Wavelets and their Applications

Edited by
Michel Misiti
Yves Misiti
Georges Oppenheim
Jean-Michel Poggi

iSTE

Wavelets and their Applications

Part of this book adapted from “Les ondelettes et leurs applications” published in France by Hermes Science/Lavoisier in 2003
First published in Great Britain and the United States in 2007 by ISTE Ltd

ISTE Ltd
6 Fitzroy Square
London W1T 5DX
UK

ISTE USA
4308 Patrice Road
Newport Beach, CA 92663
USA

www.iste.co.uk

Library of Congress Cataloging-in-Publication Data

Wavelets and their applications/edited by Michel Misiti ... [et al.].
p. cm.
"Part of this book adapted from "Les ondelettes et leurs applications"
published in France by Hermes Science/Lavoisier in 2003.
ISBN-13: 978-1-905209-31-6
ISBN-10: 1-905209-31-2
1. Wavelets (Mathematics) I. Misiti, Michel.
QA403.3.W3625 2006
515'.2433--dc22

2006032725

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-905209-31-6

Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire.

Table of Contents

Notations

$\hat{f}$	Fourier translation of function f: $\hat{f}(\omega) = \int_{\mathbb{R}} f(t) e^{-2i\pi\omega t}\, dt$
$\mathbf{F}$	Fourier transform
$\bar{\mathbf{F}}$	Inverse Fourier transform
$L^1(\mathbb{R})$	Space integrable function on $\mathbb{R}$
$L^2(\mathbb{R})$	Space of square integrable functions over $\mathbb{R}$ or set of finite energy signals over $\mathbb{R}$ $\|f\|_{L^2}^2 = \int_{\mathbb{R}} \|f(t)\|^2\, dt < +\infty$
$(\ ,\ )$	Scalar product in space $L^2(\mathbb{R})$: $(f,g) = \int_{\mathbb{R}} f(t)\,\overline{g(t)}\, dt$
$\|\ \|_{L^2}$	Norm in space $L^2(\mathbb{R})$
$l^2(\mathbb{Z})$	Space of square-summable sequences indexed by $\mathbb{Z}$
$C^0(\mathbb{R})$	Set of continuous functions over $\mathbb{R}$
$C^n(\mathbb{R})$	Set of n times continuously differentiable functions over $\mathbb{R}$
$S(\mathbb{R})$	Set of rapidly decreasing indefinitely differentiable functions over $\mathbb{R}$
$S'(\mathbb{R})$	Dual space of $S(\mathbb{R})$; set of tempered distributions over $\mathbb{R}$

C_n^k	Binomial coefficients for $0 \leq k \leq n$
$\overline{V}$	Closure of set V (the smaller closed subspace containing V)
$V \oplus W$	Direct sum of spaces V and W; set of elements of the form $v + w$ with $v \in V$ and $w \in W$
$V \perp W$	Spaces V and W are orthogonal
$\overline{z}$	Complex conjugate number of $z \in \mathbb{C}$
ε	White noise
δ	Dirac distribution (at point 0)
V_j	Space of approximations at level *j*
W_j	Space of details at level *j*
A_j or A^j	Approximation at level *j* (reconstructed in V_0)
D_j or D^j	Detail at level *j* (reconstructed in V_0)
ψ	Wavelet associated with a multiresolution analysis (discrete analysis) or satisfying the admissibility condition (continuous analysis)
$\psi_{a,b}$	Wavelet family associated with ψ: $\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad a \in \mathbb{R}^+, b \in \mathbb{R}$
$C_f(a,b)$	Coefficient of function f on wavelet function $\psi_{a,b}$
$\psi_{j,k}$	$\psi_{j,k}(x) = 2^{-j/2}\psi\left(2^{-j}x - k\right), \quad$ for $(j,k) \in \mathbb{Z}^2$ Dyadically dilated-translated versions of the wavelet
φ	Scaling function associated with a multiresolution analysis
$\varphi_{j,k}$	$\varphi_{j,k}(x) = 2^{-j/2}\varphi\left(2^{-j}x - k\right), \quad$ for $(j,k) \in \mathbb{Z}^2$ Dyadically dilated-translated versions of the scaling function
LoD	Low-pass decomposition filter
HiD	High-pass decomposition filter

LoR	Low-pass reconstruction filter
HiR	High-pass reconstruction filter
$\downarrow 2$ or *dec*	Down-sampling operator: $Y = dec(X) = [\downarrow 2]X$ with $Y_n = X_{2n}$
$\uparrow 2$ or *ins*	Up-sampling operator: $Y = ins(X) = [\uparrow 2]X$ avec $Y_{2n} = X_n$ and $Y_{2n+1} = 0$
$*$	Convolution operator: $Y = X * F$ avec $Y_n = \sum_{k\in\mathbb{Z}} X_{n-k}F_k$
$\delta^{(p)}$	Sequence defined by $\delta_k^{(p)} = 0$ if $k \neq p$ and $\delta_p^{(p)} = 1$ for $p,k \in \mathbb{Z}$
T_p	Translation operator for sequences $b = T_p(a)$ is the sequence defined by $b_k = a_{k-p}$ for $p,k \in \mathbb{Z}$

Introduction

Wavelets

Wavelets are a recently developed signal processing tool enabling the analysis on several timescales of the local properties of complex signals that can present non-stationary zones. They lead to a huge number of applications in various fields, such as, for example, geophysics, astrophysics, telecommunications, imagery and video coding. They are the foundation for new techniques of signal analysis and synthesis and find beautiful applications to general problems such as compression and denoising.

The propagation of wavelets in the scientific community, academic as well as industrial, is surprising. First of all, it is linked to their capacity to constitute a tool adapted to a very broad spectrum of theoretical as well as practical questions. Let us try to make an analogy: the emergence of wavelets could become as important as that of Fourier analysis. A second element has to be noted: wavelets have benefited from an undoubtedly unprecedented trend in the history of applied mathematics. Indeed, very soon after the grounds of the mathematical theory had been laid in the middle of the 1980s [MEY 90], the fast algorithm and the connection with signal processing [MAL 89] appeared at the same time as Daubechies orthogonal wavelets [DAU 88]. This body of knowledge, diffused through the Internet and relayed by the dynamism of the research community enabled a fast development in numerous applied mathematics domains, but also in vast fields of application.

Thus, in less than 20 years, wavelets have essentially been imposed as a fruitful mathematical theory and a tool for signal and image processing. They now therefore form part of the curriculum of many pure and applied mathematics courses, in universities as well as in engineering schools.

By omitting the purely mathematical contributions and focusing on applications, we may identify three general problems for which wavelets have proven very powerful.

The first problem is analysis, for scrutinizing data and sounding out the local signal regularity on a fine scale. Indeed, a wavelet is a function oscillating as a wave but quickly damped. Being well localized simultaneously in time and frequency it makes it possible to define a family of analyzing functions by translation in time and dilation in scale. Wavelets constitute a mathematical "zoom" making it possible to simultaneously describe the properties of a signal on several timescales.

The second problem is denoising or estimation of functions. This means recovering the useful signal while we only observe a noisy version thereof. Since the denoising methods are based on representations by wavelets, they create very simple algorithms that, due to their adaptability, are often more powerful and easy to tune than the traditional methods of functional estimation. The principle consists of calculating the wavelet transform of observations, then astutely modifying the coefficients profiting from their local nature and, finally, inversing the transformation.

The third problem is compression and, in particular, compression of images where wavelets constitute a very competitive method. Due to generally very sparse representations, they make it possible to reduce the volume of information to be coded. In order to illustrate this point, we can consider two leading applications whose impact has propagated well beyond the specialists in the field. The first application relates to the storage of millions of fingerprints by the FBI and the second is linked to the new standard of image compression JPEG 2000, which is based on wavelets.

Wavelets provide particularly elegant solutions to a number of other problems. Let us quote, for example, the numerical solution of partial derivative equations or even, more to the point, the simulation of paths for fractional Brownian processes. Numerous types of software have appeared since the beginning of the 1990s and, particularly over the last few years, a complete list can be found on the website www.amara.com/current/wavesoft.html.

Why this book?

Our aim is to be somewhere in the space that separates the foundations and the computerized implementation. Operating with wavelets means understanding the origins of the tool and, at the same time, its application to signals. We attempt to

show the link between knowing and acting by introducing a very large number of images illustrating the text (perhaps 200, often simple ones).

Who is this book for?

Who are the ideal readers? This is always difficult to say. From the most theoretical point of view, a specialist in mathematical analysis is likely not to find his due here. Other very good texts already exist. As for the applications, the programming of algorithms is not included in this book; a lot of software is available. We position ourselves somewhere between these two extremes: theoretical mathematics and applied algorithms.

This work, intended for a large audience of scientists, is directed towards learning and understanding wavelets through their applications. It can be useful by complementing the works strictly dedicated to mathematical approaches for students of engineering schools, those undertaking graduate and postgraduate research, as well as to engineers and researchers eager to have a compact yet broad view of wavelets in action.

What can be found in this work?

The organization of this work is reflected in the table below. The last column indicates the principal orientation of each chapter and shows the place occupied by the applications.

Chapter	Orientation
1	Applications
2	Theory
3	Theory and algorithms
4	Theory
5	Theory and applications
6	Applications
7	Theory and applications
8	Applications
9	Applications

Let us summarize the contents of the book in a few sentences by grouping the naturally associated chapters.

Chapter 1 proposes a quick overview of what could be called wavelets in action, of three key problems: analysis, denoising and compression.

Chapters 2 and 3 provide the mathematical and algorithmic framework for wavelets transform.

Chapters 4 and 5 concern wavelets families. The former proposes a genealogy of the well-known families, whereas the latter explains how to design a new wavelet.

Chapters 6, 7 and 8 make it possible to "handle the tool" through broad application topics. Chapter 6 examines the analysis for 1D signals, while denoising and compression are covered in Chapter 7. For images, the three topics are covered in Chapter 8.

Chapter 9 presents the analysis of real applications placed in their academic and industrial context, illustrating the variety of the fields concerned and the problems covered.

The following table enables the identification of the logical dependence of chapters.

Chapter	**Required reading**
1	
2	
3	2
4	(2, 3) or 1
5	2, 3
6	2 or 1
7	(2, 3) or 1
8	(2, 3, 6) or 1
9	1 or (2, 3, 6, 7, 8)

The examples and graphs presented throughout the book have been obtained due to the MATLAB© Wavelet Toolbox[1], which is a software for signal or image processing by wavelets and packages of wavelets, developed by the authors [MIS 00].

Is our presentation of the development of wavelets timely? Despite our best efforts, the rate of appearance of applications tests renders any static assessment null and void. Between 1975 and 1990 mathematical results have appeared in large numbers, but today they are much fewer. Applications have evolved following a complementary movement and the quantity of applied work has currently become very large. Moreover, we can imagine that a considerable number of industrial applications have not been made public.

Bibliographical references

Let us quote some references offering complete introductions to the domain of wavelets from the mathematical point of view as well as from the point of view of signal processing.

Let us start first of all with books strictly dedicated to mathematical processing. The book by Daubechies [DAU 92] remains from this point of view, a reference text, along with those of Meyer [MEY 90] and of Kaiser [KAI 94]. These may be supplemented by two books [FRA 99] and [WAL 02], of greater accessibility and with a certain teaching potential. Still in the same spirit, but concentrating on more specific questions, we may consult [COH 92a] on bi-orthogonal wavelets and, for continuous analysis, we may refer to [TEO 98] and [TOR 95].

The book by Mallat [MAL 98] is, without a doubt, one of the most comprehensive that is currently available and constitutes an invaluable source, in particular for those looking for a presentation harmoniously uniting mathematics and signal.

In the book by Strang and Nguyen [STR 96] an original vision deliberately directed towards signal processing is adopted. Moreover, it can also be referred to for concepts and definitions traditional in this field and not presented in this book.

A compact presentation accessible to a large audience can be found in the book of scientific popularization [MEY 93]. Lastly, in the book by Burke Hubbard [BUR 95] a very enthralling text on the history of the wavelets can be found.

---

1 MATLAB© is a trade mark of The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA, 01760-2098, USA, info@mathworks.com.

Naturally, more specialized references deserve to be mentioned: for example, [COI 92] and [WIC 94] for wavelet packets and [ARN 95], [ABR 97] for interesting applications of the continuous transform to turbulence or signals presenting properties of self-similarity.

With respect to denoising and, more widely, the use of wavelets in statistics, some of the typical results in this field may be found in four very different books [HAR 98, OGD 97, PER 00, VID 99]. Moreover, we can also refer to [ANT 95] in order to grasp the extent of the stimulation exerted by the ideas propelled by wavelets in the statisticians' community since a few years ago. This latter point can be usefully supplemented by two more recent articles: [ANT 97] and [ANT 01].

For compression and, in particular compression of images, we may refer to [DEV 92] for general ideas, to [BRI 95] for application to fingerprints and [USE 01] and [JPE 00] for the standard JPEG 2000.

Let us mention, finally, that at the end of Chapter 9 a list of books exclusively devoted to the applications of wavelets can be found.

This book has benefited from lessons on wavelets taught by the authors at the Ecole centrale de Lyon, at the Ecole nationale supérieure des techniques avancées (ENSTA), at the mathematical engineering DESS (equivalent to MSc) at the University of Paris XI, Orsay, at the Fudan University of Shanghai and at the University of Havana.

The authors would like to express their particular gratitude to Liliane Bel, Nathalie Chèze, Bernard Kaplan and Bruno Portier for reading this work attentively and questioningly. Of course, the authors remain responsible for the remaining errors.

Chapter 1

A Guided Tour

1.1. Introduction

In this first chapter[1], we propose an overview with a short introduction to wavelets. We will focus on several applications with priority given to aspects related to statistics or signal and image processing. Wavelets are thus observed in action without preliminary knowledge. The chapter is organized as follows: apart from the introduction to wavelets, each section centers on a figure around which a comment is articulated.

First of all, the concept of wavelets and their capacity to describe the local behavior of signals at various time scales is presented. Discretizing time and scales, we then focus on orthonormal wavelet bases making it possible at the same time:

– to supplement the analysis of irregularities with those of local approximations;

– to organize wavelets by scale, from the finest to the coarsest;

– to define fast algorithms of linear complexity.

Next we treat concrete examples of real one-dimensional signals and then two-dimensional (images) to illustrate the three following topics:

– *analysis* or how to use the wavelet transform to scan the data and determine the pathways for a later stage of processing. Indeed, wavelets provide a framework for signal decomposition in the form of a sequence of signals known as approximation

1 This chapter is a translated, slightly modified version of the [MIS 98] article published in the French scientific journal, *Journal de la Société Française de Statistique*, which the authors thank for their kind authorization.

signals with decreasing resolution supplemented by a sequence of additional touches called details. A study of an electrical signal illustrates this point;

– *denoising* or estimation of functions. This involves reconstituting the signal as well as possible on the basis of the observations of a useful signal corrupted by noise. The methods based on wavelet representations yield very simple algorithms that are often more powerful and easy to work with than traditional methods of function estimation. They consist of decomposing the observed signal into wavelets and using thresholds to select the coefficients, from which a signal is synthesized. The ideas are introduced through a synthetic Doppler signal and are then applied to the electrical signal;

– *compression* and, in particular, image compressions where wavelets constitute a very competitive method. The major reason for this effectiveness stems from the ability of wavelets to generally concentrate signal energy in few significantly non-zero coefficients. Decomposition structure is then sparse and can be coded with very little information. These methods prove useful for signals (an example thereof is examined), as well as for images. The use of wavelets for images is introduced through a real image, which is then compressed. Lastly, a fingerprint is compressed using wavelet packets which generalize wavelets.

The rapid flow of these topics focuses on main ideas and merely outlines the many theoretical and practical aspects tackled. These are detailed in other chapters of this book: Chapter 2 for the mathematical framework, Chapters 5, 6 and 8 for the analysis, and Chapters 7 and 8 for denoising and compression.

1.2. Wavelets

1.2.1. *General aspects*

Let ψ be a sufficiently regular and well localized function. This function $\psi \in L^1 \cap L^2$ will be called a wavelet if it verifies the following admissibility condition in the frequency domain:

$$\int_{\mathbb{R}^+} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} d\omega = \int_{\mathbb{R}^-} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} d\omega < +\infty$$

where $\hat{\psi}$ indicates the Fourier transform of ψ. This condition involves, in particular, that the wavelet integrates to zero. This basic requirement is often

reinforced by requiring that the wavelet has m vanishing moments, i.e. verify $\int_{\mathbb{R}} t^k \psi(t)\,dt = 0$ for $k = 0,\ldots,m$.

A sufficient admissibility condition that is much simpler to verify is, for a real wavelet ψ, provided by:

$$\psi\ ,\ \psi \in L^1 \cap L^2\ ,\ t\,\psi \in L^1 \text{ and } \int_{\mathbb{R}} \psi(t)\,dt = 0$$

To consolidate the ideas let us say that during a certain time a wavelet oscillates like a wave and is then localized due to a damping. The oscillation of a wavelet is measured by the number of vanishing moments and its localization is evaluated by the interval where it takes values significantly different from zero.

From this single function ψ using translation and dilation we build a family of functions that form the basic atoms:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \qquad a \in \mathbb{R}^+, b \in \mathbb{R}$$

For a function f of finite energy we define its continuous wavelet transform by the function C_f :

$$C_f(a,b) = \int_{\mathbb{R}} f(t)\,\overline{\psi_{a,b}(t)}\,dt$$

Calculating this function C_f amounts to analyzing f with the wavelet ψ . The function f is then described by its wavelet coefficients $C_f(a,b)$, where $a \in \mathbb{R}^+$ and $b \in \mathbb{R}$. They measure the fluctuations of function f at scale a . The trend at scale a containing slower evolutions is essentially eliminated in $C_f(a,b)$. The analysis in wavelets makes a local analysis of f possible, as well as the description of scale effects comparing the $C_f(a,b)$ for various values of a . Indeed, let us suppose that ψ is zero outside of $[-M,+M]$, so $\psi_{a,b}$ is zero outside the interval $[-Ma+b, Ma+b]$. Consequently, the value of $C_f(a,b)$ depends on the values of f in a neighborhood of b with a length proportional to a .

In this respect let us note that the situation with wavelets differs from the Fourier analysis, since the value of the Fourier transform $\hat{f}(\omega)$ of f in a point ω depends on the values of f on the entire $\mathbb{R}$. Qualitatively, large values of $C_f(a,b)$ provide information on the local irregularity of f around position b and at scale a . In this

sense, wavelet analysis is an analysis of the fluctuations of f at all scales. Additional information on quantifying the concept of localization and the comparison between the Fourier and wavelet analyses may be found in [MAL 98] or in [ABR 97].

The continuous transform (see, for example, [TOR 95] or [TEO 98]) defined above makes it possible to characterize the Holderian regularity of functions and its statistical use for the detection of transient phenomena or change-points fruitful (see Chapter 6).

In many situations (and throughout this chapter) we limit ourselves to the following values of a and b:

$$a = 2^j,\ b = k2^j = ka \text{ for } (j,k) \in \mathbb{Z}^2$$

In this case and for wavelets verifying stronger properties than merely the admissibility condition – in particular, in the orthogonal case (specified below), which we shall consider from now on – a function called a scaling function and denoted φ is associated with ψ. We dilate and translate it as ψ. On the whole, the φ function is for local approximations what the ψ function is for fluctuations around the local approximation, also called the local trend.

We then define the basic atoms of wavelets which are also sometimes called wavelets:

$$\begin{cases} \psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k), & \text{for } (j,k) \in \mathbb{Z}^2 \\ \varphi_{j,k}(x) = 2^{-j/2}\varphi(2^{-j}x - k), & \text{for } (j,k) \in \mathbb{Z}^2 \end{cases}$$

In this context, the wavelet coefficients of a signal s are provided by

$$\alpha_{j,k} = \int_{\mathbb{R}} s(t)\psi_{j,k}(t)\,dt$$

and, under certain conditions (verified for orthogonal wavelets), these coefficients are enough to reconstruct the signal by:

$$s(t) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \alpha_{j,k}\, \psi_{j,k}(t)$$

The existence of a function ψ such that the family $\left\{\psi_{j,k}\right\}_{(j,k)\in\mathbb{Z}^2}$ is an orthonormal basis of $L^2(\mathbb{R})$ is closely related to the concept of multi-resolution analysis (MRA) (see [MEY 90], [MAL 89] and [MAL 98], and also Chapter 2). An MRA of the space $L^2(\mathbb{R})$ of finite energy signals is a sequence $\left\{V_j\right\}_{j\in\mathbb{Z}}$ of nested closed subspaces:

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

of $L^2(\mathbb{R})$ whose intersection is reduced to $\{0\}$ and the union is dense in $L^2(\mathbb{R})$. These spaces are all deduced from the "central" space V_0 by contraction (for $j < 0$) or dilation (for $j > 0$), i.e.:

$$f(t) \in V_j \Leftrightarrow f(2t) \in V_{j-1} \quad \text{for} \quad j \in \mathbb{Z}$$

Lastly, there is a function φ of V_0, which generates V_0 by integer translations, that is so that:

$$V_0 = \left\{ f \in L^2(\mathbb{R}) \;\middle|\; f(t) = \sum_{k\in\mathbb{Z}} e_k \varphi(t-k), \left(e_k\right) \in l^2(\mathbb{Z}) \right\}$$

where the φ function is the scaling function introduced above.

Subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of MRA are used as approximation (or trend) spaces. In addition, we also define detail spaces noted as $\{W_j\}_{j\in\mathbb{Z}}$. For fixed $j \in \mathbb{Z}$, the W_j space is the orthogonal complement of V_j in V_{j-1}:

$$V_{j-1} = V_j \oplus W_j \quad \text{for} \quad j \in \mathbb{Z}$$

An element of the approximation space of level $j-1$ is decomposed into an approximation at level j, which is less accurate, and a detail at level j.

The integer translates of φ $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ generate V_j while those of ψ $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ generate W_j. Since $L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j$, any signal is the sum of all its details and $\{\psi_{j,k}\}_{(j,k)\in\mathbb{Z}^2}$ form an orthonormal wavelet basis of $L^2(\mathbb{R})$. Thus, $\alpha_{j,k}$ is the coefficient associated with $\psi_{j,k}$ in the orthogonal projection of s onto W_j.

The respective roles of the φ and ψ functions, as well as the concepts of detail and approximation, will be illustrated below using examples. Let us now examine an "orthogonal" wavelet.

1.2.2. *A wavelet*

Figure 1.1 relates to the wavelet noted *db4* after I. Daubechies.

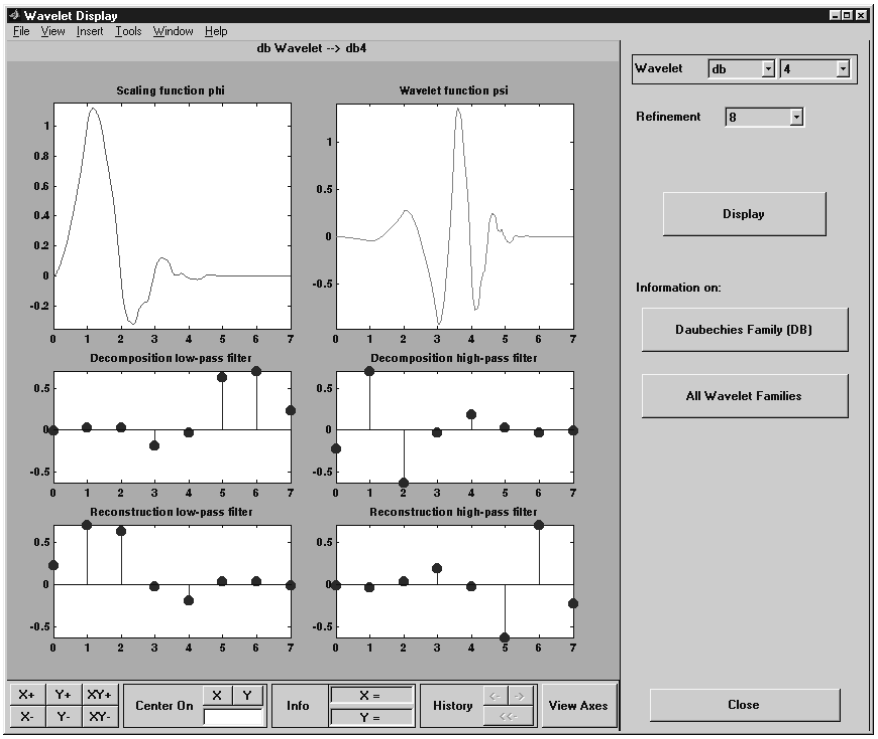


Figure 1.1. *The wavelet noted db4 after I. Daubechies*

At the top, from left to right, we find the scaling function and then the wavelet. As can be seen, the wavelet ψ oscillates and integrates to zero, while the scaling function φ oscillates less and has a positive integral (in fact, equal to 1). Consequently, calculating the scalar product of a function or a signal with a wavelet makes it possible to analyze the fluctuations of the signal around a local average provided by the calculation of the scalar product of the signal and the scaling function.

Below we provide the four filters used to carry out the calculation of coefficients by a fast algorithm due to S. Mallat (see Chapter 3). These filters go by pairs. Two are associated with the scaling function φ and they appear in the first column, while the two other filters, in the second column, are associated with the wavelet ψ .

In a column we pass from one filter to another taking the mirror filter, and in a row we pass from one filter to another taking the mirror filter and multiplying the even-indexed terms by –1. Thus, just one of these four filters is enough to produce all the others. Similarly, we can show that on the basis of a given MRA and, thus, of the scaling function, we can construct the wavelet.

The filters appear in the relations concerning the basic functions associated with successive levels. These are the equations on two following scales:

$$\varphi_{j+1,0} = \sum_{k\in\mathbb{Z}} h_k \varphi_{j,k} \quad \text{and} \quad \psi_{j+1,0} = \sum_{k\in\mathbb{Z}} g_k \varphi_{j,k}$$

The sequence *h* determines the filters of the first column and the sequence *g* determines those of the second column.

The wavelet presented in Figure 1.1 forms part of a family of *dbn* wavelets indexed by $n \in \mathbb{N}^*$ introduced by I. Daubechies in 1990 (see Chapter 4 for the construction). The wavelet *db1* is simply the Haar wavelet.

The main properties of the *dbn* wavelet are as follows:

– it is an orthogonal wavelet, associated with an MRA;

– it has compact support [0, $2n - 1$] and the associated filters are of length $2n$;

– the number of vanishing moments is n and, in general, it is far from symmetric;

– the regularity is 0, 2*n* when n is sufficiently large.

1.2.3. *Organization of wavelets*

Wavelets are thus organized using two parameters:

– time k making it possible to translate the forms for a given level;

– scale 2^j making it possible to pass from a level j to the immediately lower level in the underlying tree represented in Figure 1.2.

In the first column of the figure we find the dyadic dilates (2 times, 4 times, 8 times, etc.) of the scaling function φ and in the second column, those of the wavelet ψ.

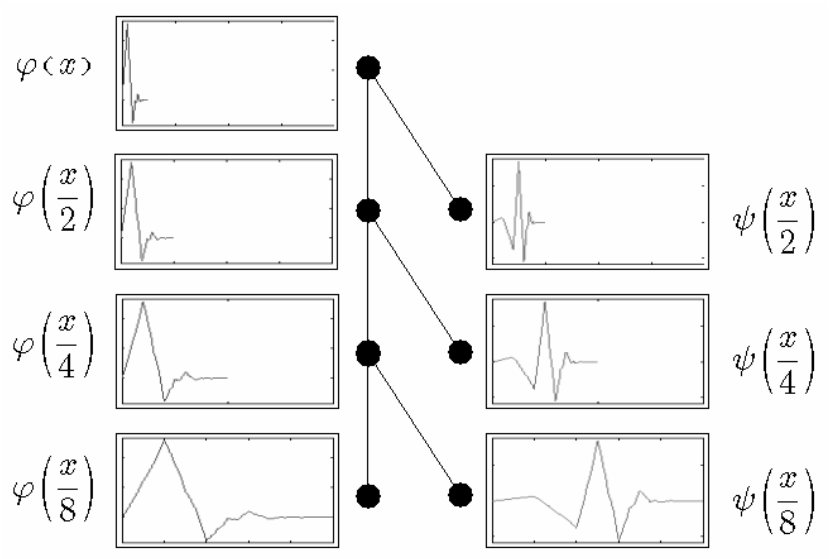


Figure 1.2. *Organization of wavelets*

The functions in the first column are used for calculating the coefficients of approximation $\beta_{j,k} = \int_{\mathbb{R}} s(t)\varphi_{j,k}(t)\,dt$, which define local averages of the signal $s(t)$. The signal $A_j(t) = \sum_{k\in\mathbb{Z}} \beta_{j,k}\,\varphi_{j,k}(t)$ is an approximation.

The functions in the second column are associated with the calculation of wavelet coefficients $\alpha_{j,k} = \int_{\mathbb{R}} s(t)\psi_{j,k}(t)\,dt$, which relate to the differences between two successive local averages. These are the final touches (we shall call them details) of the form:

$$D_j(t) = \sum_{k\in\mathbb{Z}} \alpha_{j,k}\,\psi_{j,k}(t)$$

We therefore have four kinds of objects:

– detail coefficients $\alpha_{j,k} = \int_{\mathbb{R}} s(t)\psi_{j,k}(t)\,dt$, which are also wavelet coefficients enabling us to define the details;

– detail signals themselves: $D_j(t) = \sum_{k\in\mathbb{Z}} \alpha_{j,k}\,\psi_{j,k}(t)$;

– approximation coefficients: $\beta_{j,k} = \int_{\mathbb{R}} s(t)\varphi_{j,k}(t)dt$ making it possible to calculate the approximations;

– approximation signals themselves: $A_j(t) = \sum_{k\in\mathbb{Z}} \beta_{j,k}\,\varphi_{j,k}(t)$.

Detail and approximation signals are related to t, the time of the original signal, whereas the coefficients (the $\alpha_{j,k}$ and the $\beta_{j,k}$) of the level j are in dyadic time $2^j\mathbb{Z}$. The details and approximations are interpreted in terms of the orthogonal projection onto spaces W_j and V_j, respectively. For a signal s, if we compare the values of the signal to the coefficients $\left\{\beta_{0,k}\right\}_{k\in\mathbb{Z}}$ in V_0, the $\left\{\alpha_{j,k}\right\}_{k\in\mathbb{Z}}$ and $\left\{\beta_{j,k}\right\}_{k\in\mathbb{Z}}$ are the coefficients of s with respect to the bases of W_j and V_j respectively; while D_j and A_j are elements of spaces W_j and V_j, considered as functions of V_0.

1.2.4. *The wavelet tree for a signal*

Such a tree is presented in Figure 1.3. At its root we find a signal s (we may also say a time series). The tree can be read in various ways. The first column yields three approximations, from the finest A_1 to the coarsest A_3, as may be realized focusing on the end of the signal. The differences between two successive approximations are captured in the details denoted D_1 to D_3. More precisely, we have $D_1 = s - A_1, D_2 = A_1 - A_2$ and, thus, $s = A_2 + D_2 + D_1$.

Let us return to the case of a signal s, which is in continuous time. Starting from the equality $s = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \alpha_{j,k}\,\psi_{j,k}$, meaning that we can reconstruct the signal on from its coefficients $\alpha_{j,k}$, we may use it to define the detail at level j differently. Let us fix j and sum up using k. Again we find the detail D_j:

$$D_j = \sum_{k\in\mathbb{Z}} \alpha_{j,k}\,\psi_{j,k}$$

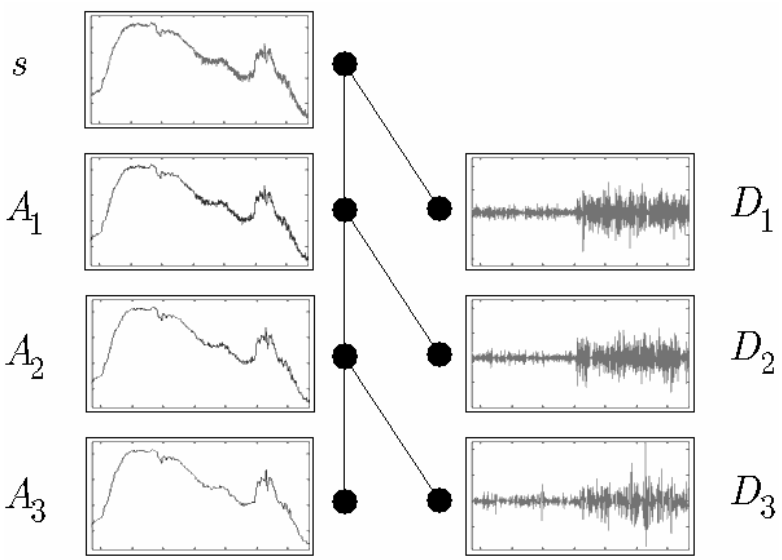


Figure 1.3. *The wavelet tree for a signal*

Let us then sum up using j. We rediscover that the signal is the sum of its details:

$$s = \sum_{j \in \mathbb{Z}} D_j$$

The details are defined. Let us now take a reference level marked J ($J = 3$ in this example). There are two kinds of details: those associated with the indices $j \leq J$ corresponding to the scales $a = 2^j \leq 2^J$, which are finer details than the resolution corresponding to J; and those for which $j > J$, are coarser. Let us aggregate the latter:

$$A_J = \sum_{j > J} D_j$$

This sum defines what will be referred to as the approximation at level J of the signal s. Moreover:

$$s = A_J + \sum_{j \leq J} D_j$$

This relation means that s is the sum of its approximation A_J and of the finer details. It can be deduced from it that the approximations are linked by:

$$A_{J-1} = A_J + D_J$$

In the orthogonal case, the family $\left\{ \psi_{j,k} \right\}_{j,k \in \mathbb{Z}}$ is orthogonal and we have:

- A_J is orthogonal to $D_J, D_{J-1}, D_{J-2}, \cdots$;
- s is the sum of two orthogonal signals: A_J and $\sum_{j \leq J} D_j$;
- the quality Q_J of the approximation of s by A_J is equal to $Q_J = \frac{\| A_J \|^2}{\| s \|^2}$ and we have $Q_{J-1} = Q_J + \frac{\| D_J \|^2}{\| s \|^2}$.

1.3. An electrical consumption signal analyzed by wavelets

As a first example we consider a minute per minute record of the electrical consumption of France; the problem is presented in detail in [MIS 94].

In Figure 1.4 we find, from top to bottom, the original signal (**s**), the approximation at level 5 ($\mathbf{a_5}$) and the details from the coarsest level ($\mathbf{d_5}$) to the finest level ($\mathbf{d_1}$). All the signals are expressed in the same of time unit, which allows a synchronous reading of all the graphs. The wavelet used is *db3*.

The analyzed signal represents to the nearest transformation three days of electrical consumption during summer in France. The three days are Thursday followed by Friday, having very similar shape and amplitude, then Saturday, which

has a much lower average level due to the start of the weekend and ebbing economic activity. We may thus reinterpret the time scales corresponding to each level.

Ignoring the effects of wavelet choice, we can roughly state that $\mathbf{d_1}$ contains the components of the signal of period between 1 and 2 minutes, $\mathbf{d_2}$ those of period between 2 and 4 minutes and so on until $\mathbf{d_5}$ which contains the components of the signal of period between 16 and 32 minutes. The $\mathbf{a_5}$ approximation contains signal components of period greater than 32 minutes.

A quick examination shows that:

– the analysis makes it possible to track possible outliers, which are detected thanks to the very large values of $\mathbf{d_1}$ around the position 1,200;

– in the graphs we can distinguish the details $\mathbf{d_1}$ and $\mathbf{d_2}$ (measurement and state noises whose amplitude is low in normal circumstances) which yield details oscillating quickly around 0;

– on the contrary we isolate the period of sensor failure corresponding to the long sequence of abnormally large values of $\mathbf{d_1}$, $\mathbf{d_2}$ and, more slightly, $\mathbf{d_3}$, between the position 2,500 and the position 3,500. We have here an additional noise;

– moreover, we notice that in the details $\mathbf{d_4}$ and $\mathbf{d_5}$ we no longer distinguish this period, and the difference due to an exceptional sensors noise disappears giving an indication of the frequency contents of the additional noise;

– the daily pseudo-periodicity evident in the analyzed trajectory ($\mathbf{s}$) can be read, for the scales examined, from levels 4 and 5 through the appearance of periodic patterns in the details at equally spaced positions. From that we deduce that the details $\mathbf{d_4}$ and $\mathbf{d_5}$ contain components of the useful signal as opposed to the non-informative noise. This phenomenon does not occur for the finer levels and we could believe that the details consist almost entirely of noise.

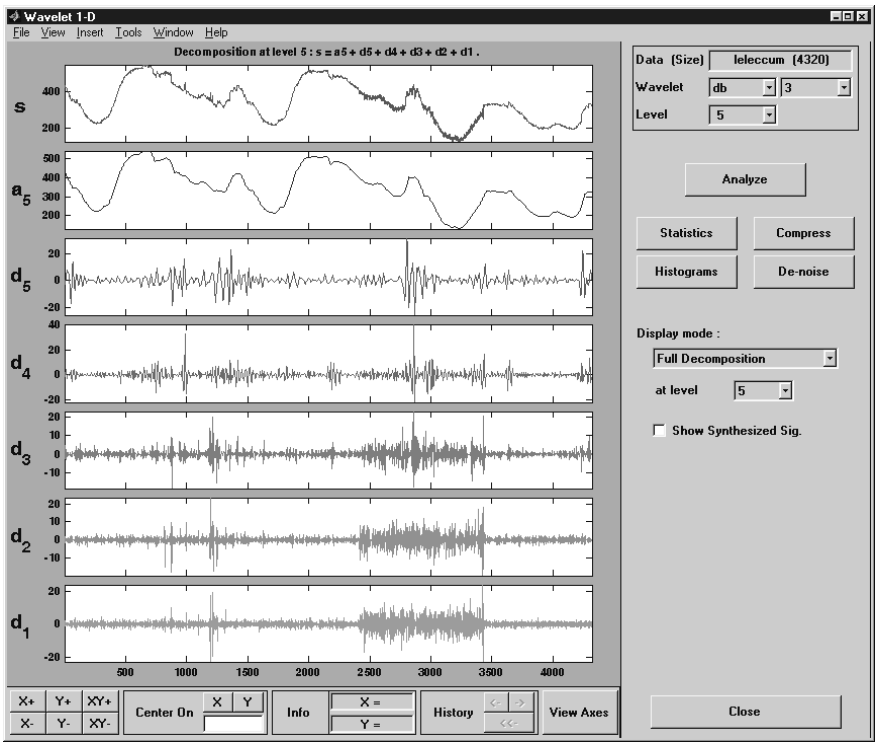


Figure 1.4. *An electrical consumption signal analyzed by wavelets*

This quick example illustrates that a simple analysis by wavelets can yield many pathways for finer processing and direct the strategy of solving classical problems, such as outlier detection, denoising reduction or signal extraction.

