \nu + h], \ |\varepsilon(x)| \le \frac{|x-\nu|^N}{N!} \sup_{y \in [\nu-h,\nu+h]} |f^{(N)}(y)|$$

The notion of Hölder regularity generalizes the previous inequality to non integer exponents.

Definition 11 (Hölder regularity) *A function is pointwise Hölder-α at ν if there exists $K > 0$ and a polynomial p_ν of degree $N = \lfloor \alpha \rfloor$[22] such that*

$$\forall x \in \mathbb{R}, \ |f(x) - p_\nu(x)| \le K|x - \nu|^\alpha \tag{39}$$

A function is uniformly Hölder-α over an interval $[a, b]$ if it satisfies the previous equation $\forall \nu \in [a, b]$ with a constant K that does not depend on ν. The Hölder regularity of f at ν over $[a, b]$ is the sup of the α such that f is Hölder-α.

12.2.2   A few remarks

If f is uniformly Hölder-α ($\alpha > N$) in the neighbourhood of ν then f is N times continuously differentiable in the neighbourhood of ν. If $0 \le \alpha < 1$ then $p_\nu(x) = f(\nu)$ and equation (39) becomes

$$\forall x \in \mathbb{R}, \ |f(x) - f(\nu)| \le K|x - \nu|^\alpha$$

If $\alpha < 1$, f is not differentiable in the neighbourhood of ν and the Hölder exponent characterizes the type of singularity. For example [Dau92, MZ92] Heaviside-like singularities are Hölder-0 while Dirac-like ones are Hölder-(-1). Note that the uniform Hölder regularity of f over $\mathbb{R}$ is related to a condition on the decay of its Fourier transform via the following theorem (proofs in [Mal98, Dau92]).

Theorem 10 *A function f is bounded and uniformly Hölder-α over $\mathbb{R}$ if*

$$\int_{-\infty}^{+\infty} |\hat{f}(\xi)|(1 + |\xi|^\alpha) d\xi < \infty$$

---

[22]$\lfloor \alpha \rfloor$ denotes the largest integer such that $N \le \alpha$.

12.3 Spaces: W^α, $B^\alpha_{p,q}$, C^α

The purpose of this section is obviously not to provide a deep analysis of the notions of Sobolev (W^α), Besov ($B^\alpha_{p,q}$) and Hölder (C^α) spaces (the notations are taken from [PB, Mey90]) spaces and their relationships with orthogonal wavelet decompositions (for that purpose the reader is directly sent to Yves Meyer's book [Mey90]). Our goal is just to (very) briefly introduce this subject.

12.3.1 Short presentation

Besov spaces are subsets of $L^p(\mathbb{R})$. They are extensions of Sobolev and Hölder spaces in which the smoothness of a given function is finer characterized. Basically, the fact that a function lies in W^α or C^α gives an idea of its global smoothness, while its membership of $B^\alpha_{p,q}$ gives some information about its local smoothness, e.g. piecewise regular functions belong to Besov spaces [Mal98].

The classical definition of Besov spaces in based on the modulus of smoothness [Zyg68]

$$\omega_p(f;h) = \|\tau_{-h}f - f\|_{L^p(\mathbb{R})}$$

and on the two following semi-norms [Del93, PB, DL92] ($1 \leq p, q < \infty$)

- $0 < \alpha < 1$:

$$N^\alpha_{p,q}(f) = \left(\int_0^{+\infty} \left(\frac{\omega_p(f;h)}{h^\alpha} \right)^q \frac{dh}{h} \right)^{\frac{1}{q}}$$

- $\alpha = 1$:

$$N^1_{p,q}(f) = \left(\int_0^{+\infty} \left(\frac{\omega^\star_p(f;h)}{h} \right)^q \frac{dh}{h} \right)^{\frac{1}{q}}$$

where $\omega^\star_p(f;h) = \|\tau_{-h}f - 2f + \tau_h f\|_{L^p(\mathbb{R})}$. If $q = \infty$, $N^\alpha_{p,\infty}(f) = \sup_{\mathbb{R}^{+*}} \omega_p(f;h)/h^\alpha$ with the modification for $\alpha = 1$. We then need (again for $0 < \alpha \leq 1$)

$$\|f\|_{B^\alpha_{p,b}} = \|f\|_{L^p(\mathbb{R})} + N^\alpha_{p,q}(f)$$

for completing the definition of a Besov space.

Definition 12 (Besov space) *A function* $f \in L^p(\mathbb{R})$ *belongs to the Besov space* $B^\alpha_{p,q}$ *if*

- $0 < \alpha \leq 1$: $\|f\|_{B^\alpha_{p,q}} < \infty$.

- $\alpha > 1$: $\|f^{(k)}\|_{B^{\alpha - \lfloor \alpha \rfloor}_{p,q}} < \infty$, $0 \leq k \leq \lfloor \alpha \rfloor$[23]

---

[23]The associated norm (for $\alpha > 1$) becomes $\|f\|_{B^\alpha_{p,q}} = \sum_{k=0}^{\lfloor \alpha \rfloor} \|f^{(k)}\|_{B^{\alpha - \lfloor \alpha \rfloor}_{p,q}}$.