1.4. Denoising by wavelets: before and afterwards

Denoising is the major application of wavelets in statistics. This problem admits a very elegant solution.

In Figure 1.5 (top) we see a portion of the real noisy signal analyzed in Figure 1.4.

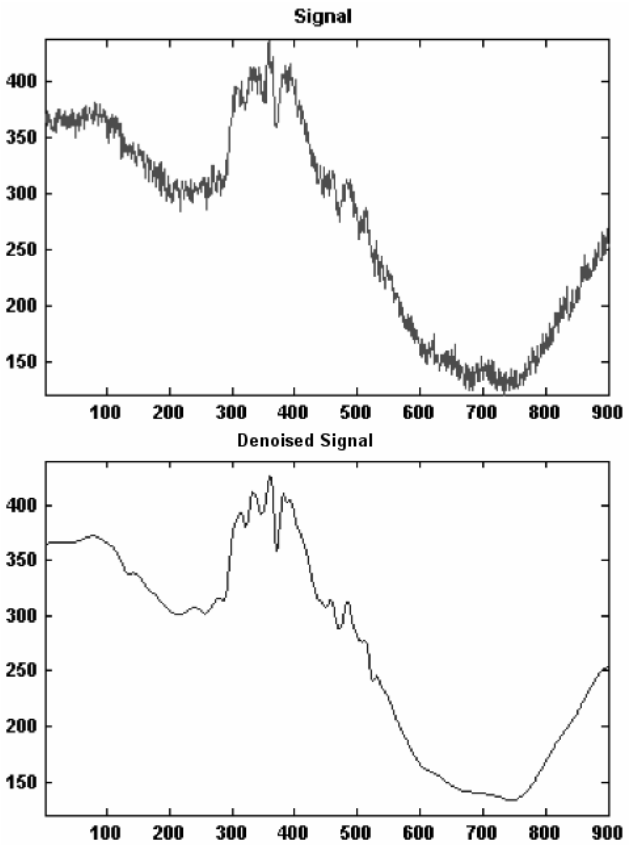


Figure 1.5. *Denoising by wavelets: before and after*

The denoised signal using wavelets is located in the lower part. It is obviously well denoised, in the zones where the signal is smooth (around positions 200 or 800, for example), as well as in the zone around instant 400 where the signal is irregular. The traditional methods of denoising are not capable of such an adaptation in time. These methods estimate the function f using the model:

$$Y_i = f(t_i) + \varepsilon_i\,, \quad t_i = \frac{i}{n}\,, \quad i = 1,\cdots,n$$

where f is unknown, $(Y_i)_i$ are observed and ε is an unobservable white noise.

1.5. A Doppler signal analyzed by wavelets

Let us first consider a simulated signal enabling us to clearly understand the spirit of the technique of denoising by wavelets (see Figure 1.6).

The screen is organized in two columns: in the first one we see the noisy signal **s**, then, in the lower part, we see the approximations from level 5 (the coarsest, $\mathbf{a_5}$) to level 1 (the finest, $\mathbf{a_1}$); in the second column at the top we see a colored version of the wavelet coefficients of levels 5 to 1 (**cfs**), followed by the noisy signal **s** and then, in the lower part, the details from level 5 (the coarsest, $\mathbf{d_5}$) to level 1 (the finest, $\mathbf{d_1}$).

The wavelet used, *sym4*, is a compactly supported almost symmetric wavelet of order 4 (see Chapter 4).

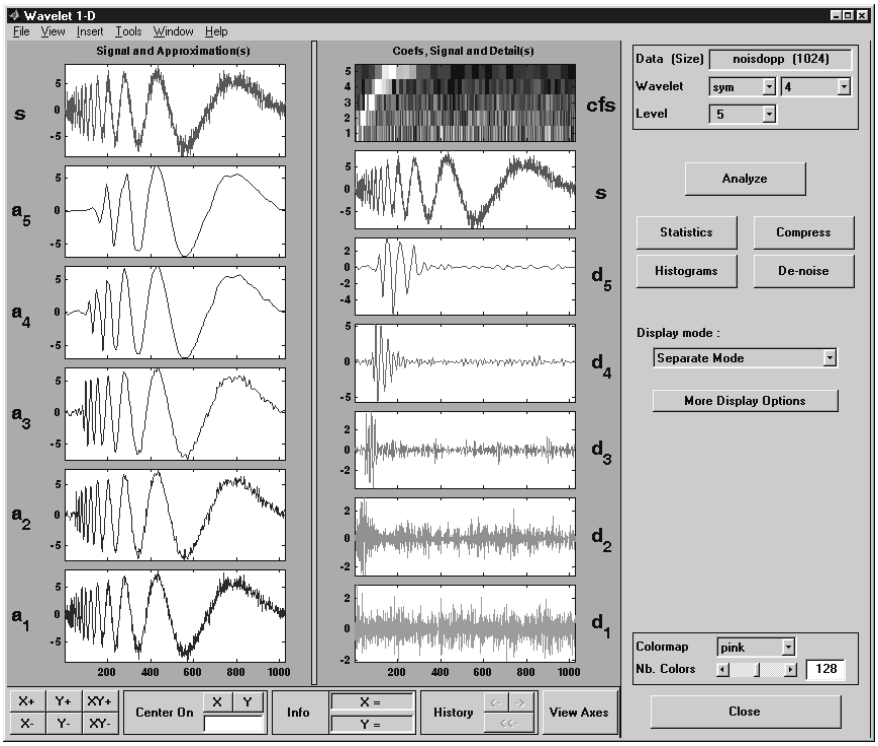


Figure 1.6. *Doppler signal analysis by wavelets*

Let us start by examining the first column and concentrate on the portion of the signal corresponding to the x-coordinates from 200 to 1,000. Starting from $\mathbf{a_1}$ let us seek, ascending, a level such that the approximation constitutes a good candidate to be an estimator of the useful signal. Levels 4 and 5 are reasonable. Nevertheless, the estimator associated with $\mathbf{a_4}$ is clearly very bad for the beginning of the signal corresponding to the x-coordinates 0 to 100. Conversely, an acceptable restitution at the beginning of the signal would result in choosing $\mathbf{a_2}$, which is visibly too noisy.

Let us now look at the details. Detail $\mathbf{d_1}$ seems to consist entirely of noise while details $\mathbf{d_2}$ to $\mathbf{d_5}$ present large values concentrated at the x-coordinates from 0 to 300. This is also visible on the graph of the wavelet coefficients (**cfs**), the largest in absolute value being the clearest. This form stems from the fact that the signal is a sinusoidal function with amplitude and period growing with time. The oscillations at the smallest scales explain the displayed details; the others are in the $\mathbf{a_5}$ approximation.

Thus, a plausible denoising strategy consists of:

– keeping an approximation such that the noise is absent or at least very attenuated ($\mathbf{a_4}$ or $\mathbf{a_5}$);

– supplementing this approximation by parts of the finer details clearly ascribable to the useful signal and rejecting the parts which are regarded as stemming from the noise.

This is precisely what the denoising by wavelets methods achieve, but in an automatic fashion. The *ad hoc* choice suggested for this particular simulated signal is that carried out by one of the most widespread methods of denoising by wavelets according to Donoho and Johnstone (see [DON 94], [DON 95a], [DON 95b]).

1.6. A Doppler signal denoised by wavelets

Let us consider Figure 1.7. The screen is organized in two columns. In the first we see wavelet coefficients from level 5 to level 1. To make them more readable, they are "repeated" 2^k times at level k (which explains the sequences of a constant gray level especially visible for $k > 3$).

In each one of these graphs we note the presence of two horizontal dotted lines: the coefficients inside the tube are zeroed by the process of denoising. In the second column at the top the noisy signal s is superimposed over the denoised signal. In the middle we find a color version of the wavelet coefficients from level 1 to 5 of the original noisy signal and in the graph at the bottom the counterpart for the

thresholded wavelet coefficients, from which the denoised signal is reconstructed. The *sym4* wavelet is used as previously.

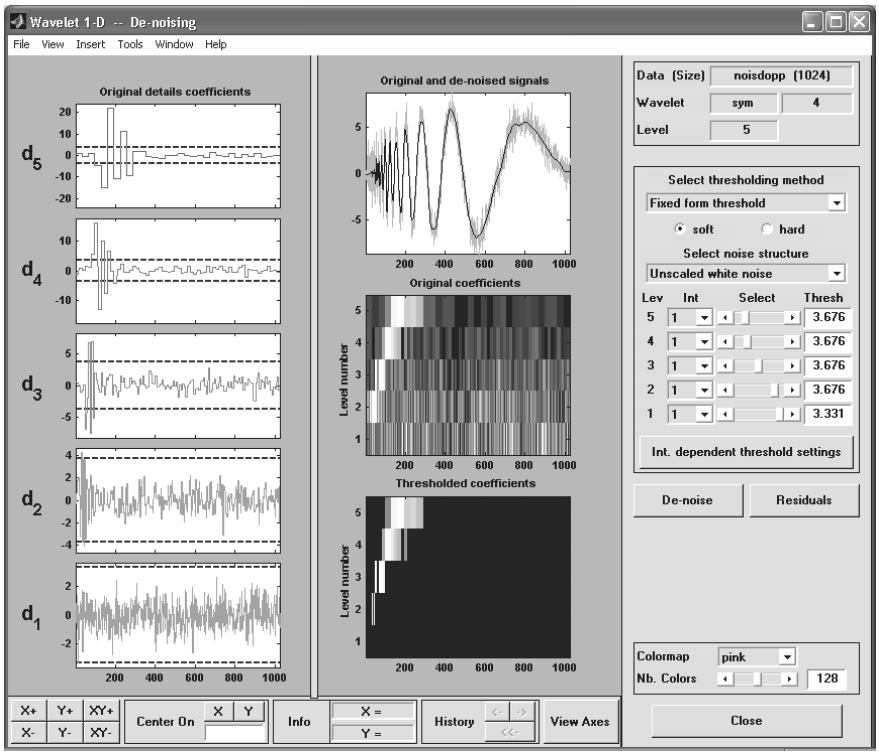


Figure 1.7. *Doppler signal denoised by wavelets*

Let us use this example to see how denoising is performed:

– all the approximation coefficients are kept, which from the onset leads to introducing the first component of the denoised signal which has the following form ($\hat{A}_d$ is not represented directly in Figure 1.7; it corresponds to the $\mathbf{a_5}$ approximation in Figure 1.6):

$$\hat{A}_d = \sum_k \hat{\beta}_{5,k} \varphi_{5,k}$$

The estimated coefficients $\hat{\beta}_{5,k}$ are simply the coefficients obtained by decomposition of the initial signal;

– only the few larger wavelet coefficients (i.e. detail coefficients) are preserved; the others are replaced by zero. In the second column we may see the coefficients before and after this thresholding operation to note the sparsity of the preserved coefficients (the black zone of the graph at the bottom corresponds to the cancelled coefficients). We thus obtain the second component of the denoised signal which has the form:

$$\hat{D}_d = \sum_{1 \leq j \leq 5} \sum_k \hat{\alpha}_{j,k} \psi_{j,k}$$

The estimated coefficients $\hat{\alpha}_{j,k}$ are simply the $\alpha_{j,k}$ coefficients obtained by decomposition of the initial signal and then thresholded. Many automatic methods are available for the choice of thresholds, depending on the form and the hypotheses concerning the model supposed to suitably represent the way of generating the data. We do not detail them here;

– the denoised signal is thus:

$$\hat{s}_d = \hat{A}_d + \hat{D}_d$$

It is visible that the result obtained at the top on the right has a good quality, except at the very beginning of the signal which oscillates too much on a small scale with fluctuations that are small compared to those of the noise.

1.7. An electrical signal denoised by wavelets

Let us return to the problem of denoising a real signal, which is more difficult since the nature of the noise is unknown. The method employed here takes the previous situation as a starting point by adapting the thresholding to the level.

The screen presented in Figure 1.8 is organized in two columns as in Figure 1.7. The wavelet used is *coif5*. Examining the first column we see that the coefficients of levels 1 to 3 have all been considered by this method as ascribable to the noise and that they have all been zeroed. This conforms to the conclusions of the analysis of the three day electrical load plot carried out before. On the other hand, for levels 4 and 5, the process mainly selects wavelet coefficients in the zone around the position 375, thus enabling an excellent restitution of the abrupt signal changes.

The key arguments to understand the effectiveness of these methods are as follows:

– the decomposition by wavelets is an additive analysis, consequently, the analysis of $Y_t = f(t) + \varepsilon_t$ is equal to the sum of the analyses of the signal $f(t)$ and the noise ε_t;

– if we suppose that the noise ε is white with a constant variance, the wavelet coefficients on all scales are white noise with the same variance. In addition, real signals are in many cases regular enough except in rare locations (start and end of transitory phenomena, ruptures for example), which renders the decomposition by wavelets of $f(t)$ very sparse and very well represented by the coefficients of a rather rough approximation and some large detail coefficients;

– if the irregularities generate coefficients larger than the scale of the noise, the process of thresholding only selects coefficients related to the signal provided that the scale of the noise can be suitably estimated. Lastly, the operation of thresholding always leads to regularize the signal.

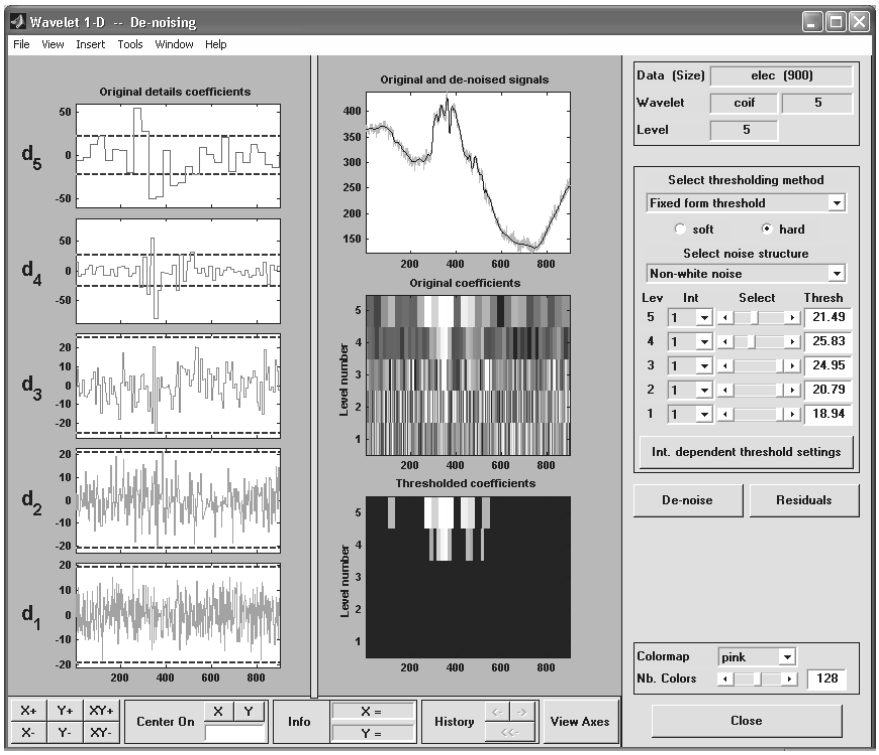


Figure 1.8. *Electrical consumption signal denoising by wavelets*

1.8. An image decomposed by wavelets

1.8.1. *Decomposition in tree form*

Let us now pass to image processing by wavelets and examine first Figure 1.9. At the top left we find the original image, in the lower part a table of images with three rows and four columns. The whole figure constitutes the tree of decomposition by wavelets; in 2D it is a quaternary tree while in 1D it is a binary tree.

In 1D the signal is decomposed into two: an approximation and a detail; in 2D the image is decomposed into four: an approximation (first column of the 3×4 table) and three details in three directions, horizontal, diagonal and vertical (three last columns of the 3×4 table). The rows of this table are indexed by levels: level 1 (the finest, noted $\mathbf{L_1}$) to level 3 (the coarsest, noted $\mathbf{L_3}$). The wavelet used here is *sym4*.

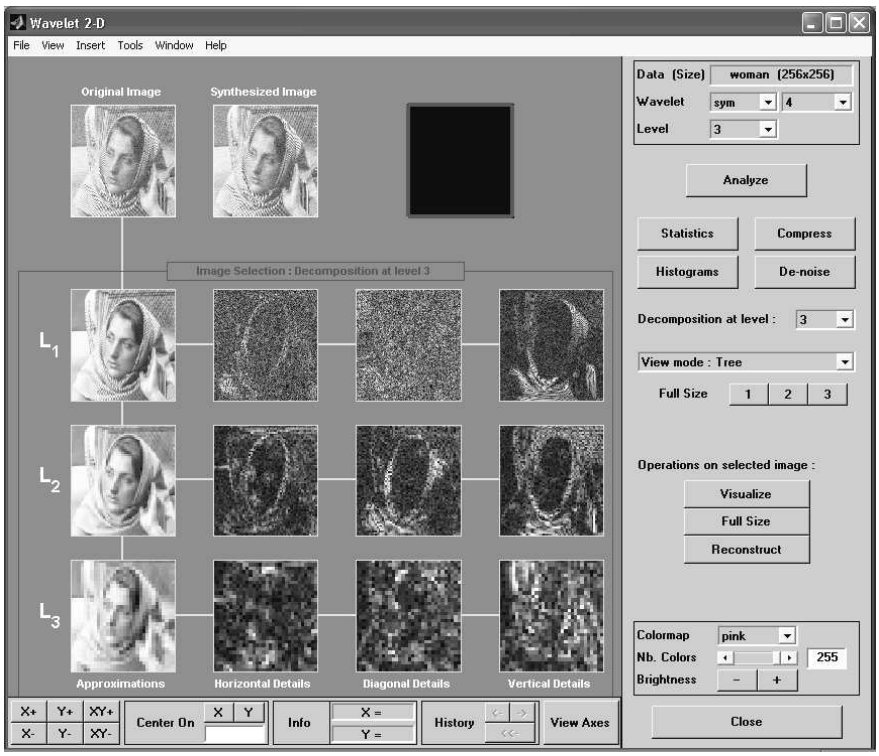


Figure 1.9. *An image ("Barbara") decomposed by wavelets. Decomposition in tree form*

Before commenting further on this screen, let us say a few words about orthogonal wavelets in 2D. These are very particular cases of wavelets, by far the most used because they lead to fast calculations, which is critical in image processing. In 1D we have two functions: φ and ψ, and the two associated filters. On this basis we produce the scaling function and the wavelets which no longer operate on the real line, but on the plane:

- $\varphi(x,y) = \varphi(x)\varphi(y)$ used to define approximations;
- $\psi_1(x,y) = \psi(x)\varphi(y)$ used to define horizontal details;
- $\psi_2(x,y) = \psi(x)\psi(y)$ used to define diagonal details;
- $\psi_3(x,y) = \varphi(x)\psi(y)$ used to define vertical details.

Algorithmic simplicity comes from the fact that it is possible to successively apply to the rows and columns of the matrix associated with the image the two filters (lowpass and highpass, respectively) useful in 1D.

Carefully examining the three approximations (in fact, their coefficients) we note that they are increasingly coarser versions of the original image.

By comparing the original image with the level 3 approximation coefficients it becomes clear that details are lost on a small scale, such as, for example, the shawl design or the weft of the cane armchair behind the face. The coefficients of detail are in general more difficult to exploit. Nevertheless, we distinguish large coefficients in the zones mentioned above, especially at level 2. We also see face features coming out in the horizontal detail of level 2.

1.8.2. *Decomposition in compact form*

The preceding representation of decomposition (see Figure 1.9) focuses on the tree structure; the one proposed in Figure 1.10 uses 2D decomposition coding.

Indeed, we find four sub-figures in this figure. The original image is top left. Let us comment on the sub-figure at the bottom right consisting of the original image decomposition coefficients.

To read this graph, we initially consider that the image is divided into quadrants. The quadrants at the bottom left, bottom right and top right are not decomposed. The three small images found represent (in the counterclockwise direction) the coefficients of vertical, diagonal and horizontal details at the finest level (level 1),

while the three following small images, located in the top left quadrant, represent the detail coefficients at level 2, and so on until we find a single small image (this is the case for the top left), which contains the approximation coefficients at the coarsest level (here 2).

The two other images, bottom left and top right, represent respectively, the image resulting from a compression process (see Figure 1.11) and the image reconstructed from only level 2 approximation coefficients. The wavelet used is still *sym4*.

The image analyzed here is a zoom of the one in Figure 1.9 and from this point of view enables finer comments. Let us concentrate on the images reproduced on the anti-diagonal of the screen. They have the same resolution as the original image and therefore can be compared directly.

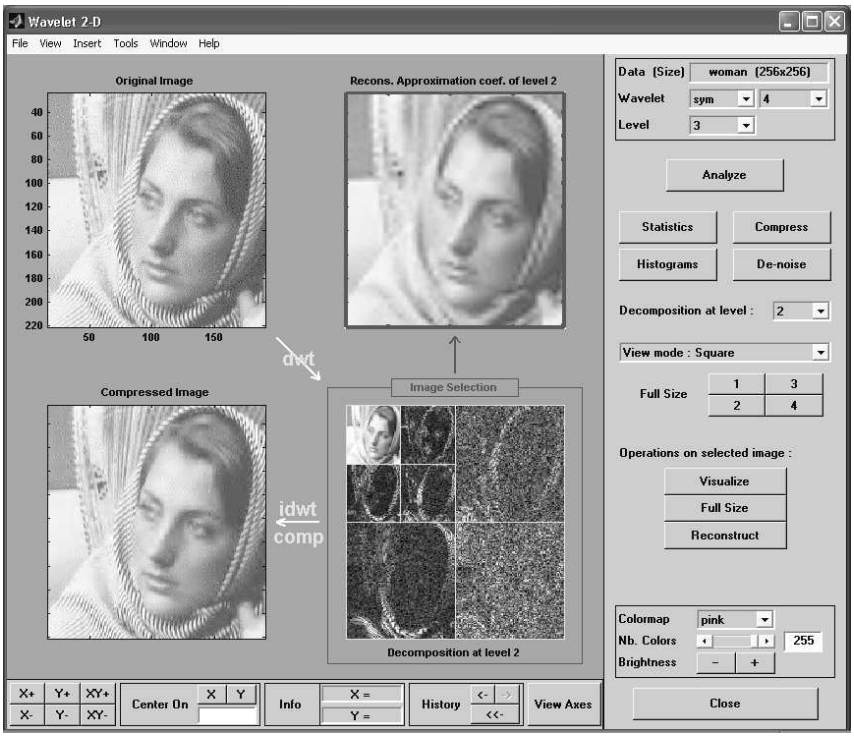


Figure 1.10. *An image ("Barbara") decomposed by wavelets. Decomposition in compact form*

Let us examine the top right image comparing it with the original image. We find the elements mentioned in the commentary on the preceding figure. The shawl design was lost in the level 2 approximation, only being discernible at scales lower

than 4 pixels. Here we recover only a rather homogenous texture with the two well restored principal folds and slightly less well-defined intermediate fold. The same holds for the armchair which is difficult to recognize. Moreover, the features of the face are fuzzier.

This general loss of contours definition can be largely improved by selectively adding coefficients of suitably selected details. After such an operation we obtain the compressed image located bottom left. It is not perfect but, nevertheless, much more accurate. Here is how it was obtained.

1.9. An image compressed by wavelets

At the top left of Figure 1.11 we find the original image, which needs to be compressed and at the top right we have the compressed image.

Below, for each level from the finest (level 1, denoted by $\mathbf{L_1}$) to the coarsest (here level 3, denoted by $\mathbf{L_3}$) we find (for each orientation: horizontal, diagonal and vertical) the grayscale histogram of the corresponding detail coefficients. In each one of these histograms we note the presence of two vertical dotted lines: the coefficients inside the vertical tube are zeroed by the process of compression.

The percentage of zeros in the representation by wavelets is more than 95%. It is an indicator of the space freed up by compression. This is a compression method with loss: approximately 98% of energy is preserved.

The criterion of energy is neither a very meaningful nor relevant indicator for the images. To evaluate the quality of restitution many numerical criteria exist but nothing is more critical than the human eye.

The method of compression is similar to that used in denoising; it is the criterion which changes. We preserve the coefficients of the roughest approximation, with which we associate the largest detail coefficients, then we reconstruct. The tuning shown in the screen presented in Figure 1.11 is manual and carried out by level and direction.

Let us examine the grayscale histograms of the detail coefficients:

– at the bottom there appear those concerning level 1. For each of the three directions all the coefficients are replaced by zero since the histogram is contained in the zone delimited by the two vertical dotted lines;

– for level 2 the strategy is a little less selective and preserves the majority of the coefficients with large absolute values;

– finally, for level 3 which is critical for image contours restitution, only the very small coefficients are eliminated, the vertical features being very close, approaching zero.

The image reconstructed by using thresholded coefficients has very good quality, despite the high percentage of coefficients replaced by zero.

These results may be improved considerably using suitably selected biorthogonal wavelets (a decomposition wavelet with a sufficient number of vanishing moments so that the representation by wavelets is the most sparse possible, associated with a dual wavelet with very regular and symmetric reconstruction in order to remove as many visual artifacts as possible). Later on we shall see an illustration thereof.

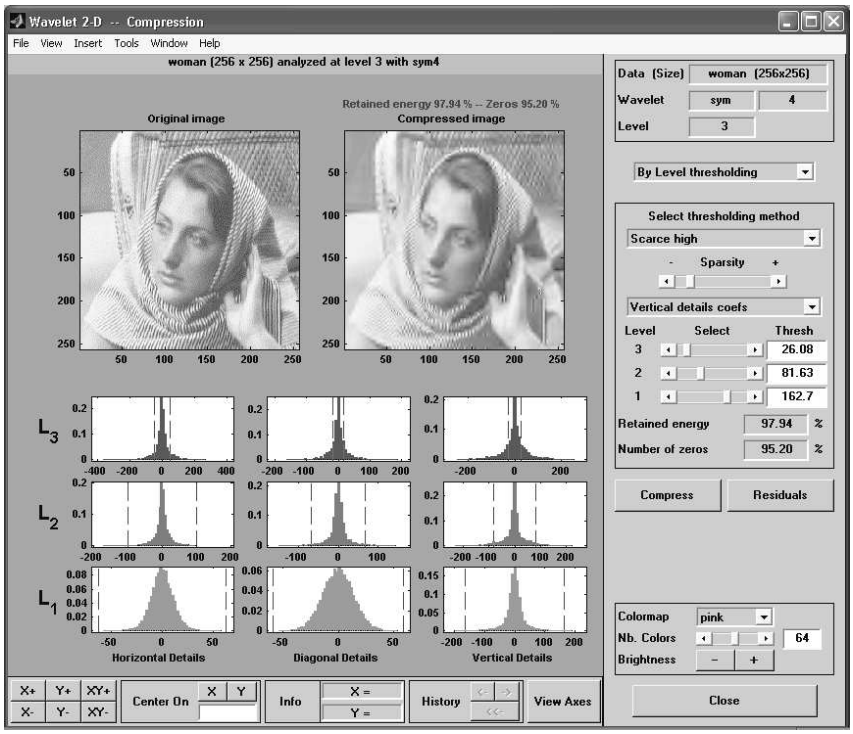


Figure 1.11. *An image ("Barbara") compressed by wavelets*

1.10. A signal compressed by wavelets

Before passing to fingerprints compression let us mention the methods of compression by wavelets of one-dimensional signals, using an artificial example.

The screen presented in Figure 1.12 is organized in two columns. In the first one we find a single graph making it possible to adjust a global threshold for the detail coefficients of the signal to be compressed (those of the approximation being preserved). All the detail coefficients whose absolute value is lower than the threshold (determined by the x-coordinate of the vertical dotted line) are zeroed by the compression. The possible values of the threshold are given by the x-axis. Two curves are drawn: one increasing which gives the percentage of zeros in the representation of the compressed signal, the other decreasing which gives the percentage of the energy preserved by the compressed signal.

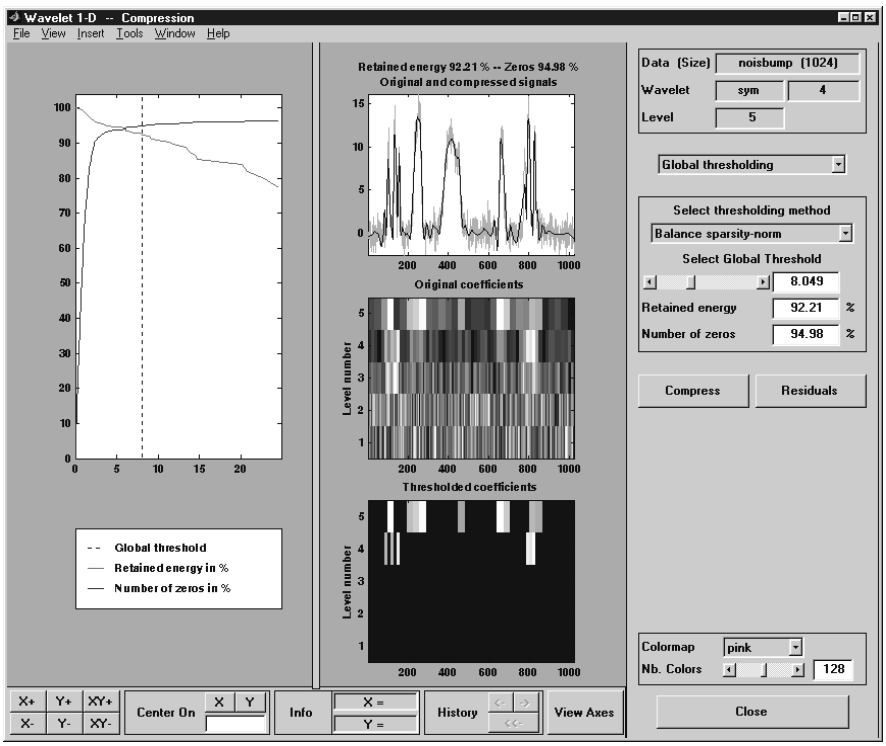


Figure 1.12. *A signal compressed by wavelets*

In the second column, the original signal is superimposed on the compressed signal. In the lower part we find a color version of the wavelet coefficients of levels 1 to 5 of the original signal and in the graph underneath is the counterpart for the thresholded wavelets coefficients, from which the compressed signal is reconstructed.

The percentage of zeros in the representation by wavelets is 95% (5% of the coefficients are preserved), for 92% of energy preserved: this stems from the fact that the signal is noisy and that therefore the very fast fluctuations are lost. The graphs (bottom right) of the coefficients before and after thresholding make it possible to notice the sparsity of the preserved coefficients. The wavelet used is *sym4*.

The compression of one-dimensional signals, although less crucial than image compression, has many applications that sometimes present a great economic interest. This is the case, for example, for companies forced to preserve the individual consumption profiles of their customers over long periods with a high degree of accuracy.

As the form of the proposed graphical interface suggests, to carry out compression by global thresholding the ideas are very close to those of denoising and the implemented algorithm is identical. It operates in three steps whose general schematic is as follows:

– decomposition by wavelets;

– we preserve the coefficients of the coarsest approximation as well as the largest wavelet coefficients in absolute value; the others are replaced by zero;

– from these modified coefficients we reconstruct the compressed signal.

1.11. A fingerprint compressed using wavelet packets

In Figure 1.13 (top right) we find the original image and (bottom left) the compressed image.

At the top left we find the wavelet packets tree and at the bottom right, the decomposition of the image to be compressed. This decomposition is obtained by decomposing into four (approximation and three details) not only the approximations but also the details. The Haar wavelet is used; the result is good but can be improved using a biorthogonal wavelet.

In Figure 1.14, at the top left we find the original image and at the top right is the compressed image. The method of compression involves a global thresholding of the wavelet packets coefficients of the image to compress. The graph underneath the original image visualizes the choice of the global threshold exactly as in 1D case. The wavelet used is a biorthogonal one.

The result is very good for a percentage of zeros of the compressed image representation equal to 95%.

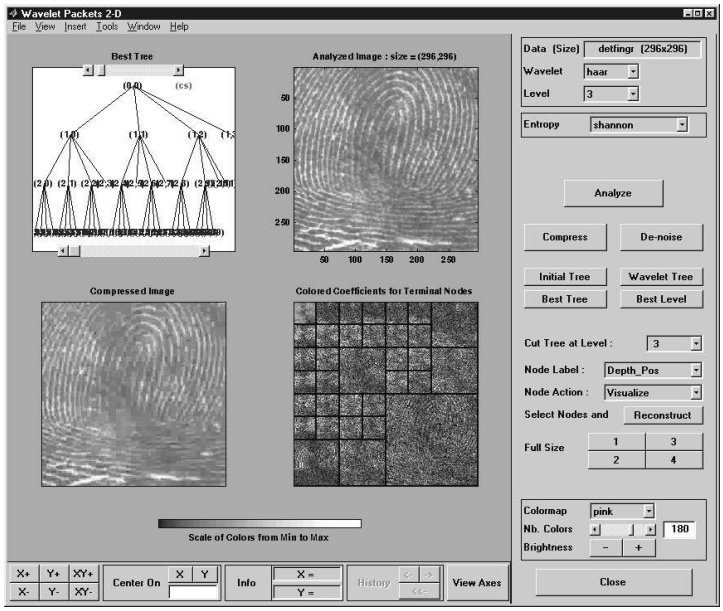


Figure 1.13. *A fingerprint analyzed and compressed by Haar wavelet packets*

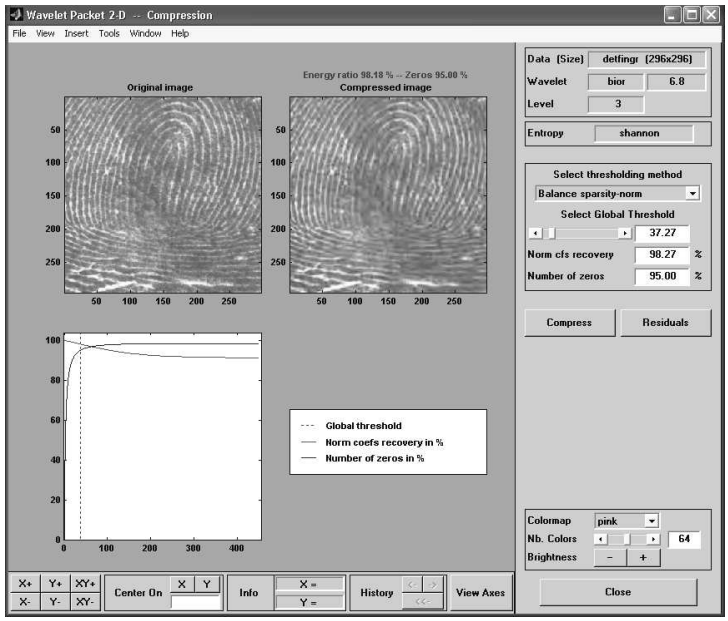


Figure 1.14. *A fingerprint compressed by biorthogonal wavelet packets*

Chapter 2

Mathematical Framework

2.1. Introduction

Several points of view are needed to realize the interest of wavelets as a mathematical tool of function analysis and representation. Two methods are proposed in this chapter.

The first method centers on the idea of building a tool for *local analysis in time*. We introduce the *continuous wavelet transform* by showing how it solves the problems that are from this point of view raised by the Fourier transform and the sliding window Fourier transform, also called Gabor transform. In fact, the first is a global transformation and second is local, but with fixed temporal resolution. On the other hand, the wavelet transform is not just a local analysis; its temporal resolution is variable. We underline its capacity to describe the local behavior of signals on various timescales.

The second method is the inversion of analysis and the search for economic representations. These are integral transforms: they are obtained by integrating the signal multiplied by the basic analyzing functions. As for any transformation, the question of reconstruction (or inversion) arises: if $g = T(f)$ is the transform of f, is it possible "to recover", to reconstruct f knowing g? In the three cases studied the answer is: yes, under appropriate conditions. In the "very good cases", it is even possible to reconstitute *f* from discrete values of the transform. This has major advantages from a numerical point of view. Thus, the *discrete wavelet transform* is introduced. Seeking "to minimize" the number of discrete information necessary to reconstruct f we are naturally led to the concept of *wavelets base*. We then introduce the orthonormal wavelet bases starting from the concept of multi-

resolution analysis of the space of the finite energy signals. This provides a framework for the decomposition of a signal in the form of a succession of approximations with decreasing resolution, supplemented by a sequence of details.

We then present the *bases of wavelet packets*, which constitute a generalization of the orthonormal wavelet bases making it possible, at the same time, to improve the frequency resolution of the analysis in wavelets and to propose a richer analysis associated with a large collection of decompositions. It is then possible to select the decomposition best adapted to a given signal, with respect to an entropy criterion.

Lastly, we introduce the *biorthogonal wavelet bases* whose idea is to slacken the strong constraints that an orthonormal wavelet basis must verify. The key is to consider two wavelets instead of just one. The duality links between the analyzed and synthesized wavelets are loose enough to partially uncouple the properties of each of the two bases depending on the objectives.

This chapter outlines a theoretical framework without providing the proofs and the results are not always expressed in a "tight" mathematical language. Let us then finish this introduction with some bibliographical indications. The issues discussed in this chapter are traditional and are thus found in all the works offering broad presentations of the wavelets, in particular in [DAU 92], [MAL 98] and [KAH 98]. On the continuous analysis in wavelets it is moreover possible to refer to [LEM 90], [TEO 98] and [TOR 95]; on the orthonormal bases to [FRA 99], [WAL 02] and [STR 89], on wavelet packets to [COI 92] and [WIC 94], and on biorthogonal bases to [COH 92b] and [COH 92c].

In the next sections we will note time as $t \in \mathbb{R}$ and frequency as $\omega \in \mathbb{R}$. The space of square integrable functions (called signals) is noted $L^2(\mathbb{R})$ or simply L^2. The square norm $\int |s(t)|^2 dt$ is called energy of the signal s. The quantifiers are frequently omitted.

2.2. From the Fourier transform to the Gabor transform

2.2.1. *Continuous Fourier transform*

The Fourier transform is noted $\mathbf{F}$ and the inverse transform $\bar{\mathbf{F}}$. The formulae of analysis and synthesis of the Fourier transform for an integrable function are given by:

$$\textit{Analysis}: \hat{f}(\omega) = \left(\mathbf{F}f\right)(\omega) = \int_{\mathbb{R}} f(t) e^{-2i\pi\omega t} dt, \qquad \omega \in \mathbb{R}$$

$$\textit{Synthesis}: f(t) = \left(\overline{\mathbf{F}}\hat{f}\right)(t) = \int_{\mathbb{R}} \hat{f}(\omega) e^{+2i\pi\omega t} d\omega\,, \qquad t \in \mathbb{R}$$

NOTE 2.1.– the synthesis formula is formal, without additional assumptions about the function f. The Fourier transform extends to finite energy functions and convergence in the synthesis formula is in L^2.

■

In signal processing $\hat{f}(\omega)$ is often called the spectrum of f, $\left|\hat{f}(\omega)\right|$ is the energy spectrum and $Arg(\hat{f}(\omega))$ is the phase.

For signals (or functions) with finite energy the transformation $\mathbf{F}$ has the following properties: it is linear, continuous, admits a linear and continuous inverse transform $\overline{\mathbf{F}}$, and preserves the angles and energies (scalar products and norms). This is stated formally as follows:

THEOREM 2.1.

The Fourier transform $\mathbf{F}$ is a bijective continuous linear application of L^2 on L^2. The inverse bijection is $\overline{\mathbf{F}}$ and we thus have:

$$f = \overline{\mathbf{F}}\mathbf{F}f = \mathbf{F}\overline{\mathbf{F}}f$$

In addition, the scalar product and thus the norm are preserved:

$$(f,g)_{L^2} = \left(\mathbf{F}f, \mathbf{F}g\right)_{L^2} = \left(\overline{\mathbf{F}}f, \overline{\mathbf{F}}g\right)_{L^2} \quad \text{and} \quad \|f\|_{L^2} = \left\|\mathbf{F}f\right\|_{L^2} = \left\|\overline{\mathbf{F}}f\right\|_{L^2}$$

■

The atoms of this transform are functions of t: $e^{-2i\pi\omega t}$.

While examining the analysis formula, some of the disadvantages of the Fourier transform appear.

Disadvantage 1. Removal of the time aspects

The temporal aspects of the function f disappear in $\hat{f}$. Indeed, if f is not continuous, it is almost impossible to detect it by using $\hat{f}$ as presented by the elementary example (see Figure 2.1). Let f be a square pulse signal equal to 1 in $[-a,a]$ and 0 elsewhere, noted $f = \mathbb{1}_{[-a,a]}, a > 0$. Its Fourier transform is $\hat{f}(\omega) = \sin(2\pi a\omega)/\pi\omega$.

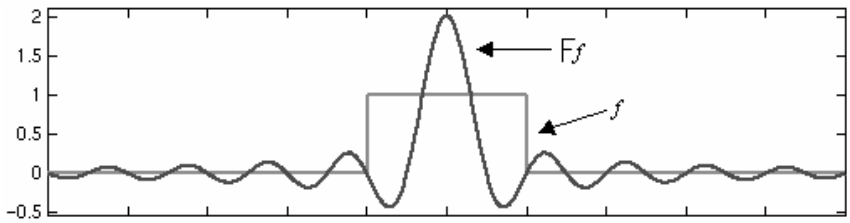


Figure 2.1. *Fourier transform of* $\mathbb{1}_{[-a,a]}$ *for a = 1*

If we know that the studied function is a square function, we can find the parameter *a* by seeking the distance between two successive zeroes of the Fourier transform. This becomes too complicated for a more composite signal, even for a simple linear combination of pulse functions, $f = \alpha\mathbb{1}_{[-a,a]} + \beta\mathbb{1}_{[-b,b]}$ with $b > a > 0$ (see Figure 2.2 for $a = 1$ $b = 2$, $\alpha = 0.9$ and $\beta = 0.1$). We cannot find a and b from $\hat{f}$, except in a rather complicated way. In addition, the two Fourier transforms presented in Figures 2.1 and 2.2 are very similar, although the second function has two discontinuities more (in $\pm b$) than the pulse function. Merely looking at the Fourier transform does not make it possible to deter the position and the number of discontinuities.

This example shows that we cannot locate the discontinuities, the changes of regularity of a function f in view of $\hat{f}$. The integration over $\mathbb{R}$ makes a kind of averaging which masks the discontinuities.

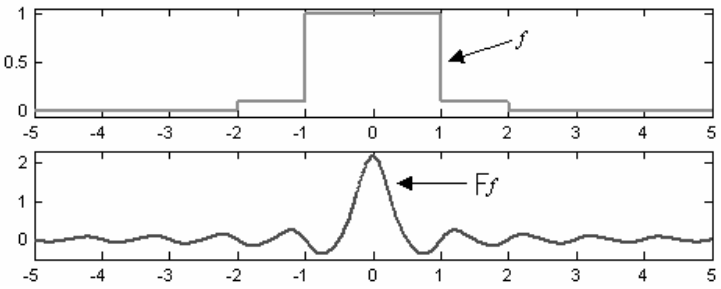


Figure 2.2. *The Fourier transform of f for a = 1 and b = 2*

Disadvantage 2. Non-causality of the Fourier transform

The calculation of $\hat{f}$ requires the knowledge of f over $\mathbb{R}$. A "progressive" calculation of the transform and, thus, a real-time analysis is impossible. Indeed, we cannot even approximately know the spectrum $\hat{f}$ of a signal f whose future we don't know. Figure 2.3 illustrates this point and presents the functions f, $\left|\hat{f}\right|$, g and $|\hat{g}|$, with $f = \mathbb{1}_{[-a,a]}$ and $g = \mathbb{1}_{[-a,0]} - \mathbb{1}_{]0,a]}$ for a = 1. It is clear that, although f and g coincide on $\mathbb{R}^-$, their transforms are very different.

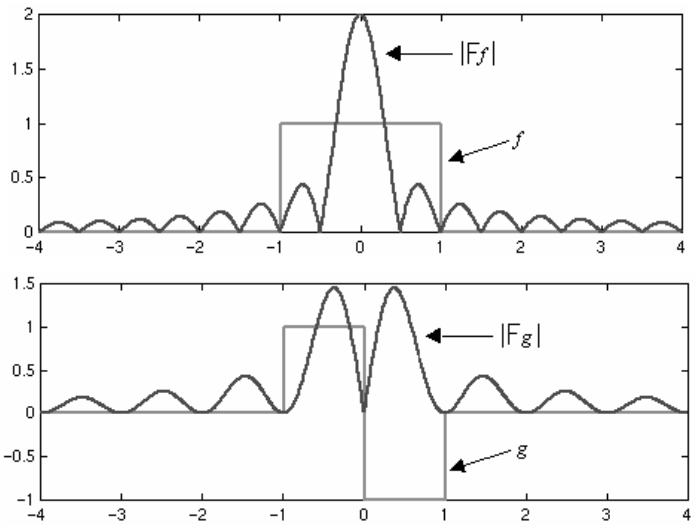


Figure 2.3. *Functions f, $\left|\hat{f}\right|$ (noted $\left|\mathbf{F}f\right|$ on the graph), g and $|\hat{g}|$ for a = 1*

Disadvantage 3. Heisenberg uncertainty principle

If the support of f is "small", then the support of $\hat{f}$ is "large" and vice versa. For example, Figure 2.4 presents f and $\hat{f}$ when $f = \mathbb{1}_{[-a,a]}$ for two values of *a*: at the top the support of f is "small" and that of $\hat{f}$ is "large".

Two limit cases are enlightening: if f is concentrated in point 0, its transform $\hat{f}$ is equal to 1 everywhere. Conversely, if f is a signal that is not localized in time, equal to 1 everywhere, $\hat{f}$ is concentrated in 0. Using the framework of temporal distributions this is expressed by: $\mathbf{F}\delta = 1$ and $\mathbf{F}1 = \delta$, where δ is a Dirac function.

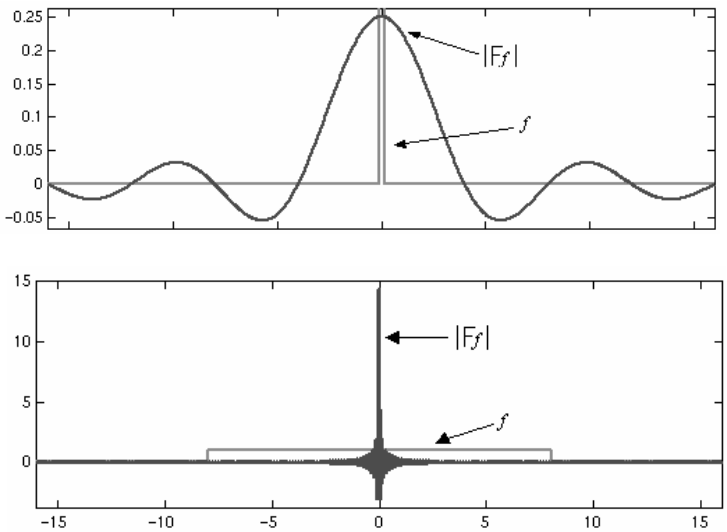


Figure 2.4. *Functions f and $\hat{f}$ for a = 1/8 and a = 8*

The localizations of f and of $\hat{f}$ are linked by the Heisenberg uncertainty principle that specifies an inequality concerning the dispersions of f and $\hat{f}$. It constrains the product of dispersions in time (σ_f) and frequency ($\sigma_{\hat{f}}$) by:

$$\sigma_f^2 \sigma_{\hat{f}}^2 \geq \frac{1}{4}$$

where these variances are defined by:

$$\sigma_f^2 = \frac{1}{\|f\|^2}\int_{\mathbb{R}}(t-u)^2\,|f(t)|^2\,dt \;,\; \sigma_{\hat{f}}^2 = \frac{1}{\|f\|^2}\int_{\mathbb{R}}(\omega-\xi)^2\left|\hat{f}(\omega)\right|^2 d\omega$$

They quantify the dispersions of $|f|^2$ and $\left|\hat{f}\right|^2$ around their respective mean values u and ξ given by:

$$u = \frac{1}{\|f\|^2}\int_{\mathbb{R}} t\,|f(t)|^2\,dt \quad \text{and} \quad \xi = \frac{1}{\|f\|^2}\int_{\mathbb{R}} \omega\left|\hat{f}(\omega)\right|^2 d\omega$$

This property implies that the inverse Fourier transform can be numerically unstable since useful information to reconstruct f from $\hat{f}$ with the synthesis formula may be located in the very high frequency domain. In particular this happens if f has compact support and is irregular. To reconstruct if we need to use $\hat{f}(\omega)$ for large values of ω.

We may summarize saying that the Fourier transform is an integral transformation of a global nature.

2.2.2. *The Gabor transform*

In order to overcome the disadvantage of the global nature of the Fourier transform, an idea consists of localizing the analysis by selecting a portion of the signal around a time position, conducting the Fourier analysis and then starting again for all the possible positions. It is the principle of the sliding window Fourier transform, also called the Gabor transform.

First of all, we take a window $\mathrm{w} \in L^1 \cap L^2$ centered in 0, with $|\hat{\mathrm{w}}|$ being even and of energy 1, used to localize the analysis in time. We note $\mathrm{w}_{\omega,b}(t) = \mathrm{w}(t-b)e^{+2i\pi\omega t}$ for $\omega, t, b \in \mathbb{R}$. The continuous Gabor transform of a signal f is defined by the following formula:

$$\textit{Analysis}:\ \mathbf{G}f(\omega,b) = \int_{\mathbb{R}} f(t)\,\overline{\mathrm{w}_{\omega,b}(t)}\,dt \qquad \omega,b \in \mathbb{R}$$

As for the Fourier transform, the Gabor transform is linear, bijective, continuous and preserves the angles and the lengths (scalar products and norms). The synthesis (or reconstruction) formula is:

$$\textit{Synthesis}:\ f(t) = \int_{\mathbb{R}^2} \left(\mathbf{G}f\right)(\omega,b)\ \mathrm{w}_{\omega,b}(t)\ d\omega\, db \ \text{ in } L^2 \qquad t \in \mathbb{R}$$

The Gabor transform is a Fourier transform local in time, since for each value of b we calculate the Fourier transform of $f(t)\,\overline{\mathrm{w}(t-b)}$. Indeed:

$$\mathbf{G}f(\omega,b) = \int_{\mathbb{R}} f(t)\,\overline{\mathrm{w}_{\omega,b}(t)}\,dt = \mathbf{F}\left[f(t)\,\overline{\mathrm{w}(t-b)}\right](\omega)$$

The window w thus restrains the analysis to a domain around the position b. If, for example, the window is localized on a segment as in the case of the square function $\mathrm{w} = \frac{1}{\sqrt{2a}}\mathbb{1}_{[-a,a]}$, the value of $\mathbf{G}f(\omega,b)$ for fixed b depends only on the values of f on the segment centered in b : $[b-a, b+a]$.

The Gabor transform is a time-frequency analysis. The L^2 scalar product in can be written according to time or frequency:

$$\mathbf{G}f(\omega,b) = \left(f, \mathrm{w}_{\omega,b}\right)_{L^2} = \left(\mathbf{F}f, \mathbf{F}\mathrm{w}_{\omega,b}\right)_{L^2} \quad \text{wherefrom:}$$

$$\mathbf{G}f(\omega,b) = \int_{\mathbb{R}} \hat{f}(\xi)\,\overline{\hat{\mathrm{w}}_{\omega,b}(\xi)}\,d\xi = \int_{\mathbb{R}} \hat{f}(\xi)\,\overline{\hat{\mathrm{w}}(\xi-\omega)}\,e^{+2i\pi b(\xi-\omega)}\,d\xi$$

What happens if the window is $\mathrm{w} = e^{-\pi t^2}$, which is its own Fourier transform? The function $\mathrm{w}(t-b)$ is localized in the vicinity of b, $\mathbf{G}f(\omega,b)$ therefore contains information about f in the vicinity of the position (time) b. Like $\mathbf{F}\mathrm{w} = \mathrm{w}$, $\mathbf{G}f(\omega,b)$ also has information on $\hat{f}$ in the vicinity of the frequency ω.

The *atoms* $\left\{\mathrm{w}_{\omega,b}\right\}_{\omega,b\in\mathbb{R}}$ of this transform are sometimes called Gabor wavelets. They are complex exponentials, as for the Fourier transform, but attenuated by the window w positioned in b. The latter is zero or essentially zero (i.e. very quickly decreasing) apart from an interval centered in 0. It localizes the analyzed function in

this interval. For the Gaussian window, Figure 2.5 represents three atoms defined by:

$$\mathrm{w}_{\omega,b}(t) = e^{-\pi(t-b)^2} e^{+2i\pi\omega t} \text{ with } b = 0 \text{ and } \omega = 1, 2 \text{ and } 5$$

In the three cases the figure shows the envelope ($E = \pm e^{-\pi t^2}$) and the real part of $\mathrm{w}_{\omega,b}$.

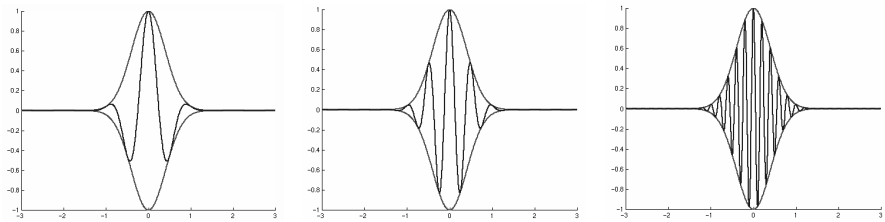


Figure 2.5. *Envelope and real part of the Gabor wavelets for* $\omega = 1, 2$ *and* 5

The number of oscillations increases with the frequency ω but the envelope is rigid and therefore the temporal resolution remains fixed. It is the major defect of this transform.

2.3. The continuous transform in wavelets

In order to overcome the disadvantage of the Gabor transform, it is necessary to seek a transform allowing a similar analysis but able to act in a whole range of temporal resolutions simultaneously. In a certain sense this is achieved by the wavelet transform.

We call a wavelet (or "mother" wavelet) a function $\psi \in L^1 \cap L^2$, admitting $n+1$ zero moments (where $n \in \mathbb{N}$), i.e. verifying:

$$\int_{\mathbb{R}} t^p \psi(t)\, dt = 0 \quad p = 0, \cdots, n$$

The function ψ has a zero integral. It is also orthogonal to the polynomials of a degree lower or equal to n . The function ψ oscillates, taking positive and negative

values. The number n controls the oscillations of ψ, in the sense that the larger n is, the more ψ oscillates.

By translation and dilation of the wavelet ψ we define the atoms of the wavelet transform. For any *scale* $a \in \mathbb{R}^{+*}$ and any *position* $b \in \mathbb{R}$ we define an atom of the transform by:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

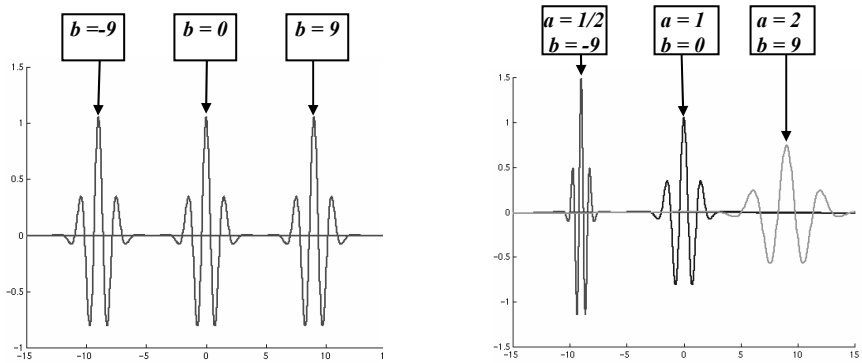


Figure 2.6. *Translation and translation-dilation of a wavelet*

The family $\left\{\psi_{a,b}\right\}_{a,b}$ is the family of wavelets associated with ψ. Taking ψ with an energy of 1 ($\|\psi\|_{L^2} = 1$) all the functions $\psi_{a,b}$ then have a norm of 1.

Wavelets no longer have a rigid envelope, contrary to the atoms of the Gabor transform. They exhibit an accordion behavior: they keep the same form and the same number of oscillations and are translated-dilated versions of the same function. In the left part of Figure 2.6 we see three wavelets $\psi_{1,-9}$, $\psi_{1,0}$ and $\psi_{1,9}$, obtained by translation of the mother wavelet and on the right we have $\psi_{1/2,-9}$, $\psi_{1,0}$ and $\psi_{2,9}$ obtained by translation-dilation of ψ.

The continuous wavelet transform of the finite energy function of f is the family of coefficients $C_f(a,b)$ defined by:

$$\textit{Analysis}:\ C_f(a,b) = \int_{\mathbb{R}} f(t)\,\overline{\psi_{a,b}(t)}\,dt = \left(f, \psi_{a,b}\right)_{L^2} \qquad a \in \mathbb{R}^{+*},\ b \in \mathbb{R}$$

The transformation admits an inverse, under an additional condition known as admissibility (see the end of this section) and the synthesis (or reconstruction) formula is:

$$\textit{Synthesis}:\ f(t) = \frac{1}{K_\psi} \int_{]0,+\infty[\times\mathbb{R}} C_f(a,b)\,\psi_{a,b}(t)\,\frac{da\,db}{a^2} \ \text{ in } L^2 \qquad t \in \mathbb{R}$$

In a certain way $C_f(a,b)$, the coefficient of f on the wavelet $\psi_{a,b}$, characterizes the "fluctuations" of the function f around the position *b*, on scale *a*. Let us suppose that ψ is nil outside of $[-M, M]$, then $\psi_{a,b}$ is nil outside of the interval $[-Ma + b, Ma + b]$. The value of $C_f(a,b)$ then only depends on the values on f around *b* in a segment whose length is proportional to *a*. Let us illustrate this idea in Figure 2.7.

The first graph presents the analyzed function $f(b)$ and the support of three wavelets positioned around b_0, b_1 and b_2 for a fixed scale a. The function is continuous except in b_0. The second graph contains the wavelet coefficients of $C_f(a,b)$ for the same scale a. Around b_1 and b_2 the coefficients are zero, since the function f is constant on the supports of ψ_{a,b_1} and of ψ_{a,b_2} yielding coefficients equal to the integral of the wavelets. On the other hand, due to discontinuity, around b_0 the coefficients are non-zero in a zone with a size proportional to the support of ψ and to the value of a. Thus, inspecting the wavelet coefficients we may deduce the presence of a singularity in the signal analyzed around b_0 and at scale a.

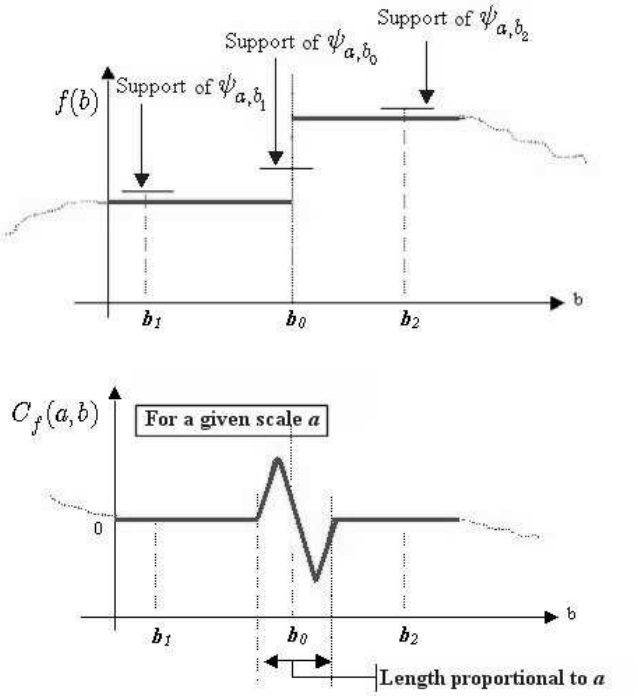


Figure 2.7. *Diagram representation of the detection of a singularity*

Many examples are examined in Chapter 6.

The inversibility of the continuous wavelet transform in L^2 supplemented by the fact that it preserves the angles and lengths (scalar products and norms), is ensured by:

THEOREM 2.2.

Let $\psi \in L^1 \cap L^2$ *verify the following admissibility condition:*

$$\int_0^{+\infty} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} d\omega = \int_{-\infty}^{0} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} d\omega < +\infty$$

Let K_ψ *be the common value of the integrals.*

Then the scalar product is preserved:

$$(f,g)_{L^2} = \frac{1}{K_\psi}\int_{]0,+\infty[\times\mathbb{R}} C_f(a,b)\,\overline{C_g(a,b)}\,\frac{da\,db}{a^2}$$

and the synthesis formula:

$$f(t) = \frac{1}{K_\psi}\int_{]0,+\infty[\times\mathbb{R}} C_f(a,b)\,\psi_{a,b}(t)\,\frac{da\,db}{a^2} \text{ in } L^2 \qquad t\in\mathbb{R}$$

■

Let us note that the condition of admissibility implies, in particular, that $\hat{\psi}(0) = 0$ and therefore $\int_{\mathbb{R}} \psi(t)\,dt = 0$.

This condition is difficult to use; often instead of it we prefer a sufficient condition of admissibility that is much simpler to verify:

$$\psi \text{ real } \psi\in L^1\cap L^2,\ t\psi(t)\in L^1 \text{ and } \int_{\mathbb{R}}\psi(t)dt = 0$$

The transform C_f associates with a function f of a real variable t, an infinite number of coefficients doubly indexed by $a\in\mathbb{R}^{+*}$ and $b\in\mathbb{R}$. From a certain point of view, the transformation goes too far: it is redundant and is sometimes desirable to avoid this redundancy. We introduce the discrete transform which in certain cases achieves this goal.

2.4. Orthonormal wavelet bases

2.4.1. *From continuous to discrete transform*

It is legitimate to wonder whether it is necessary to know C_f everywhere on $\mathbb{R}^{+*}\times\mathbb{R}$ to reconstruct f. When the answer is negative, the use of a discrete subset seems a reasonable objective. The idea is as follows: we consider discrete subsets of $\mathbb{R}^{+*}$ and $\mathbb{R}$. Let us fix $a_0 > 1$, $b_0 > 0$ and take $a\in\{a_0^p\}_{p\in\mathbb{Z}}$ and $b\in\{na_0^p b_0\}_{p,n\in\mathbb{Z}}$. Instead of using the family of wavelets:

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

for the discrete transform we use the family of wavelets indexed by $\mathbb{Z}$:

$$\psi_{p,n}(t) = a_0^{-p/2}\,\psi(a_0^{-p}t - nb_0) \qquad a_0 > 1\ ,\ b_0 > 0 \text{ fixed and } p,n \in \mathbb{Z}$$

For $f \in L^2$ we define the discrete wavelet transform of the function f by:

$$C_f(p,n) = \int_{\mathbb{R}} f(t)\,\overline{\psi_{p,n}(t)}\,dt = \left(f, \psi_{p,n}\right)_{L^2} \qquad p,n \in \mathbb{Z}$$

In the two preceding formulae and hereafter we change notations in order to simplify, in the discrete case, the writing of the atoms and coefficients.

The usual choice $a_0 = 2$ and $b_0 = 1$ is dictated by Shannon's sampling theorem (see [MAL 98] p. 41). It is then natural to tackle a more difficult question: does there exist, and under which conditions, a function ψ such that the family $\left\{\psi_{j,k}\right\}_{(j,k)\in\mathbb{Z}^2}$ where $\psi_{j,k}(t) = 2^{-j/2}\,\psi\left(2^{-j}t - k\right)$ is an orthonormal base of $L^2(\mathbb{R})$? The answer is closely related to the concept of multi-resolution analysis.

2.4.2. *Multi-resolution analysis and orthonormal wavelet bases*

A multi-resolution analysis of $L^2(\mathbb{R})$ is a family $M = \left\{V_j\right\}_{j\in\mathbb{Z}}$ of embedded vectorial subspaces with the properties [2.1] to [2.5] below that we can group in three blocks:

– $\left\{V_j\right\}_{j\in\mathbb{Z}}$ is a set of approximation spaces, i.e.:

V_j is a closed subspace of L^2 [2.1]

$$V_j \subset V_{j-1} \qquad [2.2]$$

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2 \text{ and } \bigcap_{j\in\mathbb{Z}} V_j = \{0\} \qquad [2.3]$$

Property [2.1] ensures the existence of the orthogonal projection of f on each space V_j, a projection that approaches f; [2.2] is the decreasing property of spaces and the improvement of the approximation when j decrease; [2.3] ensures that the $\{V_j\}$ sequence converges towards the entire L^2 and, thus, that the sequence of projections converges towards f;

– the V_j spaces are obtained by dyadic dilation or contraction of the functions of the single space (for example V_0):

$$\forall j \in \mathbb{Z}\,,\ \ v(t) \in V_j \Leftrightarrow v(2t) \in V_{j-1} \qquad [2.4]$$

This property characterizes the multi-resolution aspects of the M sequence and plays a crucial part in the construction of wavelet bases;

– a last property relates to the translation of functions. It supposes the existence of a function, which makes it possible to build a base of V_0 by integer translation:

$$\exists g \in V_0 \text{ such that } \{g(t-k)\}_{k\in\mathbb{Z}} \text{ is a Riesz base of } V_0 \qquad [2.5]$$

In order to supplement [2.5], let us specify what a Riesz base of L^2 is. The family $\{e_k\}_{k\in\mathbb{Z}} \subset L^2$ is a *Riesz base* of L^2 if:

$$\forall h \in L^2\,,\ \exists!\,\alpha \in l^2(\mathbb{Z}) \text{ such that } h = \sum_{k\in\mathbb{Z}} \alpha_k e_k \text{ in } L^2 \qquad [2.6]$$

There exist $0 < A \leq B < +\infty$ such that for all $h \in L^2$ we have:

$$A\left\|\alpha\right\|_{l^2} \leq \left\|h\right\|_{L^2} \leq B\left\|\alpha\right\|_{l^2} \quad \text{where } \left\|\alpha\right\|_{l^2} = \left\{\sum_{k\in\mathbb{Z}} \left|\alpha_k\right|^2\right\}^{\frac{1}{2}} \qquad [2.7]$$

A Riesz base is thus a generating, free system and, in a certain way, property [2.7] controls the angles between the basic vectors. In particular, in the case of a orthonormal base (Hilbertian base) we have $A = B = 1$ and [2.7] is then simply the Parseval equality. Choosing a Riesz base of L^2 is equivalent to choosing an isomorphism between the space of functions $L^2(\mathbb{R})$ and the space of $l^2(\mathbb{Z})$ sequences.

■

On the basis of the M family we define a second family of subspaces noted $\{W_j\}$, where W_j is the orthogonal complement of V_j in V_{j-1}:

$$V_{j-1} = V_j \oplus W_j \quad \text{with} \quad W_j \perp V_j$$

As opposed to $\{V_j\}$ spaces, which are approximation spaces, we shall say that $\{W_j\}$ spaces are detail spaces.

We obtain a series of properties for the $\{W_j\}_{j\in\mathbb{Z}}$ subspaces, which are useful for the geometrical understanding of the construction:

$$w(t) \in W_j \Leftrightarrow w(2t) \in W_{j-1} \qquad [2.8]$$

$$W_j \perp W_k \quad j \neq k \qquad [2.9]$$

$$W_j \perp V_k \quad j \leq k \qquad [2.10]$$

$$V_J = V_K \oplus W_K \oplus \cdots \oplus W_{J+1} \quad J < K \qquad [2.11]$$

$$V_J = \bigoplus_{j=J+1}^{+\infty} W_j \qquad [2.12]$$

$$L^2(\mathbb{R}) = V_J \oplus \left\{ \bigoplus_{j=-\infty}^{J} W_j \right\} \qquad [2.13]$$

$$L^2(\mathbb{R}) = \bigoplus_{j=-\infty}^{+\infty} W_j \qquad [2.14]$$

Let us comment on some of these properties. For example, [2.13] indicates that an element of L^2 can be written in the form of an orthogonal sum of a rough approximation and an infinite number of finer details. Property [2.14] in turn expresses the fact that any function of L^2 is an infinite sum of orthogonal details.

Let us note $A^j = P_{V_j} f$ and $D^j = P_{W_j} f$, orthogonal projections of $f \in L^2$ on spaces V_j and W_j respectively. We then have $A^{j-1} = A^j + D^j$ with $A^j \perp D^j$.

Spaces $\left\{V_j\right\}$ are approximation spaces in the following sense: A^j converges to f in $L^2(\mathbb{R})$ when j tends to $-\infty$; in the same way, spaces $\left\{W_j\right\}$ are detail spaces in the sense that in L^2 we have, on the one hand, D^j which converges to 0 when j tends to $-\infty$ and, on the other hand, $f = A^J + \sum_{-\infty}^{J} D^j$. In other words, for a fixed level of approximation J, the D^j are the corrections to be added to the approximation to find f.

Now let us state the fundamental result associated with multi-resolution analysis, noting $f_{j,k}(t) = 2^{-j/2} f\left(2^{-j}t - k\right)$ for any function f.

THEOREM 2.3.– ORTHONORMAL WAVELET BASES

Let M be a multi-resolution analysis of $L^2(\mathbb{R})$. Starting from g (see [2.5]), we can build a scaling function φ then a wavelet ψ such that:

$\forall J \in \mathbb{Z}, \left\{\left\{\varphi_{J,k}\right\}_{k\in\mathbb{Z}}, \left\{\psi_{j,k}\right\}_{j,k\in\mathbb{Z}, j\le J}\right\}$ *is an orthonormal base of L^2 and* $\left\{\psi_{j,k}\right\}_{j,k\in\mathbb{Z}}$ *is an orthonormal wavelet base of L^2.*

■

The principle of the proof of this theorem is as follows:

– starting from g and, thus, from $\{g(t-k)\}_{k\in\mathbb{Z}}$, construct a function φ such that $\{\varphi(t-k)\}_{k\in\mathbb{Z}}$ is an orthonormal base of V_0;

– deduce from it that $\left\{\varphi_{j,k}\right\}_{k\in\mathbb{Z}}$ is an orthonormal base of V_j;

– using φ, construct a function ψ such that $\{\psi(t-k)\}_{k\in\mathbb{Z}}$ is a orthonormal base of W_0;

– deduce from it that $\left\{\psi_{j,k}\right\}_{k\in\mathbb{Z}}$ is an orthonormal base of W_j;

– conclude from it that $\left\{\psi_{j,k}\right\}_{j,k\in\mathbb{Z}}$ is an orthonormal base of L^2.

The delicate points are stages 1 and 3. They are the subject of the two proposals stated later. Moreover, they present more completely the properties of the scaling function φ and the wavelet ψ. Before stating them and commenting on them, let us make two observations.

NOTE 2.2.

If M is a multi-resolution analysis there is an infinity of functions of scale and, thus, an associated infinity of wavelets leading to the same analysis.

In addition, there are orthonormal wavelet bases of L^2, i.e. orthonormal bases having the form of $\left\{\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right)\right\}_{j,k\in\mathbb{Z}}$, which are not associated with a multi-resolution analysis (see a counterexample in [DAU 92] p. 136). They are, thus, not associated with a scaling function. On the other hand, once ψ is sufficiently regular, there is an underlying multi-resolution analysis. ■

Some key elements to be memorized are summarized in Table 2.1.

	Functions	**Spaces**	**Bases**	$j \nearrow$	$j \searrow$
Approximations	Scaling function φ	V_j	$\left\{\varphi_{j,k}\right\}_{k\in\mathbb{Z}}$	Coarser	Finer
Details	Wavelet ψ	W_j	$\left\{\psi_{j,k}\right\}_{k\in\mathbb{Z}}$		

Table 2.1. *Key elements of multi-resolution analysis*

2.4.3. *The scaling function and the wavelet*

In this section, we state and comment on two proposals, which establish the links between the concepts of multi-resolution analysis and orthogonal wavelet and propose a manner of building the second starting from the first. This construction also shows the fundamental part played by the two-scale equations in the time and frequency domains. Let us start with the construction of the scaling function φ.

PROPOSAL 2.1.– CONSTRUCTION OF THE SCALING FUNCTION

Let us consider the scaling function φ defined using its Fourier transform $\hat{\varphi}$ by:

$$\hat{\varphi}(\omega) = \frac{\hat{g}(\omega)}{\left(\sum_{k\in\mathbb{Z}} |\hat{g}(\omega+k)|^2 \right)^{1/2}} \qquad [2.15]$$

Then:

– $\varphi \in V_0$;

– $\left\{\varphi_{0,k} = \varphi(t-k)\right\}_{k\in\mathbb{Z}}$ is an orthonormal base of V_0;

– two-scale equation for φ:

$\exists! a = \left\{a_k\right\}_{k\in\mathbb{Z}}$, $a \in l^2(\mathbb{Z})$ such that:

$$\frac{1}{2}\varphi\left(\frac{t}{2}\right) = \sum_{k\in\mathbb{Z}} a_k \varphi(t-k) \text{ in } L^2 \qquad [2.16]$$

– $m_0(\omega) = \sum_{k\in\mathbb{Z}} a_k e^{-2i\pi k\omega}$ is periodic with period 1, $m_0 \in L^2(0,1)$ and verifies:

$$\hat{\varphi}(2\omega) = m_0(\omega)\,\hat{\varphi}(\omega) \quad p.p.\,\omega \in \mathbb{R} \qquad [2.17]$$

$$\left|m_0(\omega)\right|^2 + \left|m_0\left(\omega+\tfrac{1}{2}\right)\right|^2 = 1 \ , \ p.p.\,\omega \in \mathbb{R} \qquad [2.18]$$

– more generally, $\forall j \in \mathbb{Z}$, $\left\{\varphi_{j,k} = 2^{-\frac{j}{2}}\varphi(2^{-j}t-k)\right\}_{k\in\mathbb{Z}}$ is an orthonormal base of V_j.

■

Let us comment on this result:

– relation [2.15] defines the scaling function φ starting from g in the frequency domain and leads to an orthonormalization in the time domain;

– the first two properties show that it is a change of basis in the space V_0;

– the third property in turn results from $\varphi\left(\frac{t}{2}\right) \in V_1$, from the inclusion of $V_1 \subset V_0$ and the fact that $\{\varphi(t-k)\}_{k\in\mathbb{Z}}$ is a base of V_0. It can be read differently: the sequence $\left\{a_k\right\}_{k\in\mathbb{Z}}$ being given, [2.16] is a functional equation, of which φ is the solution;

– relation [2.17] is the counterpart in the frequency domain of the two-scale equation. It reveals m_0, the discrete Fourier transform of the sequence a;

– relation [2.18] is the frequency translation of the orthogonality of the base $\{\varphi(t-k)\}_{k\in\mathbb{Z}}$ of V_0.

Let us now pass to the construction of the wavelet.