Provided $\|.\|_{B^\alpha_{p,q}}$, a Besov space has a Banach space[24] structure [PB]. Note that $B^\alpha_{\infty,\infty} = C^\alpha$ ($\{f \in L^\infty(\mathbb{R}) / \sup_{\mathbb{R}^{*+}} \omega_\infty(f;h)/h^\alpha < \infty\}$ [Dau92]) and that $B^\alpha_{2,2} = W^\alpha$ ($\{f \in L^2(\mathbb{R}) / \|f^{(\alpha)}\|_{L^2(\mathbb{R})} = \frac{1}{2\pi}\|(i\xi)^\alpha\hat{f}\|_{L^2(\mathbb{R})} < \infty\}$, $f^{(\alpha)}$ denotes the weak or Sobolev deritative of f, α is not necessarily an integer [Mal98]) (notably) [PB, Mey90, Tri78].

Orthogonal wavelet basis have interesting properties for analysing these classes of function. As demonstrated by Yves Meyer [Mey90], the norm $\|.\|_{B^\alpha_{p,q}}$ is equivalent to a norm on the wavelet coefficients if the multiresolution analysis generated by the scaling function/wavelet pair is r-regularly (in Meyer's sense) with $r \geq \alpha$ i.e. [Mey90, PB, Del93, Don91]

$$\|f\|_{B^\alpha_{p,q}} \asymp \|\mu_0\|_{l^p(\mathbb{Z})} + \left(\sum_{j=0}^{+\infty} 2^{jq(\alpha+\frac{1}{2}-\frac{1}{p})}\|\gamma_j\|^q_{l^p(\mathbb{Z})}\right)^{\frac{1}{q}} \tag{40}$$

recall that $\mu_{0;k} = <f, \phi_{0;k}>$ and that $\gamma_{j;k} = <f, \psi_{j;k}>$, the symbol $\asymp$ means that there exists two constants A and B such that the ratio of the two sides is bounded bertween them. It is therefore easier to determine if a given function $f \in L^p(\mathbb{R})$ belongs to $B^\alpha_{p,q}$.

A very interesting result comes from the fact that an orthogonal wavelet basis (obeying the regularity condition) provides an unconditionnal basis[25] of $B^\alpha_{p,q}$, see (again!) [Mey90]. This implies that orthogonal wavelet basis are "optimal" (in some sense) for analysing and processing the functions belonging to $B^\alpha_{p,q}$ e.g. simple (thresholding) operators, applied in the unconditionnal basis, work better for a whole class of problems (namely: compression, estimation and recovery) than they do in any other orthogonal basis (the mathematical details are available in [Don91]). This is (roughly) a consequence of the fact that a function is charactized by a few "relevant" coefficients in the unconditionnal basis. For more details on these functionnal spaces (other definitions, extensions to n dimensions, other properties, . . .) the reader is sent to [DP88, FJ85, Tri78] and to almost every books about wavelet analysis since this theory uses them for an increasing number of applications (most of these works—known to the author—have already been cited in this subsection).

---

[24]Banach spaces generalize the notion of Hilbert space (definition 9): the norm is not necessarily defined from a scalar product [RSN55].

[25]As defined in [Dau92]: a family of elements $\{e_k\}$ of a Banach space B is a Schauder basis of B if $\forall f \in B, \exists\{\mu_k\}$ (unique) $/f = \lim_{N\to\infty}\sum_{k=1}^N \mu_k e_k$. Moreover, if $\sum_k \mu_k e_k \in B \Rightarrow \sum_k |\mu_k|e_k \in B$ the family $\{e_k\}$ is said to be an unconditionnal basis of B. On a Hilbert space, an unconditionnal basis is a Riesz basis.

12.3.2 Example: *l'algèbre des bosses*

In order to give a more "intuitive" idea of the kind of functions that belong to Besov spaces, this subsection is devoted to a (short) presentation of $B_{1,1}^1$ also known as *l'algèbre des bosses gaussiennes* ("bump algebra" [DJ91]) introduced by Yves Meyer [Mey90]. In what follows, $g_{\mu;\sigma}(x)$ denotes the Gauss function

$$e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

such that $g_{\mu;\sigma}(\mu) = 1$ instead of the usual normalization (area equal to 1). *L'algèbre des bosses* (B) is defined as the class of functions (vanishing at infinity) which admit a (non-unique) decomposition of the form

$$f(x) = \sum_{k=0}^{+\infty} \lambda_k g_{\mu_k;\sigma_k}(x) \tag{41}$$

satisfying $\sum_k |\lambda_k| < \infty$. Provided the norm $\|f\|_B = \inf \sum_k |\lambda_i|$, such that $\{\lambda_k\}_{k\in\mathbb{N}}$ satisfies equation (41), B is a Banach space. In an orthogonal wavelet basis (generated by a sufficiently regular multiresolution analysis), the decomposition of a function f belonging to B must statisfy

$$\sum_{j=-\infty}^{+\infty} 2^{\frac{j}{2}} \|\gamma_j\|_{l^1(\mathbb{Z})} \tag{42}$$

and vice versa (proof in [Mey90]). Note that equation (42) corresponds to equation (40) with $p = q = \alpha = 1$: this illustrates the fact that the "bump algebra" is $B_{1,1}^1$. The class B contains some functions which may have considerable spatial hinomogeneity e.g. a function $f \in B$ can be extremely spiky in one part of its domain and completly flat in another location. This type of behavior could not be possible in a Hölder or Sobolev space since it is required that a function is "equally" smooth at every points on its domain [DJ91].

List of Figures