PROPOSAL 2.2.– CONSTRUCTION OF THE WAVELET

Wavelet ψ is defined using its Fourier transform $\hat{\psi}$. Let ρ be a periodic function with a period of $\frac{1}{2}$, $|\rho(\omega)| = 1$ for almost all $\omega \in \mathbb{R}$, and let us pose $m_1(\omega) = \rho(\omega)e^{-2i\pi\omega}\overline{m_0(\omega+\frac{1}{2})}$ and define:

$$\widehat{\psi}(\omega) = m_1\left(\frac{\omega}{2}\right)\widehat{\varphi}\left(\frac{\omega}{2}\right) \qquad [2.19]$$

Then:

– $\psi \in W_0$;

– $\left\{\psi_{0,k} = \psi(t-k)\right\}_{k\in\mathbb{Z}}$ is an orthonormal base of W_0;

– two-scale equation for ψ:

$$\exists! b = \left\{b_k\right\}_{k\in\mathbb{Z}} \; b \in l^2(\mathbb{Z}) \text{ such that } m_1(\omega) = \sum_{k\in\mathbb{Z}} b_k e^{-2i\pi k\omega}, \text{ and:}$$

$$\frac{1}{2}\psi\left(\frac{t}{2}\right) = \sum_{k\in\mathbb{Z}} b_k\,\varphi(t-k) \text{ in } L^2\,; \qquad [2.20]$$

– m_1 is periodic with a period of 1, $m_1 \in L^2(0,1)$ and:

$$\left|m_1(\omega)\right|^2 + \left|m_1\left(\omega+\tfrac{1}{2}\right)\right|^2 = 1 \ , \ p.p.\omega \in \mathbb{R} \qquad [2.21]$$

$$m_0(\omega)\overline{m_1(\omega)} + m_0\left(\omega+\tfrac{1}{2}\right)\overline{m_1\left(\omega+\tfrac{1}{2}\right)} = 0 \ , \text{ for almost all } \omega \in \mathbb{R} \quad [2.22]$$

– more generally, $\forall j \in \mathbb{Z}$, $\left\{\psi_{j,k} = 2^{-j/2}\,\psi\left(2^{-j}t-k\right)\right\}_{k\in\mathbb{Z}}$ is an orthonormal base of W_j;

– $\left\{\psi_{j,k}\right\}_{j,k\in\mathbb{Z}}$ is an orthonormal base of $L^2(\mathbb{R})$.

■

Let us comment on this result:

– relation [2.19] defines the wavelet starting from the scaling function φ in the frequency domain using a filter m_1 deduced from m_0;

– the first two properties affirm that we are producing a function of the detail space W_0 and considering all the integer translations $\{\psi(t-k)\}_{k\in\mathbb{Z}}$ an orthonormal base of W_0;

– the third property in turn results from $\psi\left(\frac{t}{2}\right) \in W_1$, from the inclusion $W_1 \subset V_0$ and the fact that $\{\varphi(t-k)\}_{k\in\mathbb{Z}}$ is a base of V_0, leading to a second two-scale equation defining the wavelet. This relation is none other than the counterpart in the time domain of relation [2.19] which defines m_1.

– relation [2.21] is the frequency translation of the orthogonality of the base $\{\psi(t-k)\}_{k\in\mathbb{Z}}$ of W_0. Relation [2.22] in turn expresses the orthogonality between spaces V_0 and W_0.

A last fundamental comment to conclude this section, relates to the filters m_0 and m_1. From relations [2.18], [2.19], [2.21] and [2.22] we deduce that:

$$\begin{aligned} &\left|m_0(\omega)\right|^2 + \left|m_1(\omega)\right|^2 = 1 \\ &m_0(\omega)\overline{m_0(\omega+\tfrac{1}{2})} + m_1(s)\overline{m_1(\omega+\tfrac{1}{2})} = 0 \end{aligned}$$

These two relations translate the conditions of perfect reconstruction of the banks of underlying filters. They make it possible to consider the construction of wavelets not using g but the filters. This will be exploited, in particular, in Chapter 5.

2.5. Wavelet packets

Wavelet packets are a generalization of orthogonal wavelets. They allow a finer analysis by breaking up detail spaces, which are never decomposed in the case of wavelets.

They were introduced by Coifman and Wickerhauser [COI 92] at the beginning of the 1990s in order to mitigate the lack of frequency resolution of wavelet analysis. The principle is to some extent to cut up detail spaces into frequency sections.

2.5.1. *Construction of wavelet packets*

Wavelet packets are generated by recurrence. We start with two filters with lengths of N, g_n and h_n associated with the orthogonal wavelet with compact support ψ and scaling function φ issued from an MRA of L^2. They are obtained on the basis of the sequences a and b from the formulae [2.16] and [2.20], so that their l^2 norm equals 1.

By induction we define the sequence of functions $(w_n)_{n\in\mathbb{N}}$ starting with $w_0 = \varphi$ by:

$$\begin{cases} w_{2n}(t) = \sqrt{2}\sum_{k=0}^{2N-1} h_k w_n(2t-k) \\ w_{2n+1}(t) = \sqrt{2}\sum_{k=0}^{2N-1} g_k w_n(2t-k) \end{cases} \qquad [2.23]$$

Equations [2.23] for $n = 0$ are reduced simply to two two-scale equations and we have $w_0 = \varphi$, the scaling function and subsequently $w_1 = \psi$, the wavelet. We thus see how packages generalize wavelets.

For the Haar wavelet we have $N = 1$, $h_0 = h_1 = \frac{1}{\sqrt{2}}$ and $g_0 = -g_1 = \frac{1}{\sqrt{2}}$. Equations [2.23] then become simply:

$$\begin{cases} w_{2n}(t) = w_n(2t) + w_n(2t-1) \\ w_{2n+1}(t) = w_n(2t) - w_n(2t-1) \end{cases}$$

w_0 is the Haar scaling function and w_1 is the Haar wavelet (see Chapter 4), both with support in $[0,1]$. Then we obtain w_{2n} by adding two copies of w_n contracted by a factor of $\frac{1}{2}$ with distinct supports $\left[0,\frac{1}{2}\right]$ for $w_n(2t)$ and $\left[\frac{1}{2},1\right]$ for $w_n(2t-1)$. We obtain w_{2n+1} by subtracting the same versions. For $n = 0$ to 7 we obtain the functions w_n seen in Figure 2.8.

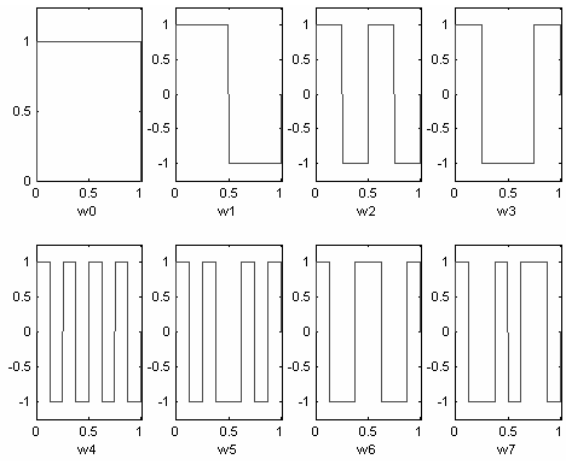


Figure 2.8. *Haar wavelet packets*

More generally, on the basis of a more regular orthogonal wavelet we obtain smoothed versions of this system of functions, all with support contained within $[0,2N-1]$. Figure 2.9 presents the functions obtained when the original wavelet is *db2* (see Chapter 4).

The doubly indexed set $\{w_n(t-k)\}_{n\in\mathbb{N},k\in\mathbb{Z}}$ is an orthonormal base of $L^2(\mathbb{R})$.

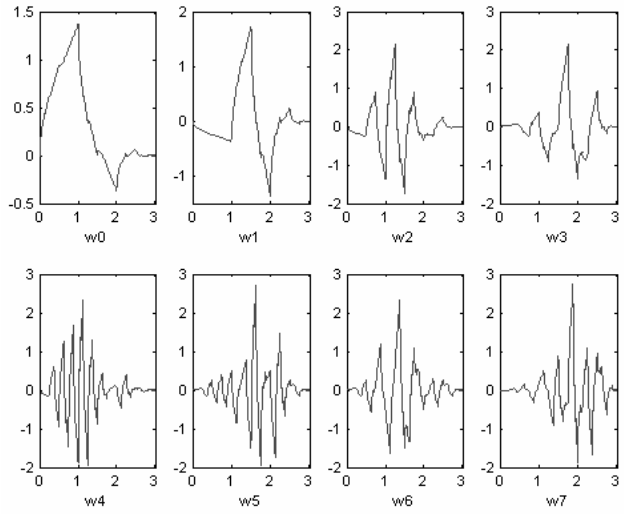


Figure 2.9. *Wavelet packets of db2*

2.5.2. *Atoms of wavelet packets*

The atoms of wavelet analysis are constructed by dyadic translation and dilation of the functions φ and ψ. For wavelet packets we proceed in a similar fashion on the basis of the functions $(w_n\,; n\in\mathbb{N})$ and considering atoms indexed by three indices:

$$(w_n)_{j,k}(t) = 2^{-j/2} w_n\left(2^{-j}t - k\right) \quad \text{for} \quad n\in\mathbb{N},(j,k)\in\mathbb{Z}^2$$

These atoms are not all useful. For fixed j the useful values of n are $0\le n\le 2^j-1$. As in the case of wavelets, k is the location parameter and j is the scale parameter. What is then the interpretation of n ?

The idea is that for fixed values of j and k, $(w_n)_{j,k}$ analyzes the fluctuations of the signal around the position $k2^j$, scale 2^j and all the frequencies corresponding to the various useful values of the last parameter n.

We may think to link these frequencies to the number of oscillations. By attentively examining the wavelet packets associated with the *db1* wavelet presented previously, the natural order of the functions $(w_n\,;\, n = 0,...,7)$ does not agree completely with the number of oscillations or passages through 0.

To restore the property that the central frequency must grow monotonously with the order it is necessary to define the frequency order of the packages. It is obtained recursively starting from the previous by a permutation of integers noted r (see [MAL 98] p. 327). For the first eight integers the latter yields Table 2.2.

Natural order *n*	0	1	2	3	4	5	6	7
Frequency order *r*(*n*)	0	1	3	2	6	7	5	4

Table 2.2. *Natural and frequency order of wavelet packets*

As we saw in the preceding figures, $w_{r(n)}(t)$ oscillates approximately n times. This re-ordering is useful for signal analysis (see Chapter 6 for examples). Thus, we construct visualizations of the coefficients of wavelet packets in the time-frequency plane.

2.5.3. *Organization of wavelet packets*

The set of functions: $P_{j,n} = \left((w_n)_{j,k}(t); k \in \mathbb{Z}\right)$ constitutes the (j,n) wavelet packet. Wavelet packets are naturally organized as a tree. For example, for a level of decomposition equal to 3, they are organized as indicated in Figure 2.10. For each scale j, which indicates the depth in the tree, the possible values of the frequency parameter n are integers from 0 to $2^j - 1$, which indicates the position in the tree. In other words, the notation $P_{j,n}$ used is consistent with the usual depth-position notation of tree nodes labeling.

Let us note the space generated by the functions of the package $P_{j,n}$ as $\overline{P_{j,n}}$. As for every j, $\overline{P_{j,0}} = V_j$ and $\overline{P_{j,1}} = W_j$, the library of wavelet packet bases contains the wavelets bases (see section 2.4.1) but also much else. Let us say a few words about some of them. Let us note $V_0 = \overline{P_{0,0}}$, then we have:

– $(P_{d,1}\,;\, d \geq 1)$ is an orthogonal base of V_0;

– for $J \geq 0$, $(P_{J,0},(P_{j,1}\,;\, 1 \leq j \leq J))$ is an orthogonal base of V_0.

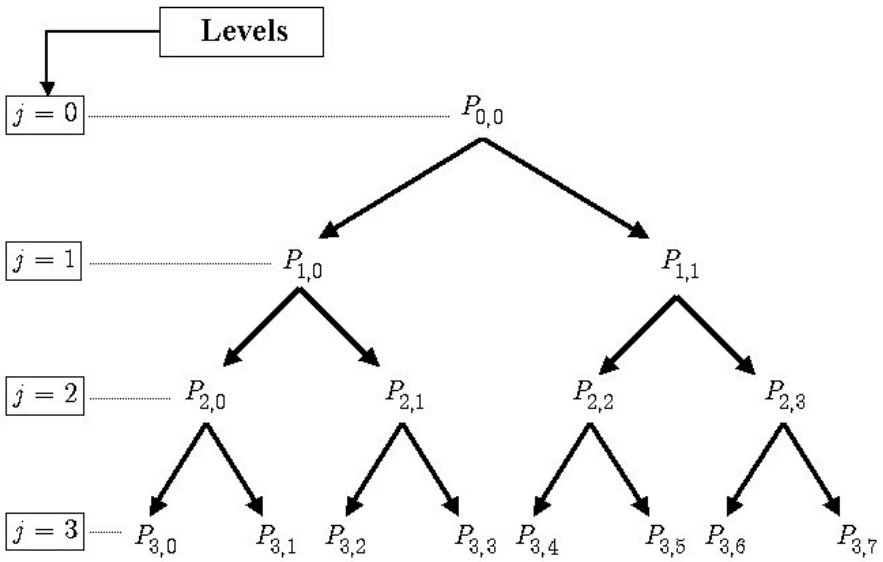


Figure 2.10. *Organization of wavelet packets*

Moreover, $(P_{j+1,2n}, P_{j+1,2n+1})$ is an orthogonal base of $\overline{P_{j,n}}$. This property gives a precise meaning to the split and, consequently, to the merge, which is the inverse process, in the tree of wavelet packets organization. Each developed node has the following form:

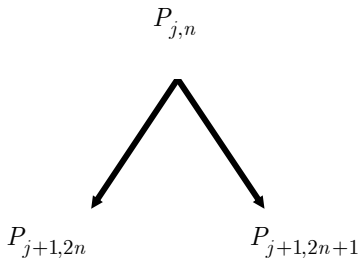


Figure 2.11. *Split and merge in the wavelet packets tree*

Consequently, the set of leaves of any binary sub-tree of a complete binary tree corresponds to an orthogonal wavelet packet basis of the initial space. That is, for a finite energy signal of V_0, each wavelet packet basis offers a specific manner of representation. We can then select the decomposition best adapted to a signal and a given objective [COI 92].

2.6. Biorthogonal wavelet bases

Orthogonal wavelets generate orthonormal bases and constitute easily handled families. However, they are difficult to construct, often irregular and implicitly defined. Slackening the orthogonality constraint makes it possible to:

– facilitate construction and, thus, increase the number of possible forms;

– improve form, symmetry and regularity while more often obtaining explicit formulae for the wavelets.

For example, if we take an interest in the filters associated with wavelets, for many problems the "good" filters are finite impulse response (FIR) filters with linear phase. Indeed, Kahane and Lemarié-Rieusset (in [KAH 98] p. 454), affirm that "it is a general belief in image processing that linear phase filters produce less visual artefacts than others. However, for the wavelet transform, linear phase corresponds to a symmetric scaling function, whereas FIR corresponds to a compact support scaling function". Can we reconcile the two properties within the framework of orthogonal wavelets? The Haar function is the only compactly supported scaling function generating a symmetric orthonormal basis [DAU 92].

In summary, the condition of orthogonality is a very strong constraint. By somewhat slackening this constraint it is possible to build wavelets presenting more attractive properties at the price of introducing a minor difficulty into the calculation. We no longer produce only one wavelet but two wavelets ψ and $\widetilde{\psi}$ in duality, called biorthogonal wavelets. The orthogonal case corresponds to $\psi = \widetilde{\psi}$.

2.6.1. *Orthogonality and biorthogonality*

Let us take a moment to discuss duality from the perspective of duality of bases. In the simple cases, like that of spaces L^2, when we have a, let us say oblique, basis $B = \{e_i\}$ we can nevertheless make calculations as if the basis were orthonormal. We can associate a second basis $\widetilde{B} = \{\tilde{e}_i\}$, whose elements are defined by the relations of duality $\left(\tilde{e}_i, e_j\right) = \delta_{ij}$. Then using the co-ordinates of X in the two bases: $x_i = (X, e_i)$ and $\tilde{x}_i = (X, \tilde{e}_i)$ we write $X = \sum x_i \tilde{e}_i = \sum \tilde{x}_i e_i$. The two bases appear in the decomposition and reconstruction formulae: the signal analysis formula ($x_i = (X, e_i)$) uses the basis B, whereas the synthesis formula ($X = \sum x_i \tilde{e}_i$) uses the dual basis $\widetilde{B}$.

Let us also note that even if B is an oblique basis, calculations are made as if it were orthonormal. Indeed, let us note two vectors X and Y by describing each one in one of the two bases: $X = \sum x_i \tilde{e}_i$ and $Y = \sum \tilde{y}_i e_i$. Then the scalar product is written:

$$(X,Y) = (\sum x_i \tilde{e}_i, \sum \tilde{y}_j e_j) = \sum_i \sum_j x_i \tilde{y}_j (\tilde{e}_i, e_j) = \sum_i \sum_j x_i \tilde{y}_j \delta_{ij} = \sum_i x_i \tilde{y}_i$$

No summation limit appears in these formulae. They are valid in the finite dimension as well as in the infinite dimension ($i, j \in \mathbb{Z}$), similarly to the case of finite energy signals spaces.

It may be the case that the basis and the dual basis are generated by two wavelets, one used for analysis, the other for synthesis. Can we use the acquired freedom intelligently? The answer is yes. It is possible to make rather different requests on ψ and $\tilde{\psi}$. The crucial point stems from the link between the analysis and synthesis wavelet which is rather loose. We can partially uncouple the properties of each of the two bases according to the objectives, and we can largely separate the constraints that we wish to formulate for the analysis and synthesis atoms. For example, in image processing for compression we will prefer a compactly supported analysis wavelet with many zero moments, so that the representation of the image is as sparse as possible. Such a wavelet will be rather asymmetric and irregular – and therefore ill-adapted to reconstruction, for which we wish to use a regular and symmetric wavelet – in order to minimize the visual artefacts. The framework of biorthogonal wavelets makes it possible to meet these two requirements simultaneously.

Complementarily, biorthogonal wavelets can be very regular (C^∞) while preserving a finite support (see, for example, [KAH 98] p. 355). In Chapter 4 we will find some examples of biorthogonal wavelets.

2.6.2. *The duality raises several questions*

In section 2.4.1 we have passed from the continuous wavelet transform to the discrete transform generating the orthogonal wavelet bases. Between these two extreme situations, one of which is very redundant and the other the most parsimonious possible, there exists an intermediate construction. It is the concept of wavelet frame [DAU 92], which will not be further detailed here. Let us note simply that these are systems generating $L^2(\mathbb{R})$ constructed using a wavelet ψ, sweeping the entire pallet from hyper redundancy to the most economic representation.

To every wavelet frame we can always associate a dual frame. In the case of the orthogonal wavelet basis the frame consists of translation-dilation forms of a single function ψ. Of course, the dual frame being identical to the original frame, it is also generated by translation-dilation forms of a function $\tilde{\psi}$ (which is simply ψ). But, in general, there is no reason for the dual frame to be generated by a wavelet $\tilde{\psi}$. Under what condition is it so?

Other questions arise: if ψ is associated with a scaling function φ in a multi-resolution analysis M can we affirm that $\tilde{\psi}$ is also associated with a scaling function $\tilde{\varphi}$ in a multi-resolution analysis $\tilde{M}$? Which links do the two multi-resolution analyses maintain?

The answers to all these questions are not provided in this book; however, the books of Cohen [COH 92c] and Kahane and Lemarié-Rieusset [KAH 98] cover a large number of these aspects.

2.6.3. *Properties of biorthogonal wavelets*

Let us suppose that wavelets are constructed and let us analyze their properties. The two families $M = \left\{V_j\right\}_{j\in\mathbb{Z}}$ and $\widetilde{M} = \left\{\widetilde{V}_j\right\}_{j\in\mathbb{Z}}$ are multi-resolution biorthogonal analyses of $L^2(\mathbb{R})$. They were first introduced in 1990 by Cohen, Daubechies and Fauveau [COH 92b] and are characterized by the property: $L^2(\mathbb{R}) = V_0 \oplus \widetilde{V_0}^{\perp}$.

Let us note by V_j, W_j, $\widetilde{V_j}$ and $\widetilde{W_j}$ the spaces generated respectively by the families of functions $\left\{\varphi_{j,k}\right\}_{k\in\mathbb{Z}}$, $\left\{\psi_{j,k}\right\}_{k\in\mathbb{Z}}$, $\left\{\widetilde{\varphi}_{j,k}\right\}_{k\in\mathbb{Z}}$ and $\left\{\widetilde{\psi}_{j,k}\right\}_{k\in\mathbb{Z}}$. These spaces and these functions verify a set of relations highlighting multi-resolution and biorthogonality properties. Let us start with the first aspect:

– for each family of spaces $\left\{E_j\right\}_{j\in\mathbb{Z}}$ we pass from E_j to E_{j-1} by dilation;

– we have the inclusions:

$$V_j \subset V_{j-1}, W_j \subset V_{j-1}, \widetilde{V}_j \subset \widetilde{V}_{j-1} \quad \text{and} \quad \widetilde{W}_j \subset \widetilde{V}_{j-1}$$

– finally, there are the decompositions:

$$V_j = V_{j+1} \oplus W_{j+1} \text{ and } \widetilde{V}_j = \widetilde{V}_{j+1} \oplus \widetilde{W}_{j+1}$$

but they are not orthogonal.

Let us now pass to the relations of duality. The set of properties below is obtained on the basis of the following result:

$$\left(\varphi_{0,k}, \widetilde{\varphi}_{0,p}\right)_{L^2} = \delta_{k,p} \quad (= 1 \text{ if } k = p\,,\, 0 \text{ if not}) \qquad [2.24]$$

– the couples of spaces (V_j , $\widetilde{V}_j$) and (W_j , $\widetilde{W}_j$) satisfy:

$$\left(\varphi_{j,k}, \widetilde{\varphi}_{j,p}\right)_{L^2} = \delta_{k,p} \quad \text{and} \quad \left(\psi_{j,k}, \widetilde{\psi}_{j,p}\right)_{L^2} = \delta_{k,p}$$

– the couples of spaces (V_j , $\widetilde{W}_j$) and ($\widetilde{V}_j$, W_j) are orthogonal and thus we have:

$$\left(\varphi_{j,k}, \widetilde{\psi}_{j,p}\right)_{L^2} = 0 \quad \text{and} \quad \left(\widetilde{\varphi}_{j,k}, \psi_{j,p}\right)_{L^2} = 0$$

– thanks to inclusions, $V_n \perp \tilde{W}_j$ and $\tilde{V}_n \perp W_j$ for $n \geq j$.

– inclusions also imply that for $n \neq j$ we have biorthogonality relations:

$$\left(\psi_{n,k}, \widetilde{\psi}_{j,p}\right)_{L^2} = \delta_{n,j}\, \delta_{k,p}\,, \text{ wherefrom } W_n \perp \tilde{W}_j \text{ for } n \neq j\,.$$

In this context, usable projections are the oblique projections P_j to V_j parallel to the direction of $(\tilde{V}_j)^{\perp}$ which are written for a signal f :

$$P_j(f) = \sum_{k \in Z} \tilde{c}_k \varphi_{j,k} \text{ where } \tilde{c}_k = (f, \widetilde{\varphi}_{j,k})$$

A question remains: how can such wavelets be constructed? The filters approach proves as effective as in the context of orthogonality and the steps are completely similar. This method is so interesting that it is the subject of a detailed description in Chapter 5 devoted to the construction of wavelets. Here we limit ourselves to a few indications.

First of all, let us note that within the framework of orthogonal wavelets two filters noted m_0 and m_1 (see proposals 2.1 and 2.2) play a crucial part. The first one, m_0 is associated with φ and the second one m_1 is associated with the wavelet ψ. In the biorthogonal case, in addition to this couple, two other filters noted $\widetilde{m}_0$ and $\widetilde{m}_1$ are associated with $\tilde{\varphi}$ and $\tilde{\psi}$ respectively.

The construction strategy consists of seeking finite impulse response filters which satisfy a condition known as perfect reconstruction, and then deducing from them the scaling functions and the wavelets.

The perfect reconstruction condition written with the four filters m_0, $\widetilde{m}_0$, m_1 and $\widetilde{m}_1$, leads to a system of two relations which is simplified for FIR filters into only one relation linking m_0 and $\widetilde{m}_0$:

$$\overline{m_0(\omega)}\widetilde{m_0}(\omega) + \overline{m_0\left(\omega+\tfrac{1}{2}\right)}\widetilde{m_0}\left(\omega+\tfrac{1}{2}\right) = 1$$

If a pair $\left(m_0, \widetilde{m_0}\right)$ satisfies this condition, for the two other filters we will take, for example, almost as in the orthogonal case (see proposal 2.2):

$$\begin{cases} m_1(\omega) = e^{-i\pi\omega}\overline{\widetilde{m_0}\left(\omega+\tfrac{1}{2}\right)} \\ \widetilde{m_1}(\omega) = e^{-i\pi\omega}\overline{m_0\left(\omega+\tfrac{1}{2}\right)} \end{cases}$$

We can then pass to wavelets. In the sense of tempered distributions and having the relations $\hat{\psi}(2\omega) = m_1(\omega)\hat{\varphi}(\omega)$ and $\hat{\tilde{\psi}}(2\omega) = \widetilde{m_1}(\omega)\hat{\tilde{\varphi}}(\omega)$ we build two functions φ and $\tilde{\varphi}$ and then two functions ψ and $\tilde{\psi}$ candidates for being wavelets, by:

$$\hat{\varphi}(\omega) = \prod_{n=1}^{\infty} m_0\left(2^{-n}\omega\right) \quad \text{and} \quad \hat{\tilde{\varphi}}(\omega) = \prod_{n=1}^{\infty} \widetilde{m_0}\left(2^{-n}\omega\right)$$

A difficulty appears, stemming from the fact that the functions produced do not always have finite energy. Additional explanations are found in Chapter 5.

Let us make a last comment on a particular case of biorthogonal wavelets: the semi-orthogonal wavelets.

2.6.4. *Semi-orthogonal wavelets*

Let φ and $\tilde{\varphi}$ be two functions such that the families $\left\{\varphi_{0,k}\right\}_{k\in\mathbb{Z}}$ and $\left\{\tilde{\varphi}_{0,k}\right\}_{k\in\mathbb{Z}}$ verify the property [2.24] and generate the same space; in other words, let us suppose that $\widetilde{V_0} = V_0$. The families $\left\{\varphi_{0,k}\right\}_{k\in\mathbb{Z}}$ and $\left\{\tilde{\varphi}_{0,k}\right\}_{k\in\mathbb{Z}}$ are thus two dual (or biorthogonal) bases of V_0. Then, for $j \in \mathbb{Z}$ we have: $\widetilde{V_j} = V_j$ and $\widetilde{W_j} = W_j$. In fact, we have a single multi-resolution analysis, the orthogonality between spaces $\left\{W_j\right\}$ is preserved and each space V_j and W_j is provided with two biorthogonal bases.

One of the simplest methods of constructing new wavelet bases consists of building bases known as semi-orthogonal wavelet bases by appropriate modification of the bases of spaces $\left\{V_j\right\}$ and $\left\{W_j\right\}$ using an orthogonal wavelet ψ associated with a scaling function φ and a multi-resolution analysis (MRA) of L^2.

More precisely, Aldroubi and Unser [ALD 93] show that if φ is a scaling function associated with an MRA and if p is an acceptable filter (essentially p invertible for the convolution), then function φ^p defined by:

$$\varphi^p = p * \varphi = \sum_{k\in\mathbb{Z}} p_k \varphi(t-k)$$

is another scaling function generating the same MRA.

Of course, the original scaling function is characterized by the two-scale relation $\frac{1}{2}\varphi\left(\frac{t}{2}\right) = \sum_{k\in\mathbb{Z}} a_k \varphi(t-k)$. A similar relation exists for φ^p via a filter expressed using p and a. However, the family of $\left\{\varphi^p(t-k)\right\}_{k\in\mathbb{Z}}$ is no longer orthogonal. In fact, we work with a single multi-resolution analysis, the orthogonality between the $\left\{W_j\right\}$ spaces is preserved, and we sacrifice the orthogonality internal to each

space: each space V_j and W_j is provided with two biorthogonal bases. In [ABR 97] and [ABR 95] we will find the processes of calculating the four filters associated with the pair of wavelets in duality using a and p.

Various techniques are used in the studies exploiting this idea: linear combinations, projections and convolutions. Let us specify the second one a little.

Let us suppose that we take an *a priori* form noted f and that we seek a scaling function φ_f approaching f. A possible solution consists of using an already existing MRA and defining φ_f by $\varphi_f = P_{V_0}(f)$, the orthogonal projection of f to the space V_0, which is the best approximation of f by an element of V_0 in the sense of least on squares. The projection φ_f is then written as $\varphi_f = \sum_{k\in\mathbb{Z}} p_k^{(f)} \varphi(t-k)$ and the problem is solved, provided that $p^{(f)}$ is an admissible function. Obviously, we may wish to seek not a scaling function but a wavelet ψ_f resembling a given form f. The solution that consists of projecting on W_0, i.e. choosing $\psi_f = P_{W_0}(f)$ leads to a solution under a similar admissibility condition.

Chapter 3

From Wavelet Bases to the Fast Algorithm

3.1. Introduction

Stéphane Mallat [MAL 89] proposed the fast algorithm of decomposition-reconstruction for the discrete wavelet transform at the end of the 1980s. He thus established the link between orthonormal wavelet bases, whose mathematical development was then recent, and traditional filter banks in signal processing.

This unifying point of view brought the two communities closer, enabled an increased development of applications for signals or images and aroused theoretical interest. For example, the fruitful filters approach led to wavelet synthesis or compression.

Moreover, two unexpected features have to be pointed out: the algorithm is remarkably simple and its complexity grows only linearly with the size of data, i.e. it is lower than that of the fast Fourier transform. This aspect is obviously crucial for applications.

The chapter begins with the discrete wavelet transform algorithm of sampled signals (shortened to DWT for discrete wavelet transform). It is a purely discrete framework, in the sense that instead of a function we decompose a finite sequence, using finite impulse response filters. In the language of signal processing, it is the implementation of two-channel [EST 77] filter banks and filter banks with perfect reconstruction (this point of view will be developed in the second part of Chapter 5).

Then we consider the justification of the algorithm coming back to the framework of continuous time signals provided by multi-resolution analyses

introduced in Chapter 2. We then show the close connection between the previous algorithm and the calculation of the components of a function on a basis of wavelets.

Next we tackle the various problems raised by the concrete implementation of the algorithm, in particular, the phase of initialization and the manner of treating edge effects. We then evaluate its algorithmic complexity, before extending the DWT to the case of decomposition-reconstruction of images.

Finally, thanks to the DWT, we define a transform invariant by translation, often called SWT (for stationary wavelet transform). Indeed, the discrete wavelet transform is not invariant by translation in time, as opposed to the continuous wavelet transform. The SWT [COI 95], [PES 96] and [NAS 95] restores this property and is used primarily for signal or image denoising (see Chapters 7 and 8).

3.2. From orthonormal bases to the Mallat algorithm

Making a decomposition means calculating the co-ordinates of the signal on the vectors of the wavelet base.

Let us recall the framework presented in Chapter 2. Let $M = \left\{V_j\right\}_{j\in\mathbb{Z}}$ be a multi-resolution analysis of $L^2(\mathbb{R})$. It is a family of decreasing closed vectorial subspaces, which enable the approximation of functions and have properties involving dilation and translation of functions.

For fixed $j \in \mathbb{Z}$ the space W_j is defined as the orthogonal complement to the space V_j in space V_{j-1}. For a function $f \in L^2$ the decomposition in wavelets consists of calculating the co-ordinates of orthogonal projections of f on V_j and W_j respectively, noted $A^j = P_{V_j} f$ and $D^j = P_{W_j} f$. These co-ordinates in spaces V_j and W_j (provided with the orthonormal bases: $\left\{\varphi_{j,k}\right\}_{k\in\mathbb{Z}}$ and $\left\{\psi_{j,k}\right\}_{k\in\mathbb{Z}}$ where φ and ψ are the scaling function and the wavelet) are nothing but approximation and detail coefficients defined by the following relations:

$$A^j = \sum_{p\in\mathbb{Z}} a_p^j \varphi_{j,p} \quad \text{with} \quad a_p^j = \left(A^j, \varphi_{j,p}\right)_{L^2} = \left(f, \varphi_{j,p}\right)_{L^2}$$

$$D^j = \sum_{p\in\mathbb{Z}} d_p^j \psi_{j,p} \quad \text{with} \quad d_p^j = \left(D^j, \psi_{j,p}\right)_{L^2} = \left(f, \psi_{j,p}\right)_{L^2}$$

The calculations of signal co-ordinates on the vectors of the wavelets basis are performed as usual by evaluation of scalar products. The cost of the numerical approximation of the corresponding integrals is high.

Fortunately, in the context of wavelet bases we may exploit three properties:

– on the one hand, there is only a finite number of co-ordinates to calculate if the signal is compactly supported;

– on the other hand, the basis is organized by level of scale;

– finally, the basic functions of a scale are related to those of the following one by a two-scale relation.

Consequently, if the co-ordinates a^{j-1} of the approximation at a scale are known, the co-ordinates a^j and d^j on the following scale may be deduced from it almost immediately, by means of a very simple transformation: a convolution, which is a linear filter, followed by dyadic sub-sampling.

Before justifying the algorithm outlined above let us present it in two phases: filters and the effective calculation of the coefficients.

3.3. Four filters

For an orthogonal wavelet ψ, the associated scaling function φ satisfies a fundamental relation, which is the following twin scale equation:

$$\frac{1}{2}\varphi\left(\frac{t}{2}\right) = \sum_{n\in\mathbb{Z}} a_n \varphi_{0,n} = \sum_{n\in\mathbb{Z}} a_n \varphi(t-n)$$

The filters involved in the discrete wavelet transform and in the inverse transform (noted IDWT) are closely linked to the $(a_n)_{n\in\mathbb{Z}}$ sequence. If φ (and, consequently, ψ) have compact support, the sequence $(a_n)_{n\in\mathbb{Z}}$ has only a finite number of non-zero elements. We may then see this sequence as a low-pass filter. This filter, noted w, thus lets the low frequencies through and retains the high ones. It has a finite impulse response (FIR), with length noted K, its sum is equal to 1 and its norm is $\frac{1}{\sqrt{2}}$.

Using the filter w we define four filters with finite impulse response, size *K* and norm 1. The decomposition filters are noted (indicated by the final *D*): *LoD* and

HiD. The first is low-pass (indicated by the initial *Lo*) and the second is high-pass (indicated by the initial *Hi*). The two reconstruction filters (indicated by the final *R*) are noted *LoR* and *HiR*.

Filters	**Low-pass**	**High-pass**
Decomposition	LoD	HiD
Reconstruction	LoR	HiR

Table 3.1. *Four filters*

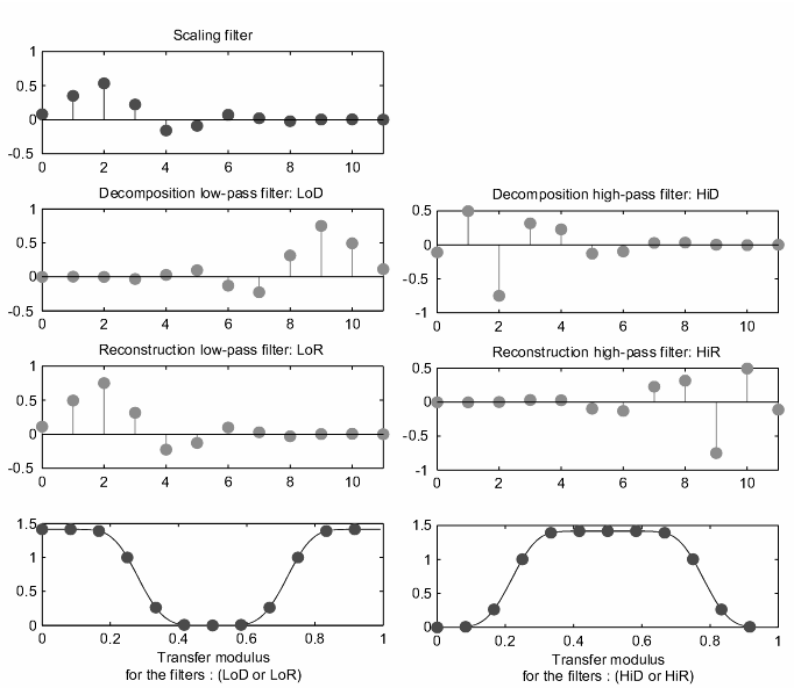


Figure 3.1. *Filters for the wavelet db6*

The two reconstruction filters are linked by:

$$LoR = \frac{w}{\|w\|} \text{ and } HiR_k = (-1)^{k-1} LoR_{K+1-k} \text{ for } k = 1,2,\ldots,K$$

They are mirror filters in quadrature. The two decomposition filters are obtained via mirror image of the reconstruction filters:

$$LoD_k = LoR_{K+1-k} \text{ and } HiD_k = HiR_{K+1-k} \text{ for } k = 1,2,\ldots,K$$

For the *db6* wavelet (see Chapter 4) we represent in Figure 3.1:

– the filter corresponding to the scaling function, at the top;

– the four filters deduced from it, in the middle;

– modules of the filters transfer functions, at the bottom.

3.4. Efficient calculation of the coefficients

The algorithm of the DWT of a signal X of length N consists of performing several elementary decomposition steps.

Starting with the signal X, the first step produces two vectors of coefficients: approximation coefficients a^1 and detail coefficients d^1. These vectors are obtained by a convolution[1] of the signal X with the low-pass filter LoD for the approximation and with the high-pass filter HiD for the detail, followed in both cases by a dyadic decimation[2]. They are approximately $\frac{N}{2}$ in length.

More precisely, the first step of the algorithm can be represented by Figure 3.2a.

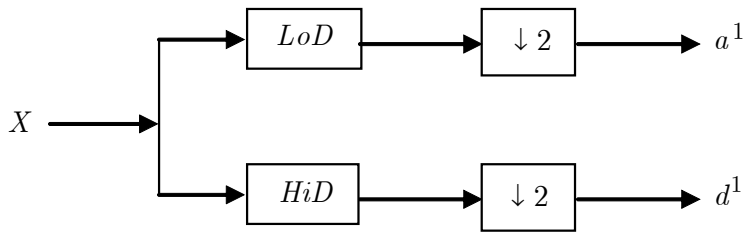


Figure 3.2a. *Representation of the first step of the DWT where* $\boxed{F}$ *represents convolution by F and* $\boxed{\downarrow 2}$ *represents decimation*

1 The convolution of a signal X by a filter F is defined by $[X * F]_n = \sum_k X_{n-k} F_k$.

2 The decimation of a signal X is defined by $Y = dec(X)$ where $Y_n = X_{2n}$.

The operations brought into play naturally lead to consider higher order decompositions. The following step thus consists of breaking up the approximation coefficients a^1 into two, replacing the signal X by a^1 and producing a^2 and d^2. The algorithm follows the same path. The decomposition enabling the passage from level j to level $j+1$ is illustrated in Figure 3.2b.

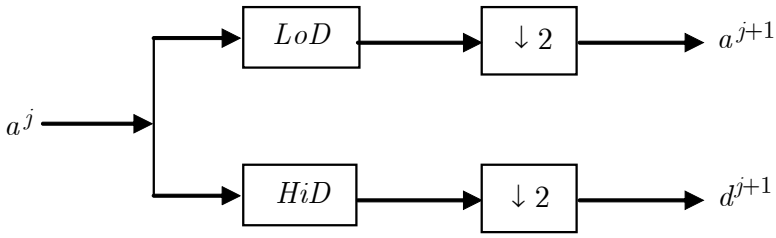


Figure 3.2b. *Representation of a DWT decomposition step with initialization at:* $a^0 = X$

It is thus a very simple algorithm. We now will justify it referring back to the elements introduced in Chapter 2 concerning orthonormal wavelet bases.

3.5. Justification: projections and twin scales

Let us note respectively by P_{V_j} and P_{W_j} the orthogonal projection operators, on V_j and W_j, and for $f \in L^2$ let us note the corresponding projections $A^j = P_{V_j} f$ and $D^j = P_{W_j} f$. The properties of spaces $\{V_j\}$ and $\{W_j\}$ imply whereas:

$$A^{j-1} = A^j + D^j \text{ with } A^j \perp D^j \qquad [3.1]$$

Spaces $\{V_j\}$ are approximation spaces in the following sense: A^j converges to f when j tends to $-\infty$; in the same way, spaces $\{W_j\}$ are detail spaces in the sense that D^j are the differences between two successive approximations: $D^j = A^{j-1} - A^j$. Moreover, for fixed J the D^j are the corrections to add to the roughest approximation A^J in order to find f, since $f = A^J + \sum_{j=-\infty}^{J} D^j$.

In addition, spaces V_j and W_j are endowed with orthonormal bases: $\left\{\varphi_{j,k}\right\}_{k\in\mathbb{Z}}$ and $\left\{\psi_{j,k}\right\}_{k\in\mathbb{Z}}$ where φ and ψ are the scaling function and wavelet associated with the multi-resolution analysis (for a function g, we note $g_{j,k} = 2^{-j/2} g\left(2^{-j}t - k\right)$). Consequently, we may describe the approximation and detail coefficients by the following relations valid for all $j \in \mathbb{Z}$:

$$A^j = \sum_{p\in\mathbb{Z}} a_p^j \varphi_{j,p} \quad \text{with} \quad a_p^j = \left(A^j, \varphi_{j,p}\right)_{L^2} = \left(f, \varphi_{j,p}\right)_{L^2} \qquad [3.2]$$

$$D^j = \sum_{p\in\mathbb{Z}} d_p^j \psi_{j,p} \quad \text{with} \quad d_p^j = \left(D^j, \psi_{j,p}\right)_{L^2} = \left(f, \psi_{j,p}\right)_{L^2} \qquad [3.3]$$

The Mallat algorithm, which is a fast algorithm of decomposition-reconstruction, is based on the three previous relations. In the decomposition part we calculate the co-ordinates a^j and d^j of approximation A^j and detail D^j using the co-ordinates a^{j-1} of the approximation A^{j-1}. In the reconstruction part the inverse operation is performed.

3.5.1. *The decomposition phase*

With relations [3.1], [3.2] and [3.3] we obtain:

$$a_k^j = \left(A^{j-1}, \varphi_{j,k}\right)_{L^2} \quad \text{and} \quad d_k^j = \left(A^{j-1}, \psi_{j,k}\right)_{L^2} \qquad [3.4]$$

Indeed, for all $k \in \mathbb{Z}$ using [3.1] and the orthogonality between V_j and W_j we have:

$$\begin{aligned}\left(A^{j-1}, \varphi_{j,k}\right)_{L^2} &= \left(A^j + D^j, \varphi_{j,k}\right)_{L^2} \\ &= \left(A^j, \varphi_{j,k}\right)_{L^2} + \left(D^j, \varphi_{j,k}\right)_{L^2} = a_k^j + 0\end{aligned}$$

In the same way, by replacing φ by ψ we obtain for the detail coefficient:

$$\begin{aligned}\left(A^{j-1},\psi_{j,k}\right)_{L^2} &= \left(A^{j}+D^{j},\psi_{j,k}\right)_{L^2}\\ &= \left(A^{j},\psi_{j,k}\right)_{L^2}+\left(D^{j},\psi_{j,k}\right)_{L^2} = 0+d_k^j\end{aligned}$$

By substituting $A^{j-1}=\sum_{p\in\mathbb{Z}} a_p^{j-1}\varphi_{j-1,p}$ to the two equalities of [3.4], it follows:

$$\begin{cases} a_k^j = \left(A^{j-1},\varphi_{j,k}\right)_{L^2} = \left(\sum_{p\in\mathbb{Z}} a_p^{j-1}\varphi_{j-1,p},\varphi_{j,k}\right)_{L^2} \\ \qquad\qquad = \sum_{p\in\mathbb{Z}} a_p^{j-1}\left(\varphi_{j-1,p},\varphi_{j,k}\right)_{L^2} \\ d_k^j = \left(A^{j-1},\psi_{j,k}\right)_{L^2} = \left(\sum_{p\in\mathbb{Z}} a_p^{j-1}\varphi_{j-1,p},\psi_{j,k}\right)_{L^2} \\ \qquad\qquad = \sum_{p\in\mathbb{Z}} a_p^{j-1}\left(\varphi_{j-1,p},\psi_{j,k}\right)_{L^2} \end{cases} \qquad [3.5]$$

Let us calculate the scalar products appearing in [3.5]:

$$\begin{cases} \left(\varphi_{j-1,p},\varphi_{j,k}\right)_{L^2} = \int_{\mathbb{R}} 2^{-(j-1)/2}\,\varphi\left(2^{-(j-1)}x-p\right)2^{-j/2}\,\varphi\left(2^{-j}x-k\right)dx \\ \left(\varphi_{j-1,p},\psi_{j,k}\right)_{L^2} = \int_{\mathbb{R}} 2^{-(j-1)/2}\,\varphi\left(2^{-(j-1)}x-p\right)2^{-j/2}\,\psi\left(2^{-j}x-k\right)dx \end{cases}$$

Changing the variable $\frac{y}{2}=2^{-j}x-k$, which is equivalent to $x=2^{j-1}y+2^jk$ we obtain:

$$\begin{cases} \left(\varphi_{j-1,p},\varphi_{j,k}\right)_{L^2} = \sqrt{2}\int_{\mathbb{R}} \varphi(y-(p-2k))\,\frac{1}{2}\varphi\left(\frac{y}{2}\right)dy \\ \left(\varphi_{j-1,p},\psi_{j,k}\right)_{L^2} = \sqrt{2}\int_{\mathbb{R}} \varphi(y-(p-2k))\,\frac{1}{2}\psi\left(\frac{y}{2}\right)dx \end{cases} \qquad [3.6]$$

We may now use the scaling equation verified by φ and the two-scale equation defining the wavelet ψ (see equations [2.16] and [2.20] in Chapter 2):

$$\begin{aligned}
&\tfrac{1}{2}\varphi\left(\tfrac{y}{2}\right) = \sum_{n\in\mathbb{Z}} a_n\, \varphi_{0,n} = \sum_{n\in\mathbb{Z}} a_n\, \varphi(y-n) \text{ and} \\
&\tfrac{1}{2}\psi\left(\tfrac{y}{2}\right) = \sum_{n\in\mathbb{Z}} b_n\, \varphi_{0,n} = \sum_{n\in\mathbb{Z}} b_n\, \varphi(y-n)
\end{aligned}$$

Since $\left\{\varphi_{0,k}\right\}_{k\in\mathbb{Z}}$ is an orthonormal base of V_0, we conclude from [3.6] that:

$$\begin{cases}
\left(\varphi_{j-1,p}, \varphi_{j,k}\right)_{L^2} = \sqrt{2}\; a_{p-2k} \\
\left(\varphi_{j-1,p}, \psi_{j,k}\right)_{L^2} = \sqrt{2}\; b_{p-2k}
\end{cases} \qquad [3.7]$$

The scalar products of the basic functions of the scales $j-1$ and j do not depend on j. Consequently, the same elementary step of decomposition applies to all the scales.

Returning to the coefficients a_k^j and d_k^j relations [3.5] and [3.7] then make it possible to write:

$$\begin{cases}
a_k^j = \sum_{p\in\mathbb{Z}} a_p^{j-1}\sqrt{2}\; a_{p-2k} \\
d_k^j = \sum_{p\in\mathbb{Z}} a_p^{j-1}\sqrt{2}\; b_{p-2k}
\end{cases} \qquad [3.8]$$

In view of computerized implementation it is interesting to define the two filters involved in this decomposition phase. They were introduced previously for the particular case of the finite filters.

$$\begin{cases}
LoD = \left\{[LoD]_n\right\}_{n\in\mathbb{Z}} \text{ with } [LoD]_n = \sqrt{2}a_{-n} \\
HiD = \left\{[HiD]_n\right\}_{n\in\mathbb{Z}} \text{ with } [HiD]_n = \sqrt{2}\, b_{-n}
\end{cases}$$

With these notations relations [3.8] are then written:

$$\begin{cases}
a_k^j = \sum_{p\in\mathbb{Z}} a_p^{j-1}\, [LoD]_{2k-p} \\
d_k^j = \sum_{p\in\mathbb{Z}} a_p^{j-1}\, [HiD]_{2k-p}
\end{cases} \qquad [3.9]$$

In the spaces of sequences we can introduce the operation of decimation $y = dec(x)$ defined by $y_n = x_{2n}$ for $n \in \mathbb{Z}$. In filter theory, decimation is often noted $\downarrow 2$. With this operation we easily verify that [3.9] is finally written:

$$\begin{cases} a^j = dec\left(a^{j-1} * LoD\right) \\ d^j = dec\left(a^{j-1} * HiD\right) \end{cases} \qquad [3.10]$$

The coefficients of approximation A^j and detail D^j are therefore calculated simply using the coefficients of approximation A^{j-1}: a convolution (a filter) followed by a decimation. We can represent [3.10] by Figure 3.3a.

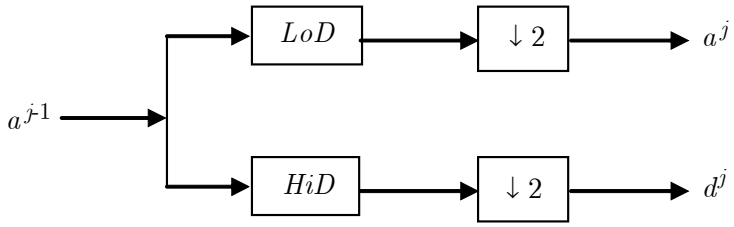


Figure 3.3a. *Representation of equation [3.10] where* $\boxed{F}$ *represents convolution by* F *and* $\boxed{\downarrow 2}$ *represents decimation*

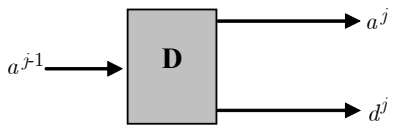


Figure 3.3b. *Compact representation of equation [3.10]*

3.5.2. *The reconstruction phase*

To calculate the co-ordinates of A^{j-1} knowing those of A^j and D^j we again start with the equality $A^{j-1} = A^j + D^j$. For all $k \in \mathbb{Z}$ we obtain:

$$\begin{aligned} a_k^{j-1} &= \left(A^{j-1}, \varphi_{j-1,k}\right)_{L^2} = \left(A^j + D^j, \varphi_{j-1,k}\right)_{L^2} \\ &= \left(A^j, \varphi_{j-1,k}\right)_{L^2} + \left(D^j, \varphi_{j-1,k}\right)_{L^2} \end{aligned} \qquad [3.11]$$

from which by developing A^j and D^j and noting that the permutation of $\sum$ and $\int$ hidden in the scalar product is valid:

$$a_k^{j-1} = \left(\sum_{p\in\mathbb{Z}} a_p^j \varphi_{j,p}, \varphi_{j-1,k} \right)_{L^2} + \left(\sum_{p\in\mathbb{Z}} d_p^j \psi_{j,p}, \varphi_{j-1,k} \right)_{L^2}$$

thus:

$$a_k^{j-1} = \sum_{p\in\mathbb{Z}} a_p^j \left(\varphi_{j,p}, \varphi_{j-1,k} \right)_{L^2} + \sum_{p\in\mathbb{Z}} d_p^j \left(\psi_{j,p}, \varphi_{j-1,k} \right)_{L^2} \qquad [3.12]$$

By again taking the result for the scalar products obtained in [3.7] and putting it in equality [3.12], we deduce that:

$$a_k^{j-1} = \sum_{p\in\mathbb{Z}} a_p^j \sqrt{2}\, a_{k-2p} + \sum_{p\in\mathbb{Z}} d_p^j \sqrt{2}\, b_{k-2p} \qquad [3.13]$$

As for the decomposition, two filters are involved in the reconstruction:

$$\begin{cases} LoR = \{ [LoR]_n \}_{n\in\mathbb{Z}} & \text{with} \quad [LoR]_n = \sqrt{2}\, a_n \\ HiR = \{ [HiR]_n \}_{n\in\mathbb{Z}} & \text{with} \quad [HiR]_n = \sqrt{2}\, b_n \end{cases} \qquad [3.14]$$

These two filters are easily deduced from those introduced during the decomposition phase:

$$LoR = inverse(LoD) \text{ and } HiR = inverse(HiD)$$

where $y = inverse(x)$ is defined by: $y(k) = x(-k)$, $k \in \mathbb{Z}$.

With the [3.14] notations, relation [3.13] is written:

$$a_k^{j-1} = \sum_{p\in\mathbb{Z}} a_p^j [LoR]_{k-2p} + \sum_{p\in\mathbb{Z}} d_p^j [HiR]_{k-2p} \qquad [3.15]$$

In the space of sequences we can also introduce the operation of zeros insertion defined by $y = ins(x)$ with $y_{2n} = x_n$ and $y_{2n+1} = 0$ for $n \in \mathbb{Z}$. In filter theory insertion is often noted $\uparrow 2$. Relation [3.15] takes the form:

$$a_k^{j-1} = \sum_{p\in\mathbb{Z}} \left[ins\left(a^j\right) \right]_p [LoR]_{k-p} + \sum_{p\in\mathbb{Z}} \left[ins\left(d^j\right) \right]_p [HiR]_{k-p}$$

and ultimately:

$$a^{j-1} = ins\left(a^j\right) * LoR + ins\left(d^j\right) * HiR \qquad [3.16]$$

The coefficients of approximation A^{j-1} are thus obtained simply from the coefficients of approximation A^j and detail D^j by an insertion followed by a convolution (of a linear filter) represented in Figures 3.4a and 3.4b.

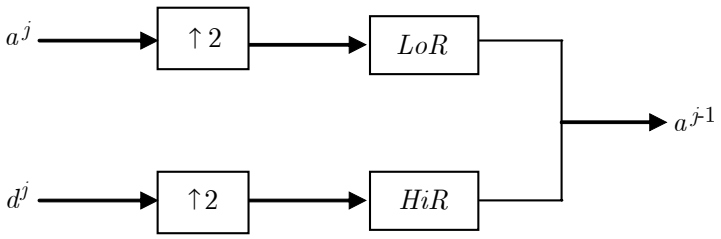


Figure 3.4a. *Representation of a DWT reconstruction step where* $\boxed{F}$ *represents the convolution by* F *and* $\boxed{\uparrow 2}$ *represents the zero insertion*

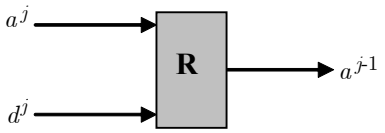


Figure 3.4b. *Compact representation of a DWT reconstruction step*

3.5.3. *Decompositions and reconstructions of a higher order*

In the two preceding sections the possibility of carrying out high order decomposition and reconstruction appeared obvious. They simply translate the following successive equalities:

$$\left|\begin{array}{l} A^{j-1} = A^{j} + D^{j} \\ A^{j-1} = A^{j+1} + D^{j+1} + D^{j} \\ \cdots \\ A^{j-1} = A^{J} + D^{J} + \cdots + D^{j} \end{array}\right. \qquad [3.17]$$

where all the decompositions are orthogonal.

Schematically we may represent the reiterated decomposition phases in the following form.

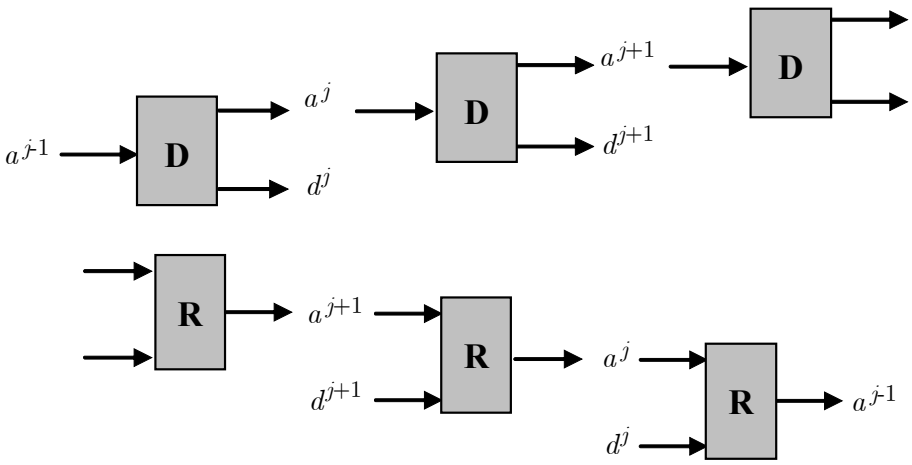


Figure 3.5. *Representation of several steps of DWT decomposition (top) and reconstruction (bottom)*

3.6. Implementation of the algorithm

The theoretical justification of the algorithm must be supplemented in order to arrive at the discrete wavelet transform of finite sequences. In fact, to be precise, we

will use finite sequences and wavelets linked to FIR filters. This section explains how to implement the DWT efficiently.

3.6.1. *Initialization of the algorithm*

The first difficulty is of course the initialization of the algorithm. For a given finite energy function f the preceding section provides a process to deduce the co-ordinates a^j and d^j of approximation A^j and of detail D^j from the co-ordinates a^{j-1} of approximation A^{j-1}. But where should we start?

A first solution, adapted if we start with a known function $f(t)$ with $t \in \mathbb{R}$, consists of selecting a desired precision $\varepsilon > 0$. The properties of multi-resolution analyses imply that f can then be approximated by its orthogonal projection onto one of the approximation spaces $\{V_j\}$, let us say V_{j_0}, noted $f_{j_0} = P_{V_{j_0}}(f)$, so that $\left\| f - f_{j_0} \right\|_{L^2} < \varepsilon$. It is then enough to calculate the co-ordinates of f_{j_0} on the basis of V_{j_0} by a numerical evaluation of the sequence a^{j_0} given by: $a_k^{j_0} = \left(A^{j_0}, \varphi_{j_0,k} \right)_{L^2} = \left(f, \varphi_{j_0,k} \right)_{L^2}$. The algorithm is then initialized.

In general, we do not know the values of $f(t)$ for $t \in \mathbb{R}$, but only the values of f in certain points: $f(t_i), i = 1,...,n$. A possible procedure to initialize the algorithm consists of returning to the preceding situation by interpolating the function between the t_i and carrying out an extrapolation outside of $[t_1, t_n]$ to construct a function $\tilde{f}$ defined on $\mathbb{R}$.

Often we only have an equispaced sampling of the f function: $f(i\delta), i = 1,...,n$ where δ is the sampling step. We could of course reuse the preceding solution but we prefer to suppose more abruptly that we observe the coefficients a^0 of the approximation A^0 directly and assimilate the sampled values of f at a^0. Naturally, we thus make an error that, in general, cannot be controlled, except, for example, for orthogonal wavelets of the coiflets family: *coifn* (see Chapter 4).

Nevertheless, this process has an interesting interpretation. Considering the sampled signal as a sequence of $l^2(\mathbb{Z})$, the decomposition has a sense within the framework of the multi-resolution analyses of $l^2(\mathbb{Z})$ (which are not explored here, but can be found in [FRA 99]). Indeed, to any multi-resolution we can associate $L^2(\mathbb{R})$, a multi-resolution analysis of $l^2(\mathbb{Z})$. We write $V_0 = l^2(\mathbb{Z})$ provided with the canonical base with sequence and the coarser spaces V_j and W_j, i.e.

associated with $j > 0$, are generated by the translations-dilations of the usual filters.

3.6.2. *Calculation on finite sequences*

The last difficulty relates to the support of observations, necessarily finite, over which we know the function or a real life signal. Indeed, the application of the algorithm passes by convolution, producing a signal, which is not a sequence indexed by $\mathbb{Z}$ but a finite sequence and the first and the last terms of the convolution product invoke sequence terms that are not defined. Two large solution families are available to resolve this difficulty.

The first one is of a theoretical nature. It consists of defining the wavelet bases not over $L^2(\mathbb{R})$, but over $L^2([a,b])$. These are known as wavelets on the interval. There exist many constructions of this type, the best known being undoubtedly that of Cohen, Daubechies and Vial [COH 93]. Theoretically interesting, it nevertheless presents difficulties when applied to a real life problem. First of all, it is relatively complicated and difficult to adapt. Moreover, it requires signal pre-processing before applying the DWT to it, as well as post-processing of the coefficients obtained, creating edge effects that may sometimes be very awkward.

For this reason we often prefer methods that are easier to implement and adapt, for example, in the case of signals whose length is not a power of 2. They consist of arbitrarily prolonging the signal by inserting pseudo-observations outside the support. This is the subject of the following section.

3.6.3. *Extra coefficients*

This extension must be carried out not only for the signal X for the first stage of the algorithm, but also for each following stage. Thus, it generates extra coefficients whose number depends linearly on the size of the wavelet. There are various usual methods:

– *extension by zeros.* In this case we suppose that the signal is zero everywhere outside of the original support. The disadvantage is to artificially create discontinuities at the edges;

– *symmetrization.* This method extends signals and images outside of their support by repeating near edge values by symmetry. This method, in general adapted to the images, creates first order derivative discontinuities;

– *regular extrapolation.* We perform a low degree polynomial extrapolation outside the support: typically 0 (repetition of the first and the last value respectively) or 1 (linear extrapolation). This method is poorly adapted to noisy signals;

– *periodic extension.* This method consists of supposing that the signal is periodic. It makes it possible to minimize the number of extra coefficients. Of course, the periodization of a signal, in general, leads to discontinuities at the edges.

These methods of extension make it possible to associate a finite sequence with a time (in $\mathbb{Z}$) signal, for which the algorithm has been theoretically justified. In particular, we have the following properties:

– preservation of the norm square by decomposition:

$$\|X\|^2 = \left\|a^1\right\|^2 + \left\|d^1\right\|^2 ;$$

– orthogonality:

$$A \perp D \text{ and } \|X\|^2 = \|A\|^2 + \|D\|^2 ;$$

– perfect reconstruction:

$$X = A + D$$

These properties verified for extended signals are lost for the initial finite sequence. Perfect reconstruction is preserved providing that the reconstruction part to the manner of truncating the coefficients obtained is adapted, for example, by selecting the central part of the convolution product.

3.7. Complexity of the algorithm

Let us suppose that the filters LoD and HiD have K non-zero coefficients and that the signal (identified as a^0) has a length N. Then the signals a^j and d^j have the length $2^{-j}N$, if we neglect the possible extra coefficients according to the adopted calculation path.

The elementary decomposition step making it possible to pass from level $j-1$ of approximation coefficients to level j of detail and approximation coefficients is given by:

$$\begin{cases} a^j = dec\left(a^{j-1} * LoD\right) \\ d^j = dec\left(a^{j-1} * HiD\right) \end{cases}$$

Each of the two above lines requires approximately $2^{-j+1}NK$ operations. This elementary decomposition step thus requires $2^{-j+2}NK$ additions and multiplications. Decomposition at level J corresponds to J successive elementary decompositions for $j=1$ to $j=J$. It therefore requires $\sum_{j=1}^{J} 2^{-j+2}NK = 2\left(\sum_{j=0}^{J-1} 2^{-j}\right)NK \leq 4NK$ additions and multiplications.

Consequently, the global complexity of the wavelet decomposition up to the J level of a signal with length N by K non-zero coefficient filter bank is of the order of $4NK$.

NOTE 3.1. – a direct implementation of formulae [3.9], whose expression is less elegant than a convolution followed by a decimation, requires twice fewer operations. The complexity obtained in this case is obviously the same, up to a constant 2. Let us mention that other algorithms are possible and improve the constant still further; let us quote in particular the polyphase method (see Chapter 5). ■

By proceeding similarly we demonstrate that the reconstruction step requires at most $4NK$ additions and multiplications. The discrete wavelet transform, thus, has a total complexity in $O(N)$, i.e. linear with respect to the size of the data, with a constant term that grows linearly with respect to the length of the filters used. It is completely remarkable, since its complexity is lower than that of the fast Fourier transform.

3.8. From 1D to 2D

For the images, a similar algorithm is usable with wavelets and scaling functions obtained by the tensor product of one-dimensional wavelets (see Chapter 8).

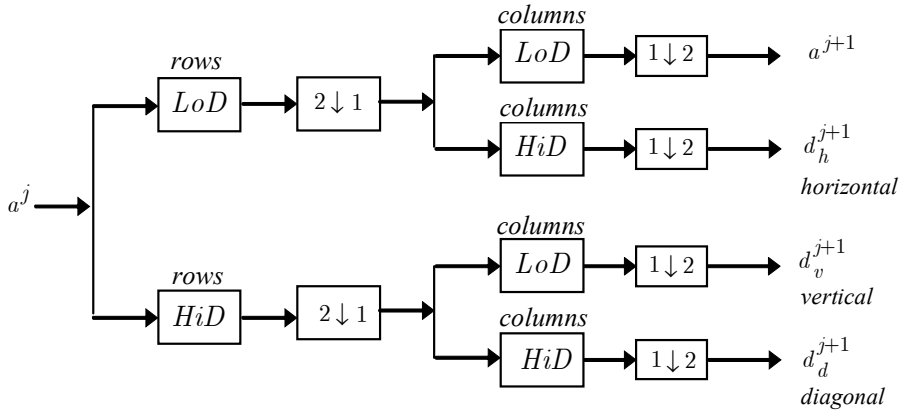


Figure 3.6a. *Basic step for the decomposition in wavelets for images*

This type of transformation using a two-dimensional DWT leads to a decomposition of the approximation coefficients at the j level in four distinct components: the approximation and details according to three orientations, horizontal, vertical and diagonal, at the level $j+1$. Calculations are simple: we filter the rows of a^j and decimate them; then, we filter and decimate the columns of the matrices obtained. Figure 3.6 presents the basic step of wavelet decomposition for images where $\overset{rows}{\boxed{F}}$ (respectively $\overset{columns}{\boxed{F}}$) represents the convolution by F of rows (respectively columns) of the matrix, $\boxed{2 \downarrow 1}$ represents the decimation of the columns (conservation of even column index) and $\boxed{1 \downarrow 2}$ represents the decimation of the lines (conservation of even index lines). For the reconstruction, the basic step is given by Figure 3.6b where $\boxed{2 \uparrow 1}$ represents upsampling (zero insertion) of the columns and $\boxed{1 \uparrow 2}$ represents the upsampling of rows.

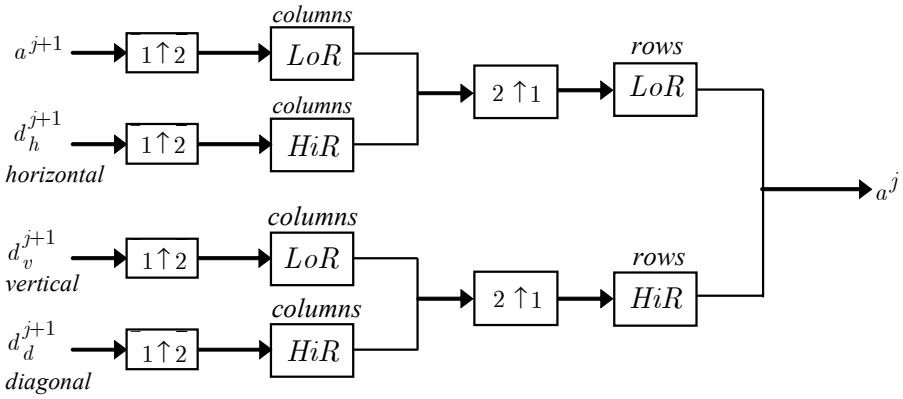


Figure 3.6b. *Basic step for wavelet reconstruction for images*

It should be mentioned that for biorthogonal wavelets the same algorithms can be used, but the decomposition filters, on the one hand, and the reconstruction filters, on the other, are obtained from two distinct scaling functions associated with two multi-resolution analyses in duality (see Chapter 2).

3.9. Translation invariant transform

The traditional DWT suffers from a known disadvantage: it is not invariant by translation in time, contrary to the continuous wavelet transform. This means that, even in the case of periodic extension of a signal X, the DWT of a translated version of X is not, in general, the translated version of the DWT of X.

A simple example of this difficulty appears in the analysis of a square form type signal X. It is composed in the following manner: 256 zero values, then 511 values equal to 1 and, finally, 257 zero values. Using the DWT, with the Haar wavelet, on the left of Figure 3.7 we find the level 1 analysis of X and on the right, the level 1 analysis of X translated by one time step.

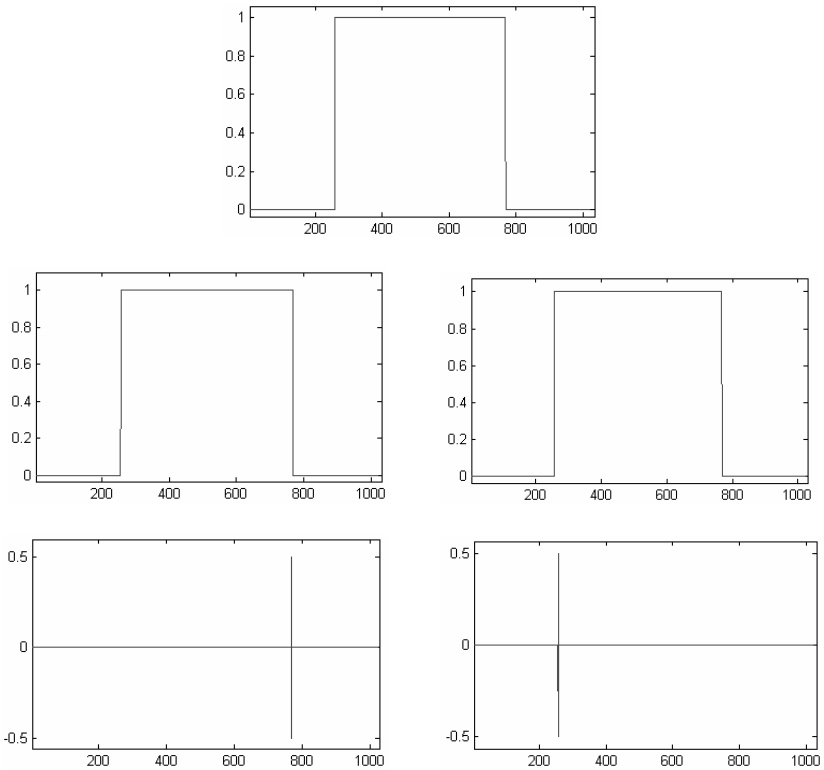


Figure 3.7. *At the top, the original signal: a square form signal. In the middle, two level 1 approximations and at the bottom, corresponding level 1 details*

The two analyses are very different, as the two details obtained show. In particular, none of them presents the two discontinuities simultaneously. For the first one (on the left), the detail is zero in the vicinity of position 256 and strong in the vicinity of 768, whereas for the second one (on the right) it is the opposite.

To restore invariance by translation the idea is to calculate the average of the DWT of all the suitably synchronized translated-periodized signals stemming from X (called ε-decimated DWT). This defines the translation invariant transform called SWT.

This property is interesting in several applications such as, for example, the detection of signal breakdown points. The principal application of the SWT is denoising (see Chapters 7 and 8). The principle consists of averaging several denoised signals. Each one of them is obtained using the usual denoising method,

but applied to the coefficients of an ε -decimated DWT. Hereafter the SWT is defined for signals whose size is divisible by 2^J, where J is the maximum level of decomposition.

3.9.1. ε *-decimated DWT*

There are several slightly different ways to use the DWT. Let us recall that the base of its calculation rests on a convolution followed by decimation. The latter preserves only the elements with an even index. In fact, decimation may very well be carried out preserving odd index elements instead of the even index elements and this alternative presents itself naturally at each step of the decomposition process. If the systematic calculation of all the possible decompositions of the original signal is carried out, we obtain 2^J decompositions at the J level.

To identify the choice made at the j stage, the variable ε_j is defined. We write $\varepsilon_j = 1$ (respectively 0), if at the j stage we decimate the elements with an odd (respectively even) index. Each decomposition is therefore identified by a sequence of 0 and 1: $\varepsilon = \varepsilon_1 \ldots \varepsilon_J$ and is called an ε-decimated DWT.

3.9.2. *Calculation of the SWT*

It is possible to obtain all the ε-decimated DWT for a given signal of a length N by calculating the approximation and detail coefficients for each possible sequence ε. This operation could be performed iteratively using a slightly modified version of the basic step of the DWT calculation. This is the standard version, for which $\varepsilon = 0$, but for $\varepsilon = 1$ the odd index elements will be retained during the decimation step. Of course, it is not the correct manner to calculate the ε -decimated DWT, since many calculations are carried out repeatedly. We will therefore describe another method: the translation invariant transform called SWT.

The algorithm used for the calculation of the SWT is very similar to that of the DWT. For level 1 all ε -decimated DWT (only two at this level) for a given signal can be obtained by convoluting the signal with the appropriate filters, as in the case of the DWT but without decimating. In this case the approximation and detail coefficients at level 1, noted here $a^{(1)}$ and $d^{(1)}$ (and no longer a^1 and d^1 as for the DWT) both have a size N, the size of the original signal. This operation can be represented by the diagram presented in Figure 3.8.

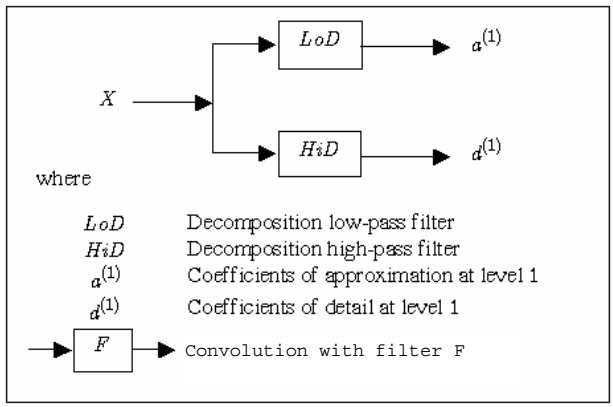


Figure 3.8. *1D SWT: decomposition at level 1*

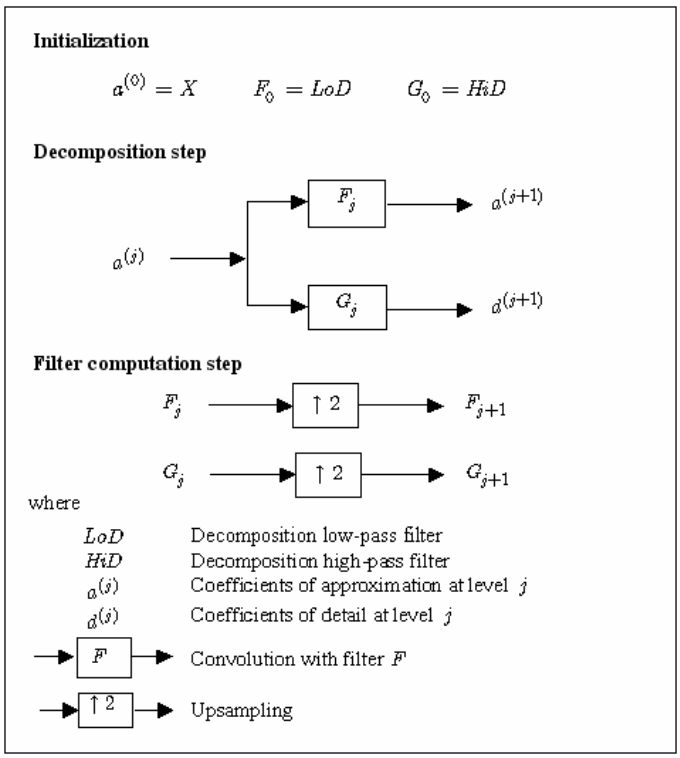


Figure 3.9. *1D SWT: decomposition algorithm*

In general, the $a^{(j)}$ approximation coefficients at the j level are convoluted with an upsampled version of the two usual filters to produce the approximation $a^{(j+1)}$ and detail $d^{(j+1)}$ coefficients at the $j+1$ level. The algorithm can be represented by the diagram presented in Figure 3.9.

Let us now show with a simple example how to extract an ε-decimated DWT on the basis of the structure containing the approximation and detail coefficients of the SWT. Let us decompose a sequence of eight numbers at the $J = 3$ level using the SWT with an orthogonal wavelet. The calculations carried out at each step by the algorithm are grouped in the following tables. In the latter we note the coefficients of approximation or detail $a(j,\varepsilon_1,\ldots,\varepsilon_j)$ or $d(j,\varepsilon_1,\ldots,\varepsilon_j)$ at the j level obtained for the ε-decimated DWT characterized by the sequence $\varepsilon = \left[\varepsilon_1,\ldots,\varepsilon_j\right]$.

Step 0 (origin data)

At the start the signal is identified with the 0 level approximation, and the eight coefficients are all noted *a*(0).

a (0)	*a* (0)	*a* (0)	*a* (0)	*a* (0)	*a* (0)	*a* (0)	*a* (0)

Step 1

By filtering process the algorithm decomposes the vector of *a*(0) into two: the first line contains the two details obtained for $\varepsilon_1 = 0$ and for $\varepsilon_1 = 1$, interlaced and noted $d\left(1,\varepsilon_1\right)$. Similarly, we find approximations in the second line.

d(1,0)	*d*(1,1)	*d*(1,0)	*d*(1,1)	*d*(1,0)	*d*(1,1)	*d*(1,0)	*d*(1,1)
a(1,0)	*a*(1,1)	*a*(1,0)	*a*(1,1)	*a*(1,0)	*a*(1,1)	*a*(1,0)	*a*(1,1)

Step 2

The algorithm no longer affects the first line of the table containing the level 1 details. On the other hand, the two approximations of level 1 are broken up into 2, and we construct:

– four level 2 details are the $d\left(2,\varepsilon_1,\varepsilon_2\right)$ and figuring in the second line;

– four level 2 approximations noted similarly.

d(1,0)	***d*(1,1)**	*d*(1,0)	***d*(1,1)**	*d*(1,0)	***d*(1,1)**	*d*(1,0)	***d*(1,1)**
d(2,0,0)	***d*(2,1,0)**	*d*(2,0,1)	*d*(2,1,1)	*d*(2,0,0)	***d*(2,1,0)**	*d*(2,0,1)	*d*(2,1,1)
a(2,0,0)	***a*(2,1,0)**	*a*(2,0,1)	*a*(2,1,1)	*a*(2,0,0)	***a*(2,1,0)**	*a*(2,0,1)	*a*(2,1,1)

Step 3

d(1,0)	***d*(1,1)**	*d*(1,0)	***d*(1,1)**	*d*(1,0)	***d*(1,1)**	*d*(1,0)	***d*(1,1)**
d(2,0,0)	***d*(2,1,0)**	*d*(2,0,1)	*d*(2,1,1)	*d*(2,0,0)	***d*(2,1,0)**	*d*(2,0,1)	*d*(2,1,1)
d(3,0,0,0)	*d*(3,1,0,0)	*d*(3,0,1,0)	*d*(3,1,1,0)	*d*(3,0,0,1)	***d*(3,1,0,1)**	*d*(3,0,1,1)	*d*(3,1,1,1)
a(3,0,0,0)	*a*(3,1,0,0)	*a*(3,0,1,0)	*a*(3,1,1,0)	*A*(3,0,0,1)	***a*(3,1,0,1)**	*a*(3,0,1,1)	*a*(3,1,1,1)

If $\varepsilon = \left[\varepsilon_1,\ldots,\varepsilon_J\right]$ with $\varepsilon_i = 0$ or 1 we then obtain $2^J = 8$ ε-decimated DWT. The choice of a sequence ε makes it possible to extract the corresponding ε-decimated DWT from the table representing the values of the SWT. Let us note C_ε the structure of decomposition of an ε-decimated DWT for a given ε. Then we can extract the latter from the SWT decomposition structure by selecting the appropriate coefficients in the following manner:

$C_\varepsilon =$

d(1,ε_1)	*d*(1,ε_1)	*d*(1,ε_1)	*d*(1,ε_1)
d(2,ε_1,ε_2)	*d*(2,ε_1,ε_2)		
d(3,ε_1,ε_2,ε_3)			
a(3,ε_1,ε_2,ε_3)			

For example, the ε-decimated DWT corresponding to $\varepsilon = \left[\varepsilon_1,\varepsilon_2,\varepsilon_3\right] = [1,0,1]$ is represented in bold characters in the sequence of tables presented in the previous example.

We may easily extend the previous results to the two-dimensional case and the SWT algorithm for images is represented by the diagram proposed in Figure 3.10.

Let us also note that the sizes of all the coefficient structures obtained by decomposition (approximation and details) are all equal to the size of the original image X.

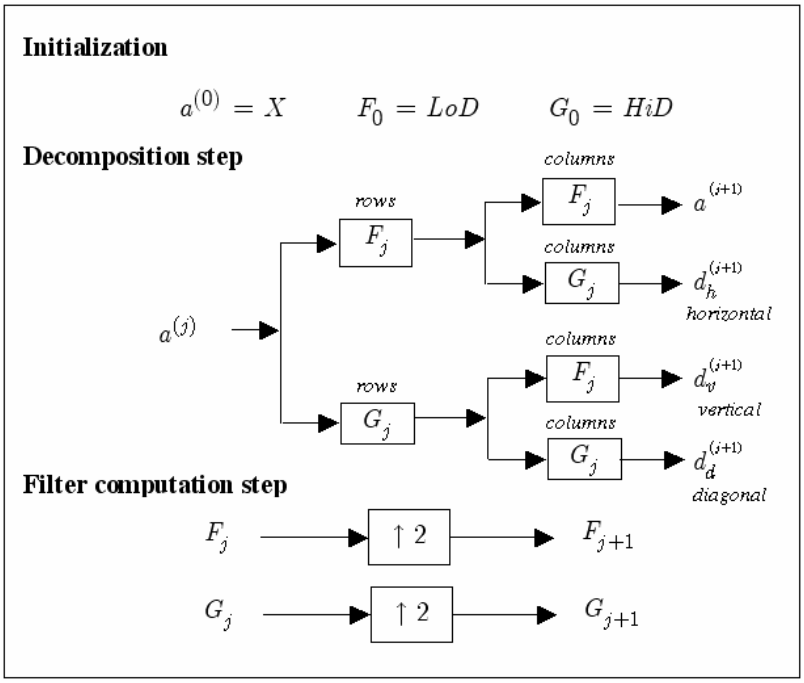


Figure 3.10. *2D SWT: decomposition algorithm*

3.9.3. *Inverse SWT*

Each ε-decimated DWT corresponding to a given ε can be inversed. We can thus reconstruct the original signal from a given ε-decimated DWT characterized by $\varepsilon = \left[\varepsilon_1,\ldots,\varepsilon_J\right]$.

The idea of the inverse SWT (ISWT) not detailed here is to average the obtained inverse forms for each ε-decimated DWT. This can be carried out recursively, moving from level J to level 1. Following the same principle this algorithm can be generalized to the two-dimensional case.

For more information on the SWT it would be useful to refer to [COI 95], [NAS 95] and [PES 96].

Chapter 4

Wavelet Families

4.1. Introduction

Wavelet analysis, as opposed to Fourier analysis, provides additional freedom since the choice of atoms of the transform deduced from the analyzing wavelet is left to the user. Moreover, according to the objectives of wavelet processing, we may prefer the continuous transform to the discrete transform, if the redundancy is useful for analyzing the signal. We would make the opposite choice, if we were looking for signal compression. In the latter case we must restrict ourselves to wavelets with filters, whereas in the former case almost any zero integral function is appropriate.

Since the Haar base appeared at the beginning of the last century, since renamed the Haar wavelet, passing by Gaussian Morlet wavelets, Meyer wavelets [MEY 90] (obtained using *ad hoc* construction) and Daubechies wavelets [DAU 88] and [DAU 92] that are the most widely used, numerous wavelets regularly appear in books and are made available in specialized software applications. Construction of new wavelets was very intense in the first ten years of their young history, but recently it has become less regular and bears on increasingly specific goals, often associated with limited application contexts. In Chapter 5 we will find some wavelet construction methods.

This chapter proposes a *mini-genealogy of well-known wavelet families.* Their presentation is organized according to the numerous properties that make it possible to differentiate between them and make a selection by prioritizing one aspect or another depending on the context and the objectives. For each wavelet family we

present the sketch of the construction process, the principal properties and the graphic representation of a typical wavelet of the family.

In the majority of cases, orthogonal or biorthogonal wavelets (which have a compact support and allow discrete decompositions using the fast algorithm) are directly defined by their associated filters. The wavelet, however, is not directly accessible, since no analytical formula defines it. Nonetheless, by using an algorithm straightforwardly deduced from the Mallat reconstruction algorithm, it is easy to obtain excellent approximations of the implicitly defined wavelet. The end of this chapter thus deals with the cascade algorithm which is also used in this work to build the graphs presenting wavelets and scaling functions associated with multi-resolution orthogonal or biorthogonal analyses.

4.2. What could we want from a wavelet?

There are several types and families of wavelets whose properties differ along the following principal criteria:

– the support of the functions ψ, $\hat{\psi}$ (and also φ, $\hat{\varphi}$) or speeds of convergence to 0 of the functions $\psi(t)$ and $\hat{\psi}(\omega)$ when the time t or the frequency ω tend to infinity, which quantifies the localization of the wavelet in time and frequency respectively;

– the symmetry, which is useful in order to avoid dephasing;

– the number of zero moments of ψ or φ, which is useful for compression;

– the regularity, which is useful in order to obtain reconstructed smooth and regular signals or images.

Other criteria are associated with two properties, which enable the use of the fast wavelet transform algorithm and parsimonious coding:

– the existence of the scaling function φ;

– the orthogonality or biorthogonality of the analysis stemming from it.

Two less crucial properties may be added: the existence of an explicit formula and the ease of calculation.

4.3. Synoptic table of the common families

Before detailing some common wavelet families, in Table 4.1 we draw up the list of those presented in this chapter with the associated abbreviations in order to make reading easier.

Names of the wavelet families	Abbreviation
Haar wavelet	*haar*
Daubechies wavelets	*Db*
Symlets	*sym*
Coiflets	*coif*
Biorthogonal wavelets	*bior*
Meyer wavelet	*meyr*
Discrete approximation of the Meyer wavelet	*dmey*
Battle and Lemarié wavelets	*btlm*
Gaussian wavelets	*gaus*
Mexican hat	*mexh*
Morlet wavelet	*morl*
Complex Gaussian wavelets	*cgau*
Complex Shannon wavelets	*shan*
Complex B-spline frequency wavelets	*fbsp*
Complex Morlet wavelets	*cmor*

Table 4.1. *List of presented wavelets*

With respect to the desirable properties stated in the first section of this chapter, these families of wavelets are distributed as indicated in Table 4.2. In this synoptic table the properties are found in rows and the wavelet families in columns; only the boxes concerned are ticked.

4.4. Some well known families

Now let us detail the definition and the construction of these families, grouping them according to four main properties: existence of associated filters, orthogonality or biorthogonality, compact or not compact support, real or complex wavelets. Table 4.3 summarizes these various properties.

	Wavelet families														
Properties	*morl*	*mexh*	*meyr*	*btlm*	*haar*	*dbN*	*SymN*	*coifN*	*biorNr.Nd*	*gaus*	*dmey*	*cgau*	*Cmor*	*fbsp*	*shan*
Admissible[1]	✓	✓								✓		✓	✓	✓	✓
Infinite regularity	✓	✓	✓							✓		✓	✓	✓	✓
Arbitrary regularity				✓		✓	✓	✓	✓						
Orthogonal with compact support					✓	✓	✓	✓							
Biorthogonal with compact support									✓						
Symmetry	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓
Asymmetry						✓									
Near symmetry							✓	✓							
Arbitrary number of zero moments						✓	✓	✓	✓						
Existence of φ			✓	✓	✓	✓	✓	✓	✓						
Zero moments for φ								✓							
Orthogonal analysis			✓	✓	✓	✓	✓	✓							
Biorthogonal analysis			✓	✓	✓	✓	✓	✓	✓						
Exact reconstruction	≃	✓	✓	✓	✓	✓	✓	✓	✓	✓	≃	✓	≃	✓	✓
FIR filters					✓	✓	✓	✓	✓		✓				
Continuous transformation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
Discrete transformation			✓	✓	✓	✓	✓	✓	✓		✓				
Fast algorithm					✓	✓	✓	✓	✓		✓				
Explicit expression	✓	✓			✓				*	✓		✓	✓	✓	✓
Complex wavelet												✓	✓	✓	✓
Complex continuous transformation												✓	✓	✓	✓
Approximation with FIR											✓				

≃ Nearly exact reconstruction * Explicit expression for splines

FIR: finite impulse response

Table 4.2. *Properties of the wavelet families*

1 See condition of admissibility of Theorem 2.2 in Chapter 2.

Wavelets with filters			**Wavelets without filters**	
With compact support		***With non-compact support***	***Real***	***Complex***
Orthogonal	***Biorthogonal***	***Orthogonal***	*gaus*, *mexh*, *morl*	*cgau*, *shan*, *fbsp*, *cmor*
dB, *haar*, *sym*, *coif*	*bior*	*meyr*, *dmey*, *btlm*		

Table 4.3. *Principal properties of wavelet families*

Wavelets with filters are associated with multi-resolution orthogonal or biorthogonal analyses; discrete transform and fast calculations using the Mallat algorithm are then possible. Wavelets without filter, on the other hand, are useful for the continuous wavelet transform.

4.4.1. *Orthogonal wavelets with compact support*

The strategy of construction of orthogonal wavelets with compact support motivated by the use of the fast Mallat algorithm is based on the direct action on the filter m_0 (see Chapter 2), which generates the scaling function. The work of Daubechies at the end of the 1980s and at the start of the 1990s has marked a decisive stage in the diffusion of wavelet methods.

4.4.1.1. *Daubechies wavelets: dbN*

This family of wavelets with one parameter, due to Daubechies (see [DAU 92] p. 115, 132, 194 and 242) is the first one to make it possible to handle orthogonal wavelets with compact support and arbitrary regularity. We will call *N* the order of the *dbN* wavelet.

This family contains the Haar wavelet, *db1*, which is the simplest and certainly the oldest of wavelets. It is discontinuous, resembling a square form.

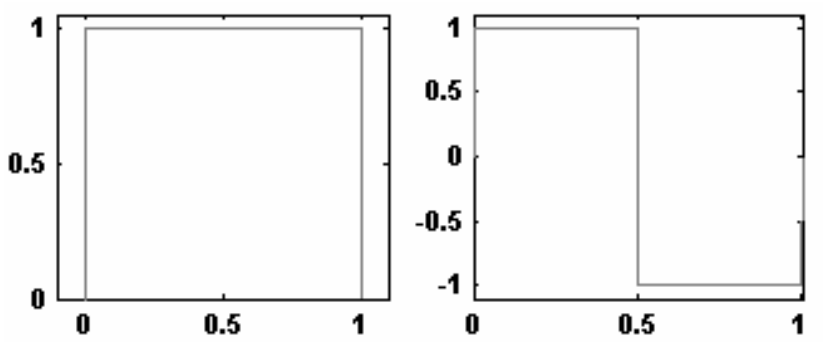


Figure 4.1. *The scaling function (on the left) and the wavelet (on the right): Haar*

The Haar wavelet is defined by:

$$\psi(x) = 1 \text{ if } x \in [0, 0.5[\,,\ \psi(x) = -1 \text{ if } x \in [0.5, 1[\text{ and 0 if not.}$$

The associated scaling function is the function:

$$\varphi(x) = 1 \text{ if } x \in [0,1] \text{ and 0 if not.}$$

Except for *db1*, the wavelets of this family do not have an explicit expression. However, the square modulus of the transfer function of the associated filter h is explicit and relatively simple:

let $P(y) = \sum_{k=0}^{N-1} C_{N-1+k}^{k}\, y^k$, where C_{N-1+k}^{k} are the binomial coefficients, then: $\left|m_0(\omega)\right|^2 = (\cos^2(\frac{\omega}{2}))^N P(\sin^2(\frac{\omega}{2}))$ where $m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k=0}^{2N-1} h_k e^{-ik\omega}$.

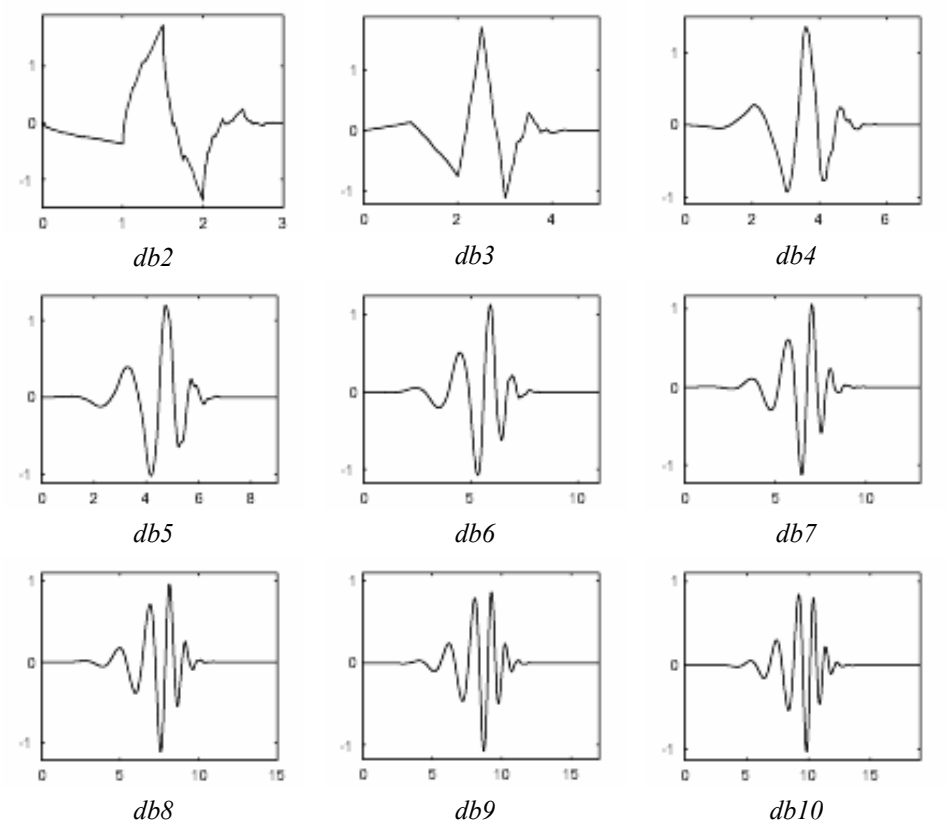


Figure 4.2. *Daubechies wavelets: dbN*

This family has the following properties:

– the ψ and φ support length is $2N-1$. The number of zero moments of ψ is *N*;

– *dbN* wavelets are asymmetric (in particular for low values of *N*) except for the Haar wavelet;

– the regularity increases with order. When *N* becomes very large, ψ and φ belong to $C^{\mu N}$ where $\mu \approx 0.2$. This value μN is too pessimistic for relatively small orders, as it underestimates the regularity;

– the analysis is orthogonal.

The wavelets of this family for the orders from 2 to 10 are presented in Figure 4.2. Moreover, for two of them (*db4* and *db8*), we also find in Figure 4.3, apart from the wavelet, the scaling function and the four associated filters (two for decomposition, two for reconstruction).

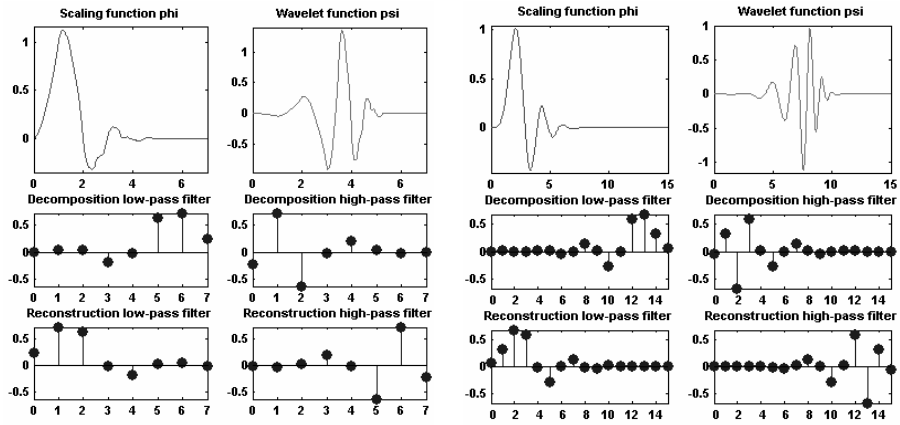


Figure 4.3. *Two Daubechies wavelets: db4 (on the left) and db8 (on the right)*

4.4.1.2. *Symlets: symN*

Symlets constitute a family of almost symmetric wavelets proposed by Daubechies by modifying the construction of the *dbN*. Apart from the symmetry, the other properties of the two families are similar. Symlets of orders 2 to 8 (*sym1* is simply the Haar wavelet) are presented in Figure 4.4.

The idea of construction consists of re-using the m_0 function introduced for *dbN*, considering $|m_0(\omega)|^2$ as a function W of the variable $z = e^{i\omega}$. We can factorize W in various manners in the form of $W(z) = U(z)\overline{U(z^{-1})}$, since the roots of W with module different from 1 go in pairs: if z_1 is a root then z_1^{-1} is also a root. By constructing U so that its roots are all of module < 1 we construct the Daubechies wavelets *dbN*. The filter U has a minimal phase. Another option, attained by optimizing factorization so that the filter U has an almost linear phase, produces much more symmetric filters: the symlets.

For two of them (*sym4* and *sym8*), in Figure 4.5 we find the wavelet, the scaling function and the four associated filters.

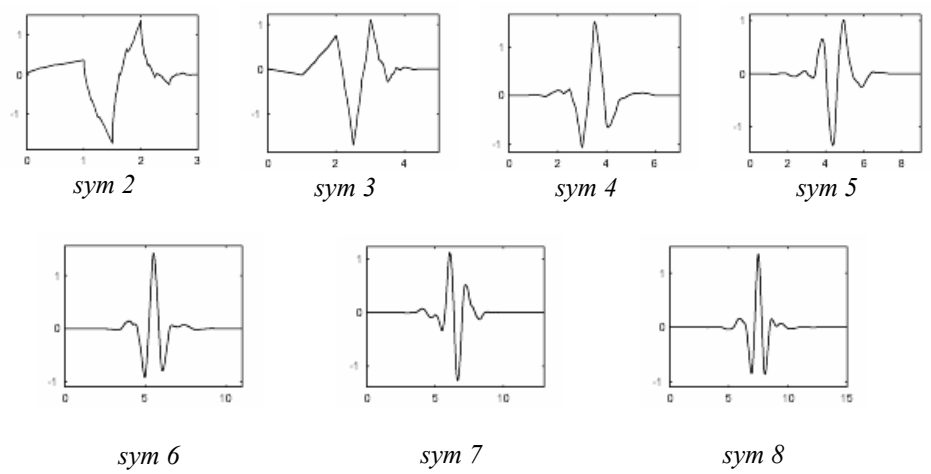


Figure 4.4. *Symlets: symN*

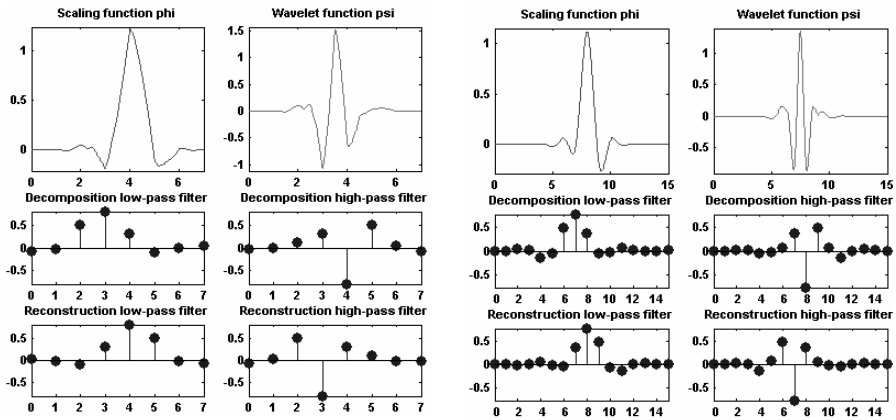


Figure 4.5. *Two symlets: sym4 (on the left) and sym8 (on the right)*

4.4.1.3. *Coiflets: coifN*

Constructed by Daubechies at the request of Coifman (see [DAU 92] p. 258), coiflets constitute a family of wavelets with an unusual property. Not only, as for the two previous families, the wavelet ψ associated with *coifN* has $2N$ zero moments, but also the scaling function φ that has an integral equal to 1 has $2N-1$ zero moments. The two functions φ and ψ have a support with a length of $6N-1$. Coiflets of orders 1 to 5 are represented in Figure 4.6.

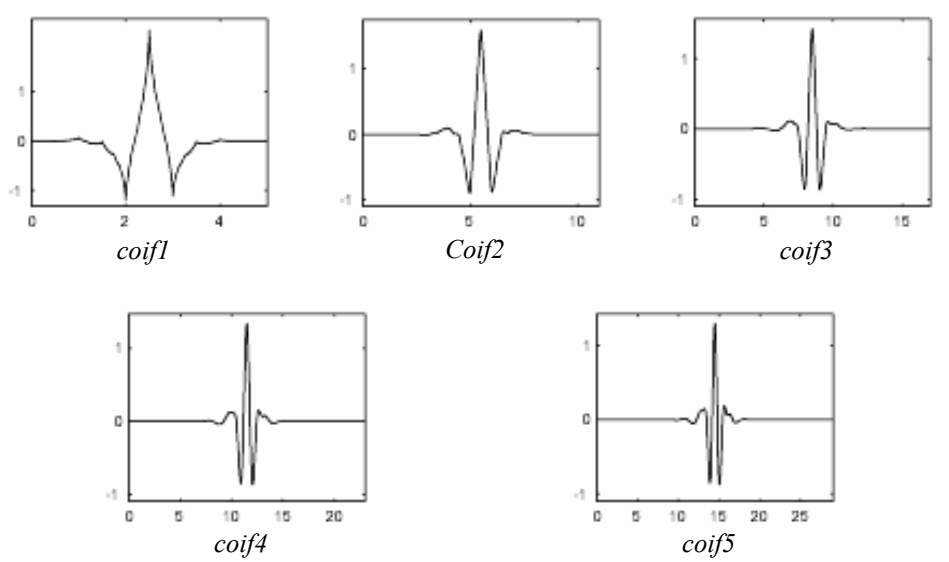


Figure 4.6. *Coiflets: coifN*

For two of them (*coif3* and *coif5*) in Figure 4.7 we find the wavelet, the scaling function and the four associated filters.

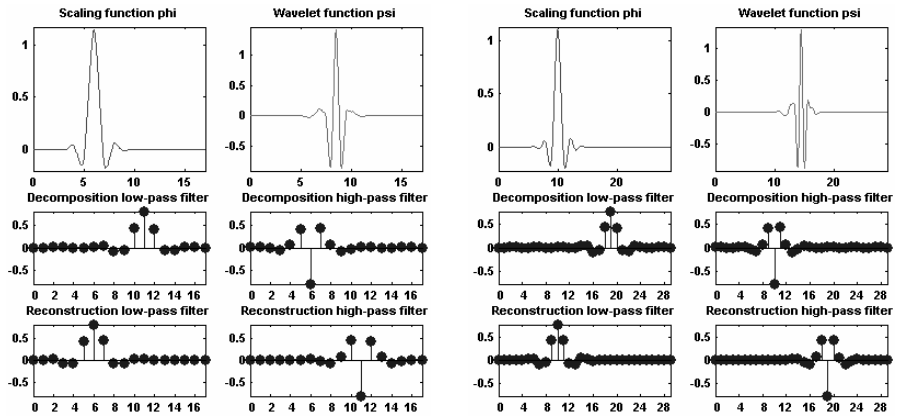


Figure 4.7. *Two coiflets: coif3 (on the left) and coif5 (on the right)*

As we can see, coiflets, as well as the associated scaling functions, are much more symmetric than the Daubechies wavelets *dbN*.

With respect to the length of the support, *coifN* can be compared to *db3N* or *sym3N*. From the point of view of the number of zero moments of ψ, *coifN* can be compared to *db2N* or *sym2N*.

The principal interest of coiflets lies in the following property: if f is a sufficiently regular function, then the approximation coefficient $\left\langle f, \varphi_{j,k} \right\rangle$ for a sufficiently large j is estimated well by $2^{-j/2} f(2^{-j}k)$, i.e. a sampled value of the function f. Moreover, if f is a polynomial of a degree equal to $N-1$, the estimate corresponds to the exact value. This property is used to control the difference between the development of a given signal on $\varphi_{j,k}$ and a sampled version of the signal.

4.4.2. *Biorthogonal wavelets with compact support: bior*

Biorthogonal wavelets extend the families of orthogonal wavelets (see Chapter 2). It is a well-known fact in the filter theory community that symmetry and perfect reconstruction are incompatible (except for the Haar wavelet) when the same FIR filters are used for decomposition and for reconstruction process. To circumvent this difficulty two wavelets are introduced instead of one:

– the first one, $\tilde{\psi}$, is used for analysis, and the coefficients of a signal s are $\tilde{c}_{j,k} = \int s(x)\tilde{\psi}_{j,k}(x)dx$;

– the other one, ψ, is used for synthesis: $s = \sum_{j,k} \tilde{c}_{j,k}\psi_{j,k}$.

The wavelets ψ and $\tilde{\psi}$ are linked by duality relations:

$$\int \tilde{\psi}_{j,k}(x)\psi_{j',k'}(x)dx = 0 \quad \text{for } k \neq k' \text{ or } \quad j \neq j'$$

$$\int \tilde{\varphi}_{0,k}(x)\varphi_{0,k'}(x)dx = 0 \quad \text{for } k \neq k'$$

It thus becomes possible to concentrate the properties desirable for analysis (the number of zero moments for example) in the wavelet $\tilde{\psi}$, while the properties interesting for synthesis (regularity, symmetry) can be concentrated in ψ.

Figure 4.8 presents biorthogonal wavelets constructed by Daubechies (see [DAU 92] p. 259 and 262-285). For each of them we find the graphs of the ψ and $\tilde{\psi}$ functions represented in Figure 4.8.

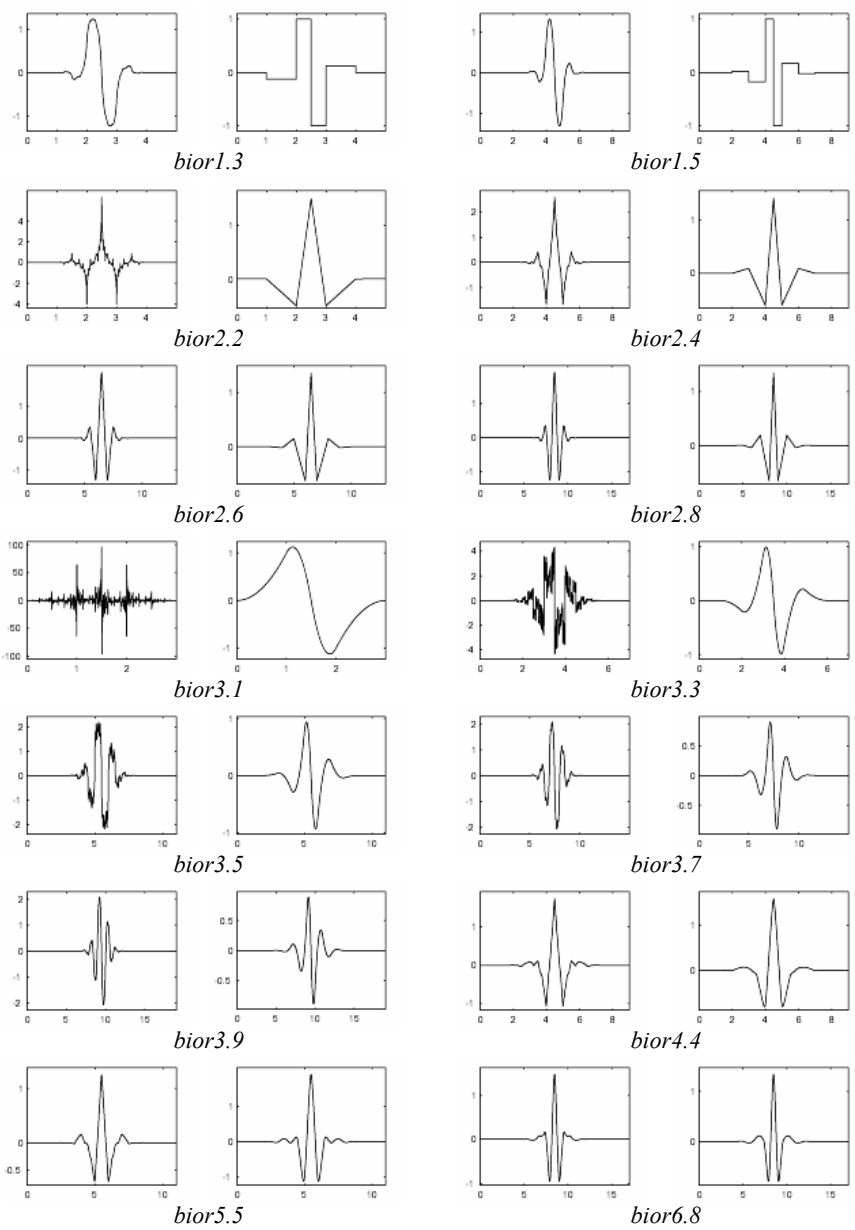


Figure 4.8. *Biorthogonal wavelets with compact support: bior*

As we can see, the ψ and $\tilde{\psi}$ functions can have very different regularities. The wavelets ψ and $\tilde{\psi}$, as well as the scaling functions φ and $\tilde{\varphi}$, have compact support generating two dual multi-resolution analyses which enable the use of the fast Mallat algorithm for calculations. Let us note that wavelets $\tilde{\psi}$ and ψ of *biorNr.Nd* have *Nr* and *Nd* zero moments respectively.

Let us note, finally, that the roles of the pairs (φ, ψ) and $(\tilde{\varphi}, \tilde{\psi})$ are symmetric and that they may therefore be exchanged to obtain another couple of biorthogonal wavelets.

4.4.3. *Orthogonal wavelets with non-compact support*

4.4.3.1. *The Meyer wavelet: meyr*

The *meyr* wavelet is one of the first wavelets. It was constructed by Meyer in the mid-1980s (see [DAU 92] p. 117, 119, 137 and 152).

It is an infinitely derivable orthogonal wavelet but it does not have compact support. The graphs of the ψ and φ functions are represented in Figure 4.9.

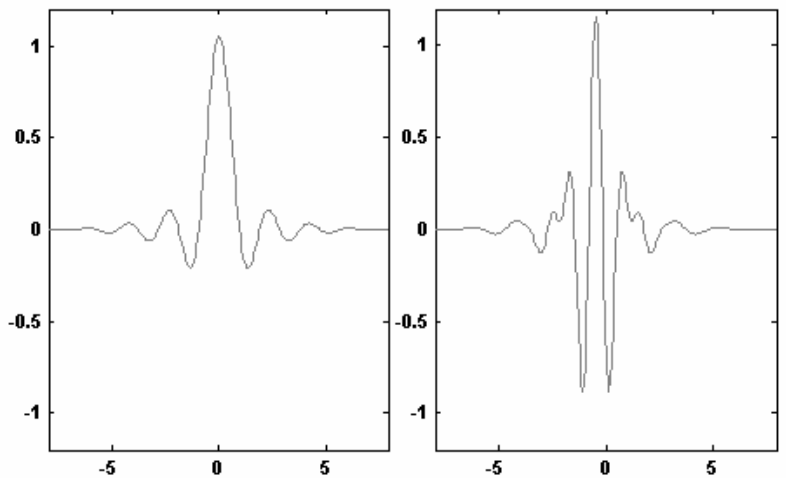


Figure 4.9. *Meyer wavelet: scaling function (on the left) and wavelet (on the right)*

The ψ and φ functions are both defined in the frequency domain, using an auxiliary function noted ν:

$$\hat{\psi}(\omega) = \begin{cases} (2\pi)^{-1/2} e^{i\omega/2} \sin\left(\frac{\pi}{2}\nu\left(\frac{3}{2\pi}|\omega| - 1\right)\right) & \text{if } \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3} \\ (2\pi)^{-1/2} e^{i\omega/2} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{4\pi}|\omega| - 1\right)\right) & \text{if } \frac{4\pi}{3} \leq |\omega| \leq \frac{8\pi}{3} \\ 0 & \text{if } |\omega| \notin \left[\frac{2\pi}{3}, \frac{8\pi}{3}\right] \end{cases}$$

where $\nu(a) = a^4(35 - 84a + 70a^2 - 20a^3)$ with $a \in [0,1]$.

$$\hat{\varphi}(\omega) = \begin{cases} (2\pi)^{-1/2} & \text{if } |\omega| \leq \frac{2\pi}{3} \\ (2\pi)^{-1/2} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2\pi}|\omega| - 1\right)\right) & \text{if } \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3} \\ 0 & \text{if } |\omega| \geq \frac{4\pi}{3} \end{cases}$$

Changing the auxiliary function ν we obtain a family of functions that generate an orthogonal analysis under certain conditions for ν.

The function ψ is not compactly supported, but when $x \to \infty$ it converges to 0 as in the next formula:

$$\forall n \in \mathbb{N}, \exists C_n \text{ such that } |\psi(x)| \leq C_n(1 + |x|^2)^{-n}$$

Moreover, ψ is infinitely derivable and the same property of decrease applies to each of the derivatives. The ψ function is thus rapidly decreasing.

4.4.3.2. *An approximation of the Meyer wavelet: dmey*

Since the ψ function does not have compact support, it is difficult to use it in practice because it requires the use of infinite impulse response filters. There are, nevertheless, FIR approximations of this wavelet enabling fast decomposition (see [ABR 95] p. 268). In Figure 4.10, in addition to the wavelet, we find the scaling function and the four associated filters.

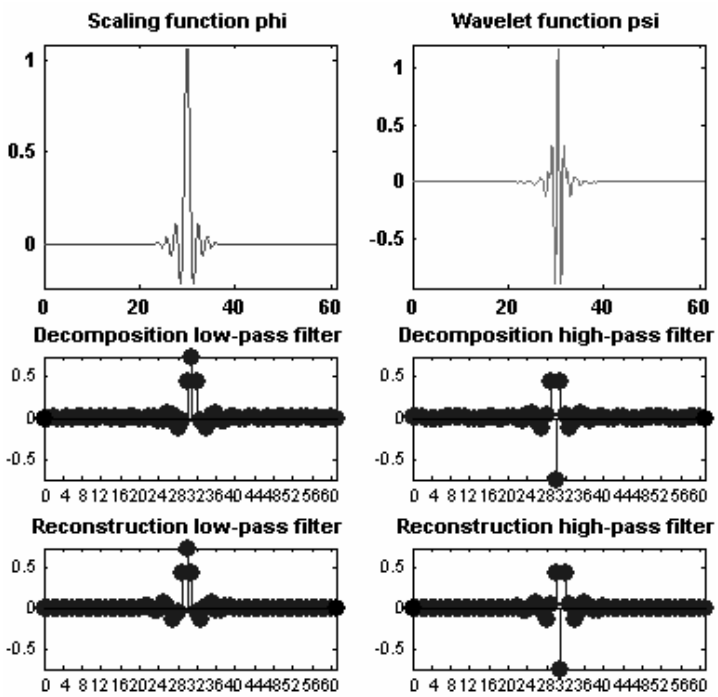


Figure 4.10. *FIR approximation of the Meyer wavelet*

4.4.3.3. *Battle and Lemarié wavelets: btlm*

The Battle and Lemarié family of wavelets is one of the first and appeared towards the end of the 1980s. Its construction is based on the spline functions (see [DAU 92] p. 146-151). In fact, there exist two forms of this wavelet: one providing an orthogonal analysis, the other not.

For $N = 1$ the scaling functions are linear splines. For $N = 2$ the scaling functions are compactly supported quadratic splines. More generally, for any *n*, $n \in N$, N, B-splines of degree N are defined by:

$$\hat{\varphi}(\omega) = (2\pi)^{-1/2} e^{i\kappa\omega/2} \left[\frac{\sin(\omega/2)}{\omega/2} \right]^{N+1} \qquad \omega \in \mathbb{R}$$

where $\kappa = 0$ if N is odd and $\kappa = 1$ if N is even.

This formula can be used to construct filters. For $N = 2M$ or $N = 2M + 1$ the two-scale relation is written as:

$$\varphi(x) = 2^{-N} \sum_{j=0}^{N+1} C_{N+1}^{j} \varphi(2x - M - 1 + j)$$

This family of wavelets also has the following properties:

– for even (or odd) N, φ is symmetric with respect to $x = 1/2$ (or $x = 0$) and ψ is antisymmetric (or symmetric) with respect to $x = 1/2$;

– the analysis becomes orthogonal if the φ and ψ functions are appropriately transformed. For $N = 1$, for example, we write:

$$\widehat{\varphi^{\perp}}(\omega) = 3^{1/2}(2\pi)^{-1/2} \frac{4\sin^2(\omega/2)}{\omega^2\left(1 + 2\cos^2(\omega/2)\right)^{1/2}};$$

– the support of φ is compact but the supports of ψ and of $\varphi^{\perp}$ are not. Nonetheless, the decrease when time goes to infinity is exponential;

– the ψ wavelets are $N - 1$ times derivable.

4.4.4. *Real wavelets without filters*

In certain situations the capacity for parsimonious coding is a secondary objective. The capacity of accurate and local analysis of the functions requires redundant time and scale analyses. For this reason we discard the requirement of generating bases and concentrate on wavelets that verify very weak admissibility conditions. These wavelets do not have an associated scaling function and thus have no filters, but have the advantage of having an explicit definition formula. We provide some examples of such wavelets with real values.

4.4.4.1. *The Mexican hat: mexh*

It is a function proportional to the second derivative of the Gaussian density function (see [DAU 92] p. 75):

$$\psi(x) = \left(\frac{2}{\sqrt{3}}\pi^{-1/4}\right)\left(1 - x^2\right)e^{-x^2/2}$$

It oscillates very little as we can see from its graph represented in Figure 4.11.

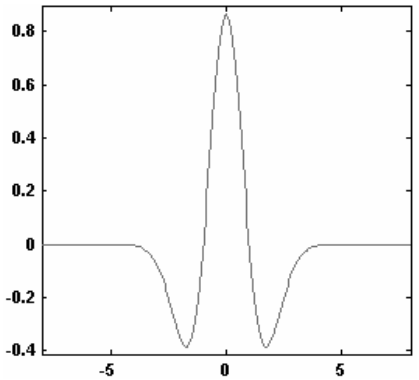


Figure 4.11. *Mexican hat mexh*

4.4.4.2. *The Morlet wavelet: morl*

This function verifies the condition of admissibility only roughly; despite the fact that it is a standard function (see [DAU 92] p. 76) defined by:

$$\psi(x) = Ce^{-x^2/2}\cos(5x)$$

where C is a normalization constant used for the reconstruction.

Its graph is represented in Figure 4.12 and it shows that it oscillates much more than the Mexican hat.

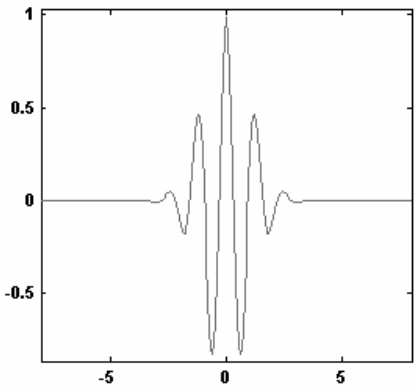


Figure 4.12. *Morlet wavelet: morl*

4.4.4.3. *Gaussian wavelets: gausN*

We may generalize the Mexican hat by introducing the one-parameter family of Gaussian derivatives. We start with $f(x) = C_p e^{-x^2}$ and consider the p order derivative of f sequence C_p so that $\left\| f^{(p)} \right\|^2 = 1$. The graph of the 8th Gaussian derivative is represented in Figure 4.13.

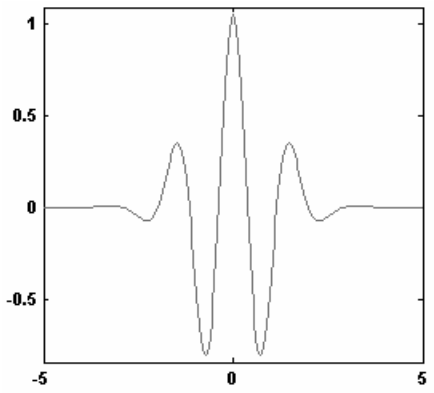


Figure 4.13. *Eighth derivative of a Gaussian function: gaus8*

4.4.5. *Complex wavelets without filters*

In this section we examine some wavelets or wavelet families with complex values, which can be found in the book by Teolis (see [TEO 98] p. 62-65).

4.4.5.1. *Complex Gaussian wavelets: cgau*

We can extend the real one-parameter family of Gaussian derivatives to the complex case. We start with $f(x) = C_p e^{-ix} e^{-x^2}$ and consider, as previously, the order p derivative of f. Here are the real and imaginary parts, the modulus and the angle of the wavelet of this family for $p = 8$.

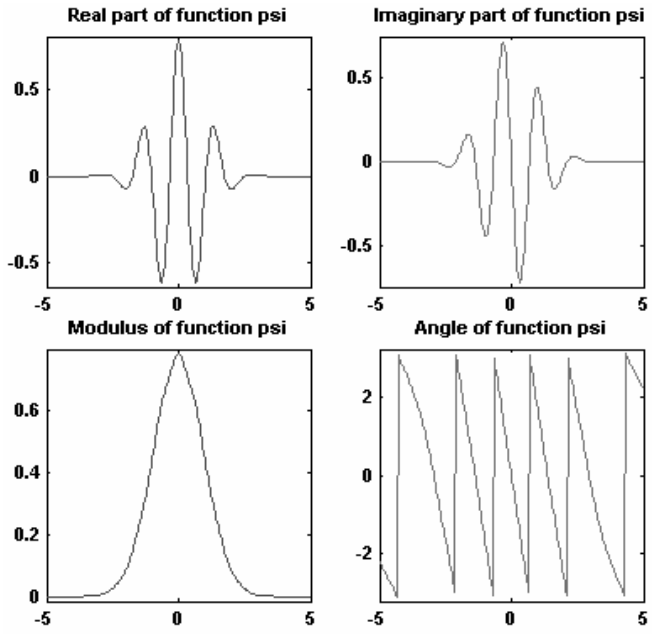


Figure 4.14. *Eighth derivative of a complex Gaussian function: cgaus8*

The real part of ψ coincides with the corresponding real wavelet up to a normalization constant.

4.4.5.2. *Complex Morlet wavelets: cmorl*

A complex Morlet wavelet is defined by:

$$\psi(x) = \frac{1}{\sqrt{\pi f_b}} e^{2i\pi f_c x} e^{-\frac{x^2}{f_b}}$$

and depends on two parameters: f_b is a window parameter and f_c is the central frequency of the wavelet. Here, in Figure 4.15, are the real and imaginary parts, the module and the angle of the Morlet wavelet for $f_b = 1.5$ and $f_c = 1$.

These functions only roughly verify the admissibility condition and are therefore not wavelets in the strict sense of the definition. Nevertheless, they are classical and commonly used.

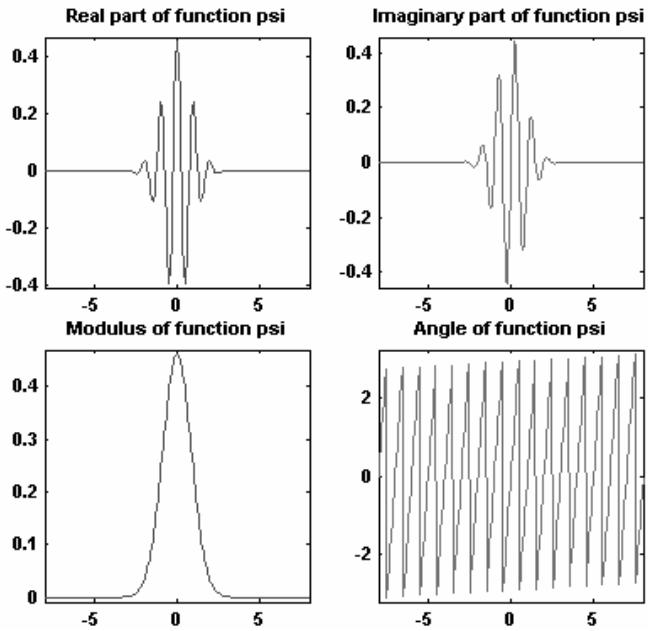


Figure 4.15. *The complex Morlet wavelet: cmor associated with* $f_b = 1.5$ *and* $f_c = 1$

4.4.5.3. *Complex frequency B-spline wavelets: fbsp*

A complex frequency B-spline wavelet is defined by:

$$\psi(x) = \sqrt{\frac{f_b}{m}} \left(\frac{\sin\left(\frac{\pi f_b x}{m}\right)}{\frac{\pi f_b x}{m}} \right)^m e^{2i\pi f_c x}$$

and depends on three parameters: the order of the wavelet given by an integer $m > 1$, f_b is a window parameter and f_c the central wavelet frequency. In Figure 4.16 we find the real and imaginary parts, the modulus and the angle of the wavelet obtained for $m = 2$, $f_b = 1$ and $f_c = 0.5$.

Let us mention, finally, that if we sequence $m = 1$, we obtain a wavelets family with two parameters, called Shannon wavelets.

We have introduced the usual wavelet families and outlined how a number of them are constructed, in particular, orthogonal wavelets. These are, in general, implicitly defined by filters, but do not have an analytical expression. Nonetheless, we can visualize them or, more exactly, visualize excellent approximations using the cascade algorithm.

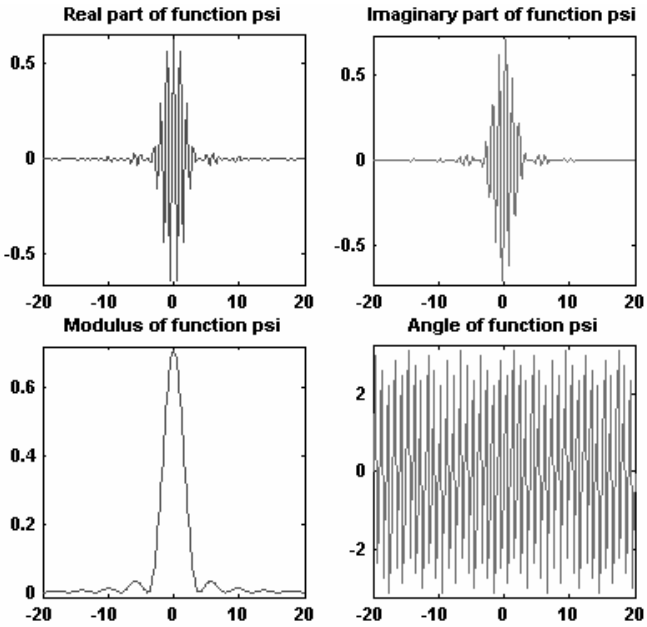


Figure 4.16. *The complex frequency B-spline wavelet: fbsp associated with* $m = 2$, $f_b = 1$ *and* $f_c = 0.5$

4.5. Cascade algorithm

In this section we present the cascade algorithm that, starting only from the associated filters φ and ψ, makes it possible to calculate and thus to trace the scaling function and the wavelet stemming from a multi-resolution analysis.

This algorithm consists of repeatedly applying the reconstruction phase of the Mallat algorithm, starting with a sequence that is 0 except for one coefficient. We start with the justification for the algorithm. The links between the cascade algorithm and the reconstruction phase of the Mallat algorithm are established. We then illustrate the behavior of the algorithm by visualizing the successive approximations of the scaling function and the wavelets associated with the *db2*

filter. In fact, the algorithm approaches the values of the two functions at the dyadic points of $\mathbb{R}$. For this reason we eventually provide a result of approximation and convergence.

4.5.1. *The algorithm and its justification*

Let us place ourselves in the context of Chapter 3. Spaces V_j and W_j are provided with orthonormal bases $\left\{\varphi_{j,k}\right\}_{k\in\mathbb{Z}}$ and $\left\{\psi_{j,k}\right\}_{k\in\mathbb{Z}}$ where φ and ψ indicate the scaling function and the wavelet. For a function $f \in L^2$ the orthogonal projections of f on V_j and W_j, respectively noted $A^j = P_{V_j} f$ and $D^j = P_{W_j} f$, are given by the relations below:

$$A^j = \sum_{p\in\mathbb{Z}} a_p^j \varphi_{j,p} \text{ with } a_p^j = \left(A^j, \varphi_{j,p}\right)_{L^2} = \left(f, \varphi_{j,p}\right)_{L^2}$$

$$D^j = \sum_{p\in\mathbb{Z}} d_p^j \psi_{j,p} \text{ with } d_p^j = \left(D^j, \psi_{j,p}\right)_{L^2} = \left(f, \psi_{j,p}\right)_{L^2}$$

The algorithm is based on the following observation. The scaling function φ is the single function $f \in L^2$ that satisfies:

$$\left(f, \varphi_{0,p}\right)_{L^2} \text{ if } p = 0 \quad \text{and for any } p, p \in \mathbb{Z}^*$$

$$\left(f, \psi_{j,p}\right)_{L^2} = 0\,,\ \forall \mathrm{p} \in \mathbb{Z}^*\,,\ \forall j \le 0$$

in other words: $a_0^0 = 1\,,\ a_p^0 = 0\,,\ \forall p \neq 0$ and $d_p^{-j} = 0\,,\ \forall p, \forall j \ge 0\,.$

We thus have:

$$\varphi \in V_0 \subset V_{-1} \subset \cdots \subset V_{-j} \subset \cdots\,,\ \forall j > 0 \quad \text{and} \quad \varphi \perp W_{-j}\,,\ \forall j \ge 0$$

or, equivalently:

$$\varphi = A^0 = A^{-1} = \cdots = A^{-j} = \cdots, \forall j > 0 \quad \text{and } D^{-j} = 0 \ , \ \forall j \geq 0 \, .$$

Consequently, by using successively $V_0, V_{-1}, \ldots, V_{-j}$, etc., we obtain:

$$\varphi = \sum_{p \in \mathbb{Z}} a_p^0 \, \varphi_{0,p} = \sum_{p \in \mathbb{Z}} a_p^{-1} \, \varphi_{-1,p} = \cdots = \sum_{p \in \mathbb{Z}} a_p^{-j} \, \varphi_{-j,p} = \cdots$$

In addition, for a fixed $J > 0$ the reconstruction phase of the Mallat algorithm makes it possible to calculate the components of A^{-J} using those of A^{-J+1} and D^{-J+1} (see relation [3.15] in Chapter 3):

$$a_m^{-J} = \sum_{p \in \mathbb{Z}} a_p^{-J+1} [LoR]_{m-2p} + \sum_{p \in \mathbb{Z}} d_p^{-J+1} [HiR]_{m-2p}$$

where LoR and HiR are the reconstruction filters associated with the scaling function and the wavelet. Since the detail D^{-J+1} is equal to zero:

$$a_m^{-J} = \sum_{p \in \mathbb{Z}} a_p^{-J+1} [LoR]_{m-2p}$$

The process is iterated until level 0, using the fact that all the details are zero.

Ultimately, starting from the sequence $\left\{ a_0^0 = 1 \, , \, a_p^0 = 0 \, , \, \forall p \neq 0 \right\}$ and applying the reconstruction phase of the Mallat algorithm with zero components for details J times, we calculate the sequence $\left\{ a_m^{-J} \right\}_m$. As we shall see in the last section, the sequence $\left\{ 2^{1/2} a_m^{-J} \right\}_m$ is then an "approximation" of the sequence of values $\left\{ \varphi\left(m 2^{-J} \right) \right\}_m$. For instance, if φ is Hölderian of the order α over $\mathbb{R}$, the approximation error is less than $D 2^{-J\alpha}$. This means that very good precision is reached after few iterations, say $J = 10$, as in the example in the following section.

Similarly, we construct an approximation of the wavelet ψ. Indeed, the function ψ is the single function $f \in L^2$ that verifies:

$$\left(f, \psi_{0,p}\right)_{L^2} = 1 \text{ if } p = 0 \quad \text{and} \quad \left(f, \psi_{0,p}\right)_{L^2} = 0 \text{ if } p \neq 0$$

$$\left(f, \psi_{j,p}\right)_{L^2} = 0 \,,\, \forall \mathrm{p} \in \mathbb{Z}^* \,,\, \forall j < 0$$

consequently:

$$d_0^0 = 1 \,,\, d_p^0 = 0 \,,\, \forall p \neq 0 \quad \text{and} \quad d_p^{-j} = 0 \,,\, \forall p, \forall j > 0$$

$$\psi \in W_0 \subset V_{-1} \subset \cdots \subset V_{-j} \subset \cdots \,,\, \forall j > 0 \quad \text{and} \quad \psi \perp W_{-j} \,,\, \forall j > 0$$

$$\psi = D^0 = A^{-1} = \cdots = A^{-j} = \cdots, \forall j > 0 \quad \text{and} \quad D^{-j} = 0 \,,\, \forall j > 0$$

The construction of the approximation of ψ is then similar to that of φ; only the initialization changes: we start here with the sequence $\left\{ d_0^0 = 1 \,,\, d_p^0 = 0 \,,\, \forall p \neq 0 \right\}$.

4.5.2. *An application*

Let us apply the algorithm to calculate successive approximations of the scaling function and wavelet associated with the filter *db2* that has a length of 4. In Figure 4.17, at the top, we present the approximations of the scaling function φ and at the bottom, those of the wavelet after 2, 3, 4 and 10 iterations of the cascade algorithm.

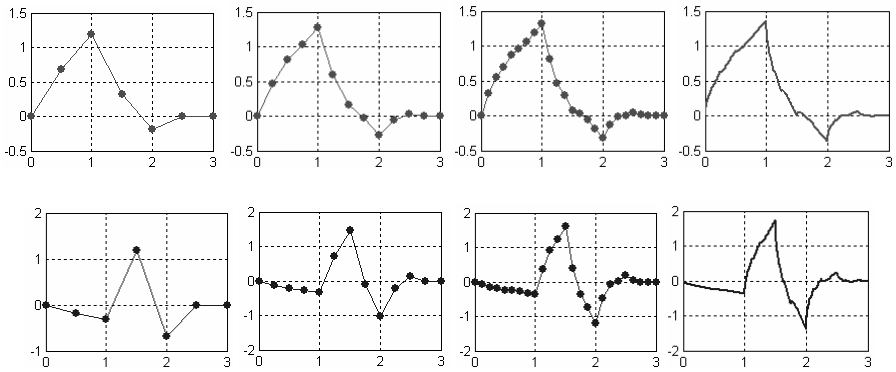


Figure 4.17. *Approximations of the scaling function φ and the wavelet ψ associated with the db2 filter after 2, 3, 4 and 10 iterations of the cascade algorithm*

4.5.3. *Quality of the approximation*

The cascade algorithm is based on the following theoretical approximation (or regularization) result.

Let us consider a function φ in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, whose integral over $\mathbb{R}$ is equal to 1. To simplify, we can also suppose that φ is real and has a compact support.

For a function f of $\mathbb{R}$ to $\mathbb{R}$, we define, at least formally, a "regularizing" sequence by:

$$\theta_j(x) = 2^j \int_{\mathbb{R}} f(x+y)\varphi\left(2^j y\right) dy = 2^{j/2} \int_{\mathbb{R}} f(z)\left[2^{j/2}\varphi\left(2^j (z-x)\right)\right] dz\,,\ j \in \mathbb{N}$$

The convergence of the sequence $\left\{\theta_j\right\}$ to f depends on the properties of f. In particular, we have the following results:

– if $f \in L^2(\mathbb{R})$, then θ_j tends to f in the space L^2. Moreover, if f is continuous at the point x_0, then $\theta_j(x_0)$ tends to $f(x_0)$;

– if $f \in C^0(\mathbb{R})$, then for any x_0 of $\mathbb{R}$, $\theta_j(x_0)$ tends to $f(x_0)$;

– if f is uniformly continuous over $\mathbb{R}$, then θ_j tends to f uniformly over $\mathbb{R}$;

– if f is a Holder function of order α over $\mathbb{R}$, i.e., if for any pair of real (x, y) we have $|f(x) - f(y)| \leq C|x - y|^{\alpha}$, then $\left|f(x) - \theta_j(x)\right| \leq D\,2^{-j\alpha}$ for all x in $\mathbb{R}$.

One uses the previous result setting $f = \varphi$ and considers the dyadic points of $\mathbb{R}$, that is, points of the form $x = K\,2^{-N}, K \in \mathbb{Z}, N \in \mathbb{N}$.

If x is a point of continuity of φ, then:

$$\begin{aligned}\varphi\left(K2^{-N}\right) &= \lim_{j\to+\infty} 2^{j/2}\int_{\mathbb{R}} f(z)\left[2^{j/2}\varphi\left(2^j z - K\,2^{j-N}\right)\right]dz \\ &= \lim_{j\to+\infty} 2^{j/2}\left(\varphi, \varphi_{-j,K2^{j-N}}\right)_{L^2}\end{aligned}$$

where $\varphi_{j,k}(t) = 2^{-j/2}\varphi\left(2^{-j}t - k\right)$.

If φ is an Holder order α over $\mathbb{R}$, there also exists an estimate of the approximation quality is provided by:

$$\left|\varphi\left(K2^{-N}\right) - 2^{j/2}\left(\varphi, \varphi_{-j,K2^{j-N}}\right)_{L^2}\right| \leq D\,2^{-j\alpha}$$

The cascade algorithm allows the sample calculation of the terms of the form of $\left(\varphi, \varphi_{-j,m}\right)_{L^2}$, provided that m is an integer. Let us note that, once $j \geq N$, the number $K\,2^{j-N}$ is an integer. We have a similar result for ψ.

Chapter 5

Finding and Designing a Wavelet

5.1. Introduction

It is possible that for certain applications we do not find a suitable wavelet among the known options. It is then natural to try to produce a new wavelet adapted to the specific problem being treated. In this chapter we are interested in two wavelet construction processes: one is useful for continuous analysis and the other is linked to the discrete case.

In the first part of this chapter we consider the construction of wavelets usable for continuous analysis. This problem does not present great difficulty as the requirements to obtain such wavelets are relatively limited. We proceed in the following manner: starting from a pattern which, in general, is not a wavelet, we seek the wavelet nearest to the given form in the least squares sense. The usefulness of the construction process is demonstrated when this is applied to a detection problem. The task consists of employing the adapted wavelet to identify in a signal the patterns stemming from the basic form via translation and dilation.

In the second part of this chapter we tackle the construction of wavelets for discrete analyses, which is, in turn, more delicate. Until a few years ago, it was still only a subject for specialist discussion. Recently, the lifting method developed by Sweldens (see [SWE 98]) has made this task easier. It makes it possible to construct an infinite number of biorthogonal wavelet bases starting from a given biorthogonal wavelet base.

In fact, the discussion of this issue is more naturally organized starting from the concept of two-channel filter banks having the property of perfect reconstruction. Indeed, lifting enables the construction of an infinite number of filters on the basis of such a filter bank. The question of knowing the conditions under which an obtained filter bank is associated with a biorthogonal wavelet base deserves to be separated from the description of lifting.

Consequently, we tackle the three topics outlined above:

– we begin with filter banks. First of all, we set the rule that a two-channel filter bank has the property of perfect reconstruction and we set the construction algorithms for such filters;

– we then introduce lifting which make it possible to generate an infinite number of such filter banks and to parameterize them;

– finally, we characterize filter banks associated with biorthogonal wavelet bases. A number of examples are provided at the end of this chapter.

5.2. Construction of wavelets for continuous analysis

In this section, we first present the construction of a new wavelet and then this method is applied to a detection problem.

5.2.1. *Construction of a new wavelet*

5.2.1.1. *The admissibility condition*

The continuous wavelet transform of a function f (see Chapter 2) is calculated using the formula:

$$C_f(a,b) = \int_{\mathbb{R}} f(t)\overline{\psi_{a,b}(t)}dt = \left(f, \psi_{a,b}\right)_{L^2}$$

with:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right),\ b \in \mathbb{R} \text{ and } a \in \mathbb{R}^{+*}$$

For f in L^2 the existence of the coefficients $C_f(a,b)$ is ensured once ψ is in L^2. Nonetheless, in order to enable the inversion of the transformation, it is necessary to choose functions ψ belonging to $L^1 \cap L^2$ and satisfying the following admissibility condition relating to the Fourier transform $\hat{\psi}$ of ψ:

$$(CNS\text{-}ad) \quad \int_0^{+\infty} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} d\omega = \int_{-\infty}^{0} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} d\omega < +\infty \qquad [5.1]$$

These hypotheses about the ψ function guarantee the possibility of calculating the inverse transform. We then say that ψ is an admissible wavelet.

To verify that a real function is a wavelet we use the following simpler sufficient admissibility condition:

$$(CS\text{-}ad) \quad \psi \text{ real, } \psi \in L^1 \cap L^2, \int_{\mathbb{R}} \left|t\,\psi(t)\right| dt < +\infty \text{ and } \int_{\mathbb{R}} \psi(t)\, dt = 0$$

The condition $(CS\text{-}ad)$ makes it possible to affirm that the Fourier transform $\hat{\psi} \in C^1$ with $\hat{\psi}(0) = 0$; consequently, the integrals of the condition $(CNS\text{-}ad)$ exist.

5.2.1.2. *Simple examples of admissible wavelets*

The admissibility condition and the condition $(CS\text{-}ad)$ are very weak and it is easy to construct wavelets usable for continuous analysis. Almost any function integrating to zero is appropriate, provided that it also verifies some elementary properties.

Let us cite some simple examples of ψ functions verifying the condition $(CS\text{-}ad)$.

If f is Gaussian, $f(t) = Ae^{-Bt^2}$ with $A \in \mathbb{R}$ and $B > 0$, then all its derivatives satisfy the sufficient admissibility condition. Indeed, for all $n \in \mathbb{N}^*$ there is a polynomial P_n, such that $f^{(n)}(t) = P_n(t)e^{-Bt^2}$; therefore, $f^{(n)} \in L^1 \cap L^2$, $tf^{(n)} \in L^1$ and $\int_{\mathbb{R}} f^{(n)}(t)dt = \left[f^{(n-1)}\right]_{-\infty}^{+\infty} = 0$.

More generally, if f belongs to the set of rapidly decreasing functions, all the derivatives of f satisfy the condition.

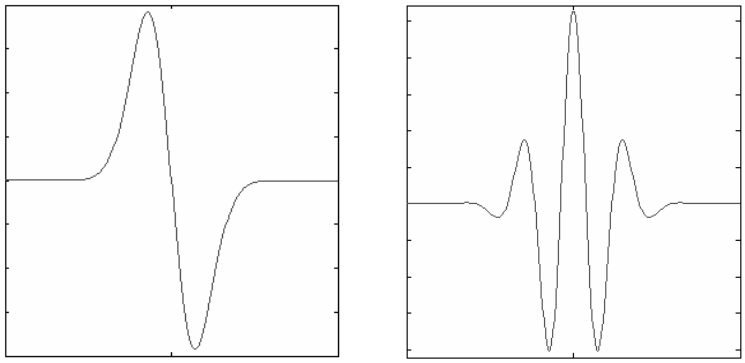


Figure 5.1. *Gaussian functions: 1st derivative on the left and 8th derivative on the right*

If $f \in L^1(\mathbb{R}^+) \cap L^2(\mathbb{R}^+)$ with $tf \in L^1(\mathbb{R}^+)$, by extending f to $\mathbb{R}^-$ in an odd fashion we also obtain a function satisfying the sufficient admissibility condition. We can, in particular, choose a continuous or piecewise continuous function f over $\mathbb{R}^+$ with a compact support, as in the example in Figure 5.2.

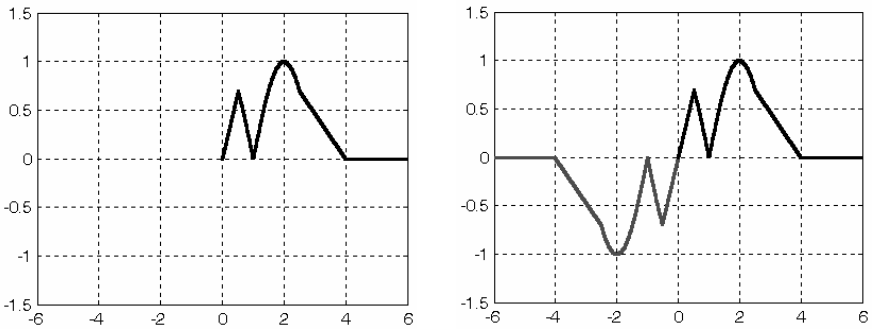


Figure 5.2. *On the left, a function f defined over* $\mathbb{R}^+$. *On the right, the wavelet* ψ *obtained by extension of f*

The construction of wavelets with compact support and arbitrary regularity is easy. Let us consider, for example, the following family with a parameter $n \in \mathbb{N}$:

$$f_n(t) = [\operatorname{sgn}(t)]^{n+1} \sin^n(\pi t) \mathbb{1}_{[-1,1]}(t)$$

The sufficient admissibility condition is satisfied. Moreover, $f_n \in C^{n-1}(\mathbb{R})$ and it is zero outside of the segment $[-1,1]$.

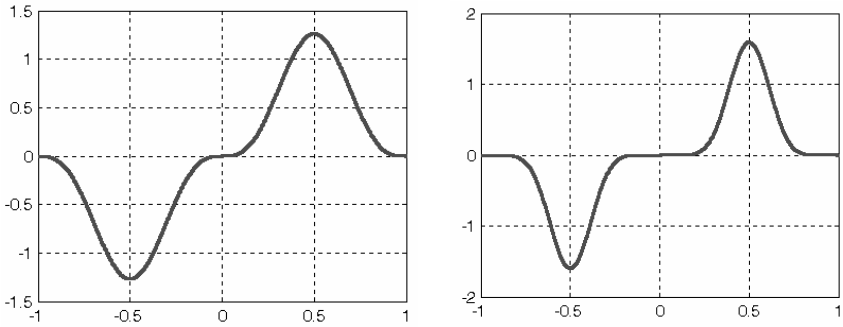


Figure 5.3. *Function* f_3 *on the left and function* f_8 *on the right*

In order to extend a support equal to $[a,b]$ we can, for example, consider the family $\{g_n\}$ defined by $g_n(t) = f_n\left(\frac{1}{2}(b(t+1)+a(t-1))\right)$ for $n \in \mathbb{N}$, or construct an adapted polynomial function.

5.2.1.3. *Construction of wavelets approaching a pattern*

On the basis of the given function f with compact support and finite energy we may consider the construction of a wavelet usable for continuous analysis, approximating this function as well as possible, in the least squares sense. Various construction methods are possible.

First method

Let us denote by $[a,b]$ an interval containing the support of f and take a family $F = \{\rho_i\}_{i=1}^{N}$ of N functions, continuous or piecewise continuous on $[a,b]$. Let us suppose that the family F is linearly independent in $L^2(a,b)$.

We want to construct the approximation in the least squares sense of the function f using a finite linear combination of the form $\psi = \sum_{i=1}^{N} \alpha_i \rho_i$. Given the hypotheses on f and F, the approximation ψ clearly satisfies the two conditions $\psi \in L^1 \cap L^2$ and $t\psi \in L^1$. The function ψ must integrate to zero (we can foresee having other vanishing moments). This condition reduces to the following linear constraint: $\sum_{i=1}^{N} \alpha_i R_i = 0$ where $R_i = \int_a^b \rho_i(t)\,dt$.

The vector $\alpha = \{\alpha_i\}_{i=1}^{N}$ and, consequently, the function $\psi = \sum_{i=1}^{N} \alpha_i \rho_i$ are obtained by solving a linear system. More precisely, by noting:

– G the Gram matrix $(N\text{x}N)$ defined by $G_{ij} = \int_a^b \rho_i(t)\rho_j(t)dt$;

– M the matrix of constraints $(1\text{x}N)$ with $M_i = R_i$;

– B the vector $(N\text{x}1)$ defined by $B_i = \int_a^b f(t)\rho_i(t)dt$;

the vector α and the Lagrange multiplier λ associated with the constraint are the solutions of the linear system:

$$\begin{pmatrix} G & M^t \\ M & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} = \begin{pmatrix} B \\ 0 \end{pmatrix}$$

When the family F is not orthogonal to the set of constants, M becomes the matrix of a surjective application of $\mathbb{R}^N$ in $\mathbb{R}$ and the solution is therefore unique.

EXAMPLE 5.1.

In this example we construct two approximating wavelets differing by the choice of the F functions family. Let us consider the function f with the support $[-1,1]$, defined by:

$$f(t) = \left\{C_1 \mathbb{1}_{[0,1]}(t) + C_2 \mathbb{1}_{[-1,0]}(t)\right\}\sin(\pi t)$$

Let us note that $\int_{\mathbb{R}} f(t)dt = \frac{2(C_1 - C_2)}{\pi}$ and thus $\int_{\mathbb{R}} f(t)dt \neq 0$, if $C_1 \neq C_2$.

Let us take, for example, $C_1 = 1$ and $C_2 = \frac{1}{2}$.

The wavelet represented in the left graph in Figure 5.4 is obtained using the family of polynomials of degree ≤ 2 over the interval $[-1,1]$ as family F. By employing as family F the same family of polynomials but on a larger interval, here $[-2,1]$, we obtain the wavelet of the right graph.

Of course, the result does not seem very satisfactory. To improve it, it is necessary to increase the degree of the F polynomials and to add continuity constraints to the ψ function. Let us examine that in the next section.

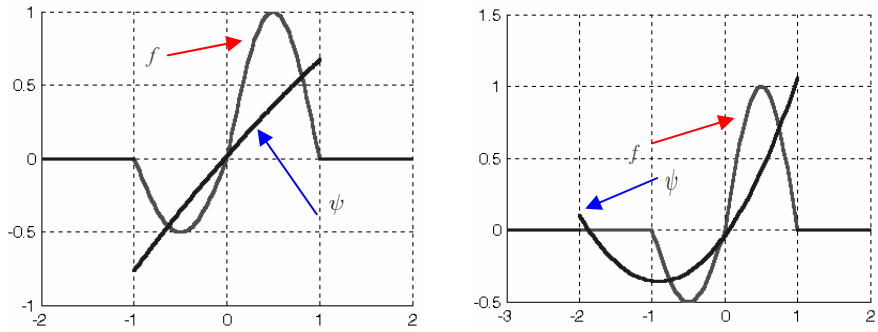


Figure 5.4. *Approximation by polynomials of degree* ≤ 2 *:*
on the left on $[-1,1]$*, on the right on* $[-2,1]$

Second method

Since the family F consists of continuous functions on $[a,b]$ we may seek a function ψ continuous on the real line $\mathbb{R}$. That amounts to imposing two additional linear constraints $\psi(a)=0$ and $\psi(b)=0$ leading to $\sum_{i=1}^{N}\alpha_i\rho_i(a)=0$ and $\sum_{i=1}^{N}\alpha_i\rho_i(b)=0$.

The constraints matrix M becomes a $(3\text{x}N)$ matrix defined by:

$$M_{1,i}=R_i,\ M_{2,i}=\rho_i(a) \text{ and } M_{3,i}=\rho_i(b),\ \forall i=1,\cdots,N$$

EXAMPLE 5.2.

Let us again take the function f from the previous example and use the polynomials of degree ≤ 3 on $[-1,1]$ to construct a wavelet ψ that would be the approximation of f in the least squares sense. We obtain the approximations represented in Figure 5.5.

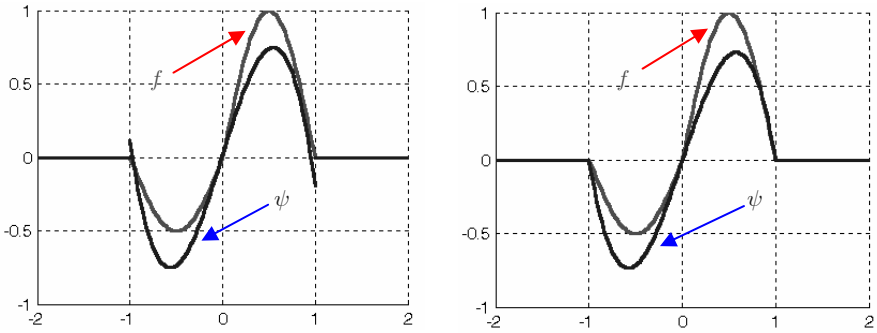


Figure 5.5. *Approximation by polynomials of degree* ≤ 3 *on* $[-1,1]$*: on the left unconstrained, on the right with the constraints* $\psi(-1) = 0$ *and* $\psi(1) = 0$

Third method

We can use a more direct construction method. We seek the best approximation of f in the least squares sense, within the space of functions orthogonal to constants. On the basis of $f \in L^2$ with $\text{supp}(f) \subset [a,b]$, the following ψ function approximates f and verifies the admissibility condition:

$$\psi = f - \left(\int_a^b f\right) \mathbb{1}_{[a,b]}$$

If f is continuous over the interval $[a,b]$, satisfying the two additional constraints $\psi(a) = 0$ and $\psi(b) = 0$ leads to seeking a wavelet ψ of the form:

$$\psi = f - \left(\alpha t^2 + \beta t + \gamma\right) \mathbb{1}_{[a,b]}$$

The coefficients α, β and γ are calculated solving the linear system:

$$\begin{pmatrix} \frac{b^3 - a^3}{3} & \frac{b^2 - a^2}{2} & b - a \\ a^2 & a & 1 \\ b^2 & b & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \int_a^b f \\ f(a) \\ f(b) \end{pmatrix}$$

Fourth method

In practice and for numerical calculations we generally know only one sampling of the function *f* over the $[a,b]$ interval. We have a finite set of values $\{(t_k, y_k)\}_{k=1,\cdots,K}$ such that:

$$a \leq t_k \leq b \text{ and } y_k \approx f(t_k).$$

As above, we consider a family $F = \{\rho_i\}_{i=1}^{N}$ linearly independent in $L^2(a,b)$, and we denote by V the span vector space of *F*. Formulated for this finite set of pairs, the problem amounts to seeking $\alpha = \{\alpha_i\}_{i=1}^{N}$ in $\mathbb{R}^N$ and thus $\psi = \sum_{i=1}^{N} \alpha_i \rho_i$ such that:

$$\sum_{k=1}^{K} [\psi(t_k) - y_k]^2 = \underset{\beta \in \mathbb{R}^N}{Min} \left\{ \sum_{k=1}^{K} [v_\beta(t_k) - y_k]^2 \text{ such that } \int_a^b v_\beta(t)dt = 0 \right\}$$

where for β in $\mathbb{R}^N$, $v_\beta = \sum_{i=1}^{N} \beta_i \rho_i$.

It is thus a problem of minimization in the least squares sense with a constraint, a problem similar to the one considered previously. The vector α and the Lagrange multiplier λ associated with the constraint are obtained by solving the linear system:

$$\begin{pmatrix} G & M^t \\ M & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} = \begin{pmatrix} B \\ 0 \end{pmatrix} \text{ with } G, M \text{ and } B \text{ defined by:}$$

$$G_{ij} = \sum_{k=1}^{K} \rho_i(t_k)\rho_j(t_k), \quad M_i = \int_a^b \rho_i(t)\,dt, \quad B_i = \sum_{k=1}^{K} y_k \rho_i(t_k).$$

5.2.2. *Application to pattern detection*

In this section we propose considering the following detection problem. We take S signals consisting of a linear combination of dilations-translations $f\left(\frac{t-b}{a}\right)$ of a basic form f and, possibly, also a noise and a "low frequency" function.

In this context, only $S(t)$ and $f(t)$ are known. We have to identify the number of patterns $\left\{f\left(\frac{t-b_k}{a_k}\right)\right\}_{k=1}^{K}$ and the value of the pairs (scale, position) $\{(a_k,b_k)\}_k$.

The following process is used:

– find an acceptable wavelet ψ approximating the form f using one of the methods presented in the previous section;

– perform an analysis of S using the adapted wavelet ψ;

– select (a_k,b_k) seeking the maxima of the coefficients' surface.

In this section we will find three increasingly complex examples, for which we demonstrate that:

– the above process performs well;

– the use of another wavelet systematically leads to worse results.

EXAMPLE 5.3.

Let us consider the basic form already encountered in example 5.1, i.e. the function f defined by:

$$f(t) = \left\{C_1 \mathbb{1}_{[0,1]}(t) + C_2 \mathbb{1}_{[-1,0]}(t)\right\} \sin(\pi t)$$

with $C_1 = 1$ and $C_2 = 0.9$. Since $\int_{\mathbb{R}} f(t)\,dt = \frac{2(C_1 - C_2)}{\pi} \simeq 0.0637 \neq 0$, this form is not a wavelet.

We construct a signal S containing two specimens dilated from this basic form f and depicted in Figure 5.6.

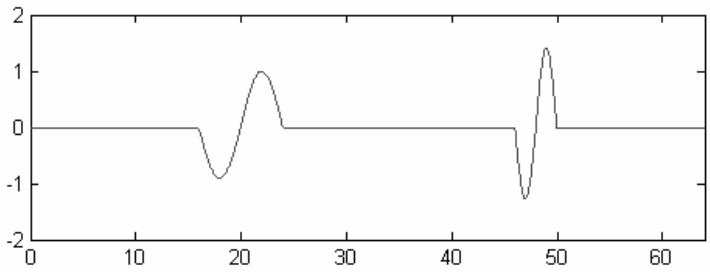


Figure 5.6. *Constructed signal:* $S(t) = f\left(\frac{t-20}{8}\right) + \sqrt{2}\, f\left(\frac{t-48}{4}\right)$

With the detection technique we hope to find two pieces of information for each dilated form:

– scale a (8 and 4 respectively);

– position b (20 and 48 respectively).

We analyze the signal S using one of the admissible wavelets approximating the basic f form constructed in example 5.1, i.e. we calculate its coefficients $C_s(a,b)$. We compare the results obtained with this adapted wavelet to those associated with the *db1* wavelet. We obtain time-scale graphs seen in Figures 5.7 and 5.8.

In each of the two figures we find three graphs. On the left there are the contour plot of the wavelet coefficients $C_s(a,b)$ with positions b along the x-axis and scales a along the y-axis. On the right are the two zooms on the two pairs $(a,b) = (8,20)$ and $(4,48)$.

The two analyses are similar: the position of the two dilated forms is detected well. Nevertheless, the adapted wavelet is more effective in locating the two scales 8 and 4: the y-coordinates of local maxima identify these scales better when the approximating wavelet is used. Although relevant, since the wavelet slightly resembles the sought form, the analysis with *db1* indicates a slight shift towards the bottom (in scale) of these maxima.

Let us now test a more regular wavelet: *sym4* (see Figure 5.9). This analysis proves much less fruitful. The position and the scale of the two dilated forms are not detected well. This phenomenon is due to the size of the wavelet support. More generally, it seems that *db1* is the best amongst the wavelets not adapted to f.

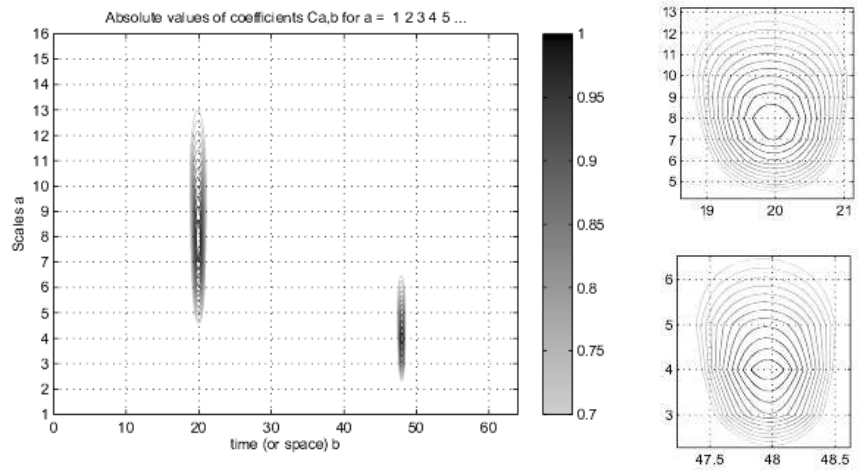


Figure 5.7. *On the left, contour plot of the coefficients of the signal analyzed with the wavelet approximating the form. On the right, zooms on the two detected pairs (scale, position)*

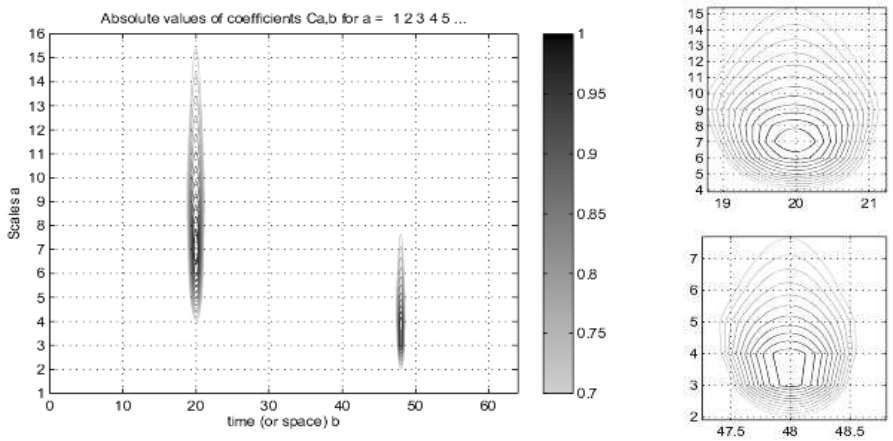


Figure 5.8. *On the left, contour plot of the coefficients of the signal analyzed with the db1 wavelet. On the right, zooms on the two detected pairs (scale, position)*

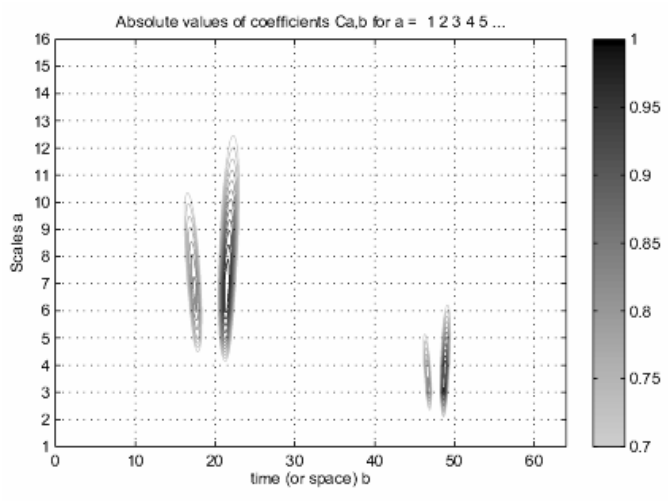


Figure 5.9. *Contour plot of the coefficients of the signal analyzed with the sym4 wavelet*

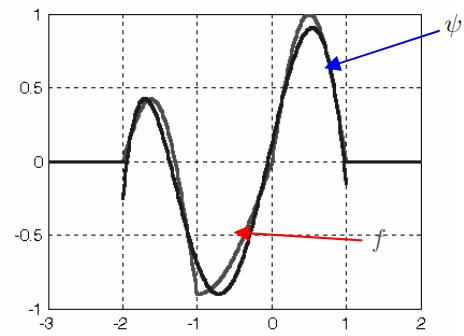


Figure 5.10. *Pattern* f *and adapted wavelet* ψ *obtained using polynomials of degree* ≤ 6 *on* $[-1,1]$

EXAMPLE 5.4.

Let us consider a second example, where we will partially superimpose two translated and dilated versions of another basic form f, defined by:

$$f(t) = \left\{C_1 \mathbb{1}_{[0,1]}(t) + C_2 \mathbb{1}_{[-1,0]}(t)\right\} \sin(\pi t) + \frac{1}{4} \mathbb{1}_{[-1,1]}(t) \sin(4\pi t)$$

with $C_1 = 1$ and $C_2 = 0.9$, and, thus, also here $\int_{\mathbb{R}} f(t)\,dt = \frac{2\left(C_1 - C_2\right)}{\pi} \simeq 0.0637$.

We construct an admissible wavelet approximating f using polynomials of degree ≤ 6 on [−1,1] (see Figure 5.10).

We construct a signal S containing two dilated and partially superimposed specimens of this basic form.

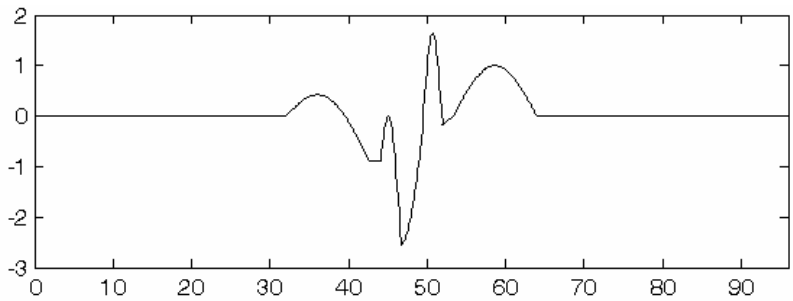


Figure 5.11. *Constructed signal:* $S(t) = f\left(\frac{t-T}{\mathbf{32}}\right) + 2\,f\left(\frac{t-T}{\mathbf{8}}\right)$, *with T = 48*

We successively analyze signal S with the adapted wavelet and the *db1* wavelet. (Figures 5.12 and 5.13, respectively). In the two analyses the joint position of the two dilated forms is clearly detected. Nonetheless, the adapted wavelet is more accurate in locating the two scales 32 and 8. Compared to the preceding one, the analysis with *db1* presents a considerable downwards shift of these maxima. Let us stress that the use of a wavelet with larger support is not efficient for the detection of the position and scale of every form.

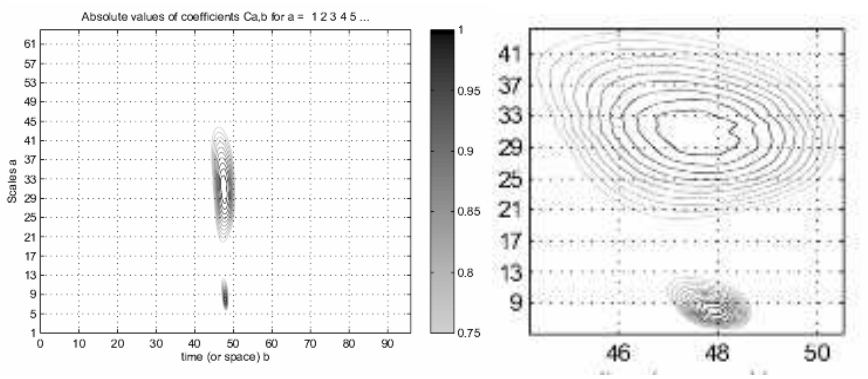


Figure 5.12. *Contour plot of the coefficients of the signal analyzed with the adapted wavelet (on the right is the zoom on the significant part)*

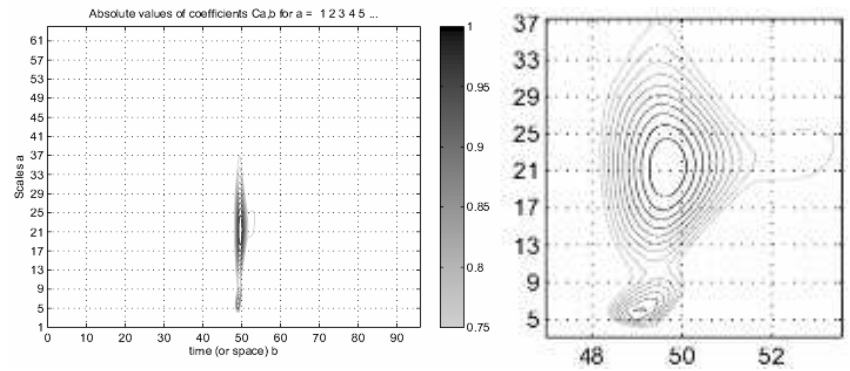


Figure 5.13. *Contour plot of the coefficients of the signal analyzed with the db1 wavelet (on the right is the zoom on the significant part)*

EXAMPLE 5.5.

In order to illustrate the stability of the behavior highlighted above, let us consider a basic signal obtained using the function f from example 5.2 and add a "triangular" signal T and a Gaussian white noise B with large variance.

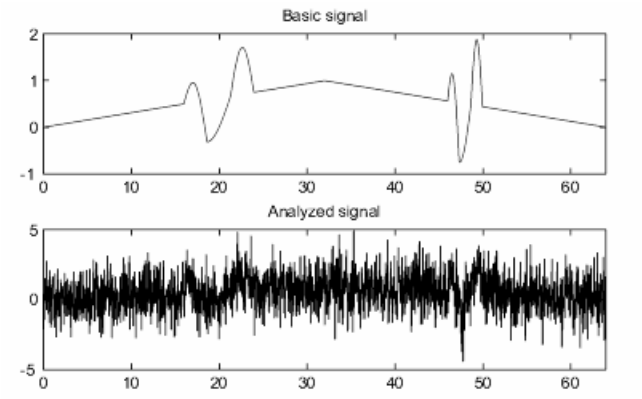


Figure 5.14. *Constructed signal:* $S(t) = f\left(\frac{t-20}{8}\right) + \sqrt{2}\, f\left(\frac{t-48}{4}\right) + T(t) + B(t)$

Let us analyze the signal S with the adapted wavelet (it is the approximating wavelet from example 5.2) and with the *db1* wavelet.

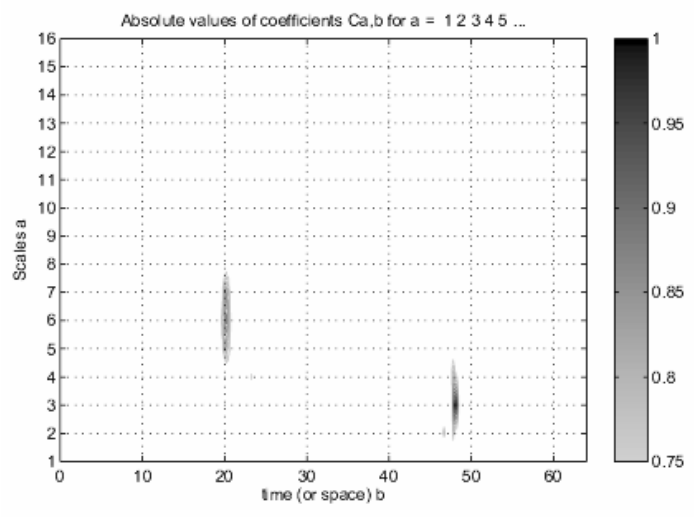


Figure 5.15. *Signal analysis with the adapted wavelet*

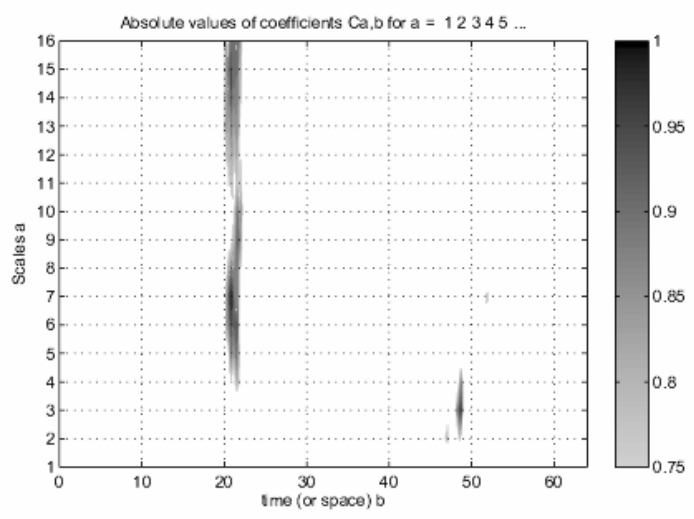


Figure 5.16. *Signal analysis with the db1 wavelet*

Despite the presence of noise, the wavelet adapted to the form remains more efficient in locating the two characteristic scales 8 and 4. A wavelet that is more regular than *db1* (for example, *dbN* or *symN*) smoothes out the effect of the noise. However, the scales of the two forms are still badly detected.

5.3. Construction of wavelets for discrete analysis

The rest of this chapter is dedicated to the construction of wavelets for discrete analysis. First of all, we widen the framework passing by the two-channel filter banks, whose properties are examined. In particular, we determine the conditions enabling perfect reconstruction. This classical signal processing approach [EST 77] leads to what is called quadrature mirror filters (QMF) [MIN 85] or also conjugated mirror filters [SMI 86], which are also often referred to as QMF.

A perfect reconstruction filter bank is known as "biorthogonal" and the associated filters as biorthogonal. Lightly constraining conditions make it possible to obtain such filter banks without too much difficulty. With the "lifting" method it is then possible to construct an infinite number thereof starting from a biorthogonal bank. Moreover, thanks to the technique known as "polyphase", we show [DAU 97] that all the biorthogonal transformations can be decomposed into elementary lifting steps.

Finally, the link with wavelets is the subject of the last section. By starting from a biorthogonal filter bank with some additional conditions we can construct biorthogonal wavelet bases. In particular, on the basis of a biorthogonal filter bank stemming from wavelets, we can construct an infinite number of filter banks of the same type, among which it is possible to distinguish those associated with biorthogonal wavelets.

5.3.1. *Filter banks*

5.3.1.1. *From the Mallat algorithm to filter banks*

Within the framework of orthogonal wavelets the decomposition-reconstruction algorithm uses operations of filtering by convolution, down-sampling and up-sampling. This algorithm uses four filters obtained from a single filter associated with the scaling function φ by the two-scale equation (see Chapter 3). Two filters are used for the decomposition and two for the reconstruction.

One of the algorithm's properties is invertibility: a decomposition followed by a reconstruction provides again the original signal. Moreover, various properties link the four filters and the (φ, ψ) pair.

Let us forget for the moment the link with wavelet bases and consider the more general context of two-channel filter banks, in order to determine the conditions ensuring invertibility. Let us consider the two-channel filter bank represented in Figure 5.17: the top part represents one channel and the bottom part the other.

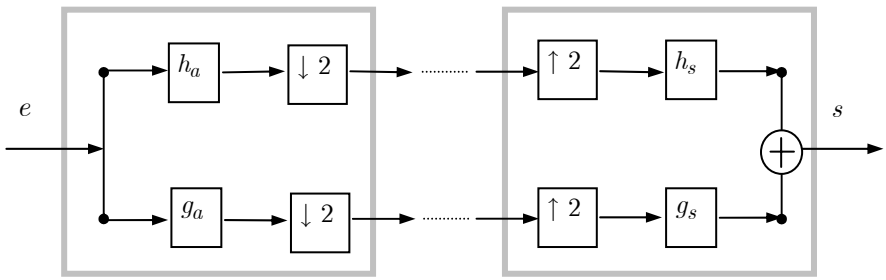


Figure 5.17. *Two-channel filter bank*

The inputs and outputs are sequences indexed by $\mathbb{Z}$.

In each channel the input sequence e is filtered using h_a (respectively g_a) and then decimated by $\downarrow 2$. This is the analysis part. In order to synthesize we start with two sequences, insert the zeros (operation noted $\uparrow 2$), we filter using h_s and g_s respectively and, finally, sum up to obtain the output sequence s.

The main property necessary for a filter bank is the property of perfect reconstruction. After a decomposition followed by a reconstruction without any intermediate operation we want to obtain an output s equal (possibly up to a close translation) to the value of input e.

What are the relations that must link the four filters h_a, h_s, g_a and g_s in order to satisfy this property?

5.3.1.2. *The perfect reconstruction condition*

Let us consider the filter bank in Figure 5.17, for the case where there is no intermediate step between decomposition and reconstruction. We may then represent it more concisely using the following figure.

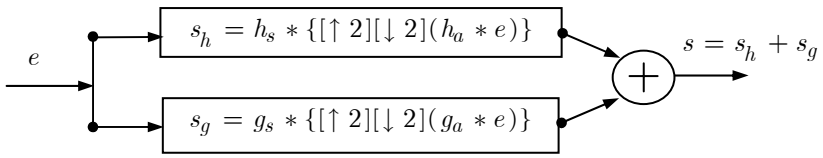


Figure 5.18. *Two-channel filter bank (input-output view)*

A necessary and sufficient condition (NSC) linking the four filters h_a, h_s, g_a and g_s to obtain a perfect reconstruction with a translation of d time units is provided by the theorem [VET 86][1] where the z-transform of a sequence $s = (s_k)_{k\in\mathbb{Z}}$ is formally defined by $S(z) = \sum_{k\in\mathbb{Z}} s_k z^{-k}$, $z \in \mathbb{C}$.

1 See also [MAL 98], p. 260.

THEOREM 5.1. (PERFECT RECONSTRUCTION)– *let* h_a, g_a, h_s, g_s *be the four bank filters of Figure 5.18. Let us note their respective z-transforms as* H_a, G_a, H_s *and* G_s.

For a given $d \in \mathbb{Z}$ *a NCS, such that for every input* e *the corresponding output* s *verifies* $s = T_d(e)$, *where* $\left(T_d(e)\right)_k = e_{k-d}$ *(i.e. the output* s *is equal to the input* e *delayed by* d *time units), is:*

$$H_s(z)\,H_a(z) + G_s(z)\,G_a(z) = 2\,z^{-d} \qquad [5.2]$$

and:

$$H_s(z)\,H_a(-z) + G_s(z)\,G_a(-z) = 0 \qquad [5.3]$$

We speak then of a perfect reconstruction and, when $d = 0$, we have perfect reconstruction without translation: $s = e$.

■

Condition [5.3] is called the anti-aliasing condition.

In the following section we present a simple method of constructing four filters ensuring the perfect reconstruction condition which is then illustrated using three examples.

5.3.1.3. *Construction of perfect reconstruction filter banks*

We can exploit the preceding result to determine the four z-transforms: H_a, H_s, G_a and G_s. We seek to easily satisfy condition [5.3], so as to express relation [5.2] more simply. However, for the simplest choices of the additional constraint, the reconstruction is necessarily carried out with a translation.

First of all, we will seek an immediate solution of equation [5.3]. Let Θ be an unspecified function (in practice it is a monomial). If H_a, H_s, G_a and G_s are selected so that:

$$H_s(z) = \Theta(z)\,G_a(-z) \text{ and } G_s(z) = -\Theta(z)\,H_a(-z) \qquad [5.4]$$

then relation [5.3] is verified automatically. Using this in [5.2] we obtain just one equation to be satisfied, where there appear only two unknown factors, H_a and H_s:

$$H_s(z)\,H_a(z) - \frac{\Theta(z)}{\Theta(-z)}\,H_s(-z)\,H_a(-z) = 2\,z^{-d}\ ,\ d \in \mathbb{Z} \qquad [5.5]$$

Relation [5.5] is simplified if we choose an even or odd Θ function, which will be done later on. The pair (H_a, H_s) then verifies:

$$H_s(z)\,H_a(z) - \sigma H_s(-z)\,H_a(-z) = 2\,z^{-d}\ ,\ d \in \mathbb{Z} \qquad [5.6]$$

with $\sigma = +1$ if Θ is even and $\sigma = -1$ if Θ is odd.

In [STR 96] (p. 105-109) and articles by Sweldens, causal filters are privileged. Relation [5.4] with $\Theta = 1$ is retained, i.e.: $H_s(z) = G_a(-z)$ and $G_s(z) = -H_a(-z)$. Relation [5.6] is then satisfied with $\sigma = +1$.

The pair (G_a, G_s) satisfies similar relations.

If $\sigma = +1$, the right-hand side of the equality [5.6] is an odd function of z. This imposes an odd value for d. Reconstruction has to be performed with a translation, a "delay". If $\sigma = -1$, the first member of the equality is an even function of z and, thus, d necessarily has an even value. In this case we can take $d = 0$, i.e. to obtain a reconstruction without translation.

Regardless of the parity of Θ, posing $X_H = H_s H_a$ and by writing $X = z^d X_H$, the equality [5.6] leads to a very simple equation:

$$X(z) + X(-z) = 2 \qquad [5.7]$$

The first member of the above equality is an even function. If we seek X as a Laurent series, then X involves only odd powers of z, except for a constant term equal to 1. In other words, we have:

$$X(z) = 1 + \sum_{p \in \mathbb{Z}} a_p\, z^{-(2p+1)}$$

Under these conditions the search for the four filters H_a, H_s, G_a and G_s that verify the Vetterli reconstruction condition can be carried out using algorithm 5.1.

Initial choices

– Choose the sequence $\{a_p\}_{p\in\mathbb{Z}}$ and write $X(z) = 1 + \sum_{p\in\mathbb{Z}} a_p\, z^{-(2p+1)}$

– Choose $d \in \mathbb{Z}$ and write $X_H(z) = X(z)\, z^{-d}$

Step 1: Determination of H_a and H_s

– Factorize X_H in the form of $X_H = H_s H_a$

Step 2[2]*:* Determination of G_a and G_s

– If *d* is odd, choose an even Θ and take $\sigma = +1$; if *d* is even, choose an odd Θ and take $\sigma = -1$

– Calculate: $G_a(z) = \sigma \dfrac{H_s(-z)}{\Theta(z)}$ and $G_s(z) = -\Theta(z)\, H_a(-z)$

Algorithm 5.1. *Construction algorithm*

This process makes it possible to construct the four filters choosing X, and thus X_H, almost arbitrarily. Then, by factorizing X_H we determine the two filters of the same channel (h_a, h_s). Afterwards the two filters of the other channel are calculated (g_a, g_s). Naturally, since the pairs (h_a, h_s) and (g_a, g_s) play symmetrical roles, we can calculate the second pair first.

For the choice of Θ we can, for example, take functions like: $\Theta(z) = \varepsilon z^p$, where $\varepsilon = \pm 1$ and $p = 2m$ or $p = 2m + 1$ with $m \in \mathbb{Z}$.

NOTE 5.1.– it is obviously simpler to start directly from a factorization $X_H = H_s H_a$. However, it requires that all the product terms have the same parity, except for one whose coefficient must be equal to 1.

On the other hand, it is also possible to formulate the problem by first seeking the two analysis filters, i.e. the pair (h_a, g_a) and then the two synthesis filters (h_s, g_s).

2 This step explains why we need non-causal filters if we want zero delay. The choice *d* = 0 implies that one of the four filters has a z-transform, which contains positive and negative powers and is therefore non-causal. Conversely, if we require causal filters, the delay cannot be zero.

Lastly, if we seek finite impulse response filters, the functions H_a, H_s, G_a and G_s are necessarily Laurent polynomials and Θ is a monomial.

■

5.3.1.4. *Examples of perfect reconstruction filter banks*

In this section we present three examples of perfect reconstruction filter banks.

First of all, we take a look at the "lazy wavelet", which is not a wavelet in the traditional sense of the term. It leads to the simplest possible perfect reconstruction filter bank: "it does not do anything". It is, in particular, the ideal candidate to initialize a lifting process (see section 5.3.2.1).

EXAMPLE 5.6. (THE "LAZY WAVELET")

Let us introduce the following notation: for $p \in \mathbb{Z}$, $\delta^{(p)}$ stands for the sequence defined by $\delta_p^{(p)} = 1$ and $\delta_k^{(p)} = 0$ for $k \neq p$. The sequence $\delta^{(0)}$ is the element neutral for the convolution and, more generally, for $p \in \mathbb{Z}$, $b = \delta^{(p)} * a = T_p(a)$, i.e. $b_k = a_{k-p}$.

Let us take $X(z) = 1$, $d = 0$ and $\Theta(z) = -z$, which leads to $X_H(z) = X(z) = 1$. We can factorize X_H in the form $H_a(z) = 1$ and $H_s(z) = 1$; we obtain $G_a(z) = z^{-1}$ and $G_s(z) = z$. Thus:

$$h_a = \delta^{(0)},\ h_s = \delta^{(0)},\ g_a = \delta^{(1)} \text{ and } g_s = \delta^{(-1)}$$

The convolutions by h_a and h_s simply correspond to the identity. The filters g_a and g_s carry out translations of 1 and (-1) respectively. How, then, does the filter bank operate?

Since $s_h = h_s * \{[\uparrow 2][\downarrow 2](h_a * e)\}$ and $s_g = g_s * \{[\uparrow 2][\downarrow 2](g_a * e)\}$, the analysis part separates the components with an even index from those with an odd index, then performs a translation by 1 to the right for the components with an odd index. According to Figure 5.17 it is illustrated by:

$$e = \{e_k\}_{k\in\mathbb{Z}} \rightarrow \left| \begin{array}{ll} \xrightarrow{Id} \{e_k\}_{k\in\mathbb{Z}} & \xrightarrow{[\downarrow 2]} \{e_{2p}\}_{p\in\mathbb{Z}} \\ \xrightarrow[T_1]{} \{e_{k-1}\}_{k\in\mathbb{Z}} & \xrightarrow[[\downarrow 2]]{} \{e_{2p-1}\}_{p\in\mathbb{Z}} \end{array} \right.$$

The synthesis part inserts zeros, translates the odd index components by a step of 1 to the left and then calculates the sum:

$$\begin{array}{l} \left\{e_{2p}\right\}_{p\in\mathbb{Z}} \xrightarrow{[\uparrow 2]} \left\{\cdots,0,e_0,0,\cdots\right\} \xrightarrow{Id} \left\{\cdots,0,e_0,0,\cdots\right\} \\ \left\{e_{2p-1}\right\}_{p\in\mathbb{Z}} \xrightarrow[[\uparrow 2]]{} \left\{\cdots,0,e_{-1},0,\cdots\right\} \xrightarrow[T_{-1}]{} \left\{\cdots,e_{-1},0,e_1,\cdots\right\} \end{array} \Bigg| \xrightarrow{\oplus} e$$

Thus, we find the original sequence.

EXAMPLE 5.7. ("HAAR FILTERS")

The second example deals with the simplest orthogonal wavelet: the Haar wavelet. A direct application of the preceding algorithm makes it possible to build the filters associated with this wavelet, but also other filters ensuring perfect reconstruction but not leading to a wavelet.

Let us take $X(z) = \frac{1}{2}\left(z^{1/2} + z^{-1/2}\right)^2 = \frac{1}{2}\left(z + 2 + z^{-1}\right)$, $d = 0$ and $\Theta(z) = -z^{-1}$.

A first factorization of $X = X_H$ is given by $X_H = H_a H_s$ with:

$$H_a(z) = \frac{1}{\sqrt{2}}(1 + z) \text{ and } H_s(z) = \frac{1}{\sqrt{2}}\left(1 + z^{-1}\right).$$

From it we deduce that: $G_a(z) = \frac{1}{\sqrt{2}}(-1 + z)$ and $G_s(z) = \frac{1}{\sqrt{2}}\left(-1 + z^{-1}\right)$.

We recognize the z-transforms of the Haar filters:

$$h_a = \frac{1}{\sqrt{2}}\left(\delta^{(0)} + \delta^{(-1)}\right),\ h_s = \frac{1}{\sqrt{2}}\left(\delta^{(0)} + \delta^{(1)}\right)$$

$$g_a = \frac{1}{\sqrt{2}}\left(-\delta^{(0)} + \delta^{(-1)}\right),\ g_s = \frac{1}{\sqrt{2}}\left(-\delta^{(0)} + \delta^{(1)}\right)$$

These filters are associated with the Haar scaling function and the Haar wavelet (see Chapter 4).

Another factorization, which does not lead to a wavelet in the classical sense of the term, is possible:

$$H_a(z) = \tfrac{1}{\sqrt{2}} \text{ and } H_s(z) = \tfrac{1}{\sqrt{2}}\left(z + 2 + z^{-1}\right)$$

which yields $G_a(z) = \tfrac{1}{\sqrt{2}}\left(z^2 - 2z + 1\right)$ and $G_s(z) = \tfrac{1}{\sqrt{2}}\, z^{-1}$, and, consequently:

$$h_a = \tfrac{1}{\sqrt{2}}\,\delta^{(0)},\ h_s = \tfrac{1}{\sqrt{2}}\left(\delta^{(-1)} + 2\delta^{(0)} + \delta^{(1)}\right)$$

$$g_a = \tfrac{1}{\sqrt{2}}\left(\delta^{(-2)} - 2\delta^{(-1)} + \delta^{(0)}\right),\ g_s = \tfrac{1}{\sqrt{2}}\,\delta^{(1)}$$

EXAMPLE 5.8. (LONGER FILTERS, THE *BIOR2.2*)

This last example applies the construction technique using a longer X filter and leads to several quadruplets with perfect reconstruction. One of them is associated with the pair of biorthogonal wavelets *bior2.2* (see Chapter 4).

Let us consider a symmetrical Laurent polynomial longer than those in the preceding examples:

$$X(z) = \tfrac{1}{16}\left(-z^{-3} + 9z^{-1} + 16 + 9z - z^3\right) \text{ with } d = 0 \text{ and } \Theta(z) = -z^{-1}.$$

The polynomial X is factorized in the form:

$$X(z) = \frac{1}{16}(z+1)^2\left(z^{-1}+1\right)^2(z-\alpha)\left(\alpha^{-1}z^{-1}-1\right) \text{ with } \alpha = 2+\sqrt{3}\,.$$

Consequently, by distributing the zeros of the polynomial z^3X in H_a and in H_s we can factorize $X = X_H$ as $X_H = H_aH_s$ in several ways. One of them, provided by factorization:

$$X(z) = \frac{\left(1 + 2z^{-1} + z^{-2}\right)}{4}\frac{\left(-z^{-1} + 2 + 6z + 2z^2 - z^3\right)}{4}\,,$$

leads to the *bior2.2* wavelet.

5.3.2. *Lifting*

5.3.2.1. *The lifting method*

In this section we demonstrate that by using a quadruplet solution of the system [5.2] and [5.3], we can build an infinite number of other solutions through a method called "lifting" introduced by Sweldens ([SWE 98], see also [STR 96] p. 214).

The technique is as follows. We start with an odd or even Θ (thus, $\sigma = \pm 1$), with $d \in \mathbb{Z}$ and two pairs (H_a, H_s) and (G_a, G_s) that are solutions of [5.4]. Due to relation [5.6], these pairs are solutions of equation:

$$F_s(z)\,F_a(z) - \sigma F_s(-z)\,F_a(-z) = 2\,z^{-d} \qquad [5.8]$$

where F_a and F_s are the unknown factors.

The key to lifting is that on the basis of a pair (F_a, F_s), which is a solution of the previous equation, we can build an infinite number of other solution pairs. Indeed, let us fix F_s and take a function η. We can define $F_a^N(z) = F_a(z) + F_s(-z)\eta(z)$ (the N exponent stands for "new" and this convention applies to the rest of the chapter). When $\sigma = 1$ and η is even, the pair $\left(F_a^N, F_s\right)$ is also the solution of [5.8]. We thus produce an infinite number of solutions parameterized by η. We have the same result for $\sigma = -1$ and an odd η.

If we only consider polynomial solutions, it suffices to take an unspecified Laurent polynomial P and to take $\eta(z) = P(z^2)$ for $\sigma = 1$ or $\eta(z) = zP(z^2)$ for $\sigma = -1$.

By symmetry, if we fix F_a, all the solutions $\left(F_a, F_s^N\right)$ of [5.8] are obtained in a similar way (this operation is sometimes called dual lifting).

Theorem 5.2 synthesizes the above idea and provides a parameterization of the infinite number of polynomial quadruplets that solve the system [5.4] and [5.5] (and thus, also the system [5.2] and [5.3]) obtained by lifting starting from a quadruplet polynomial solution (H_a, G_a, H_s, G_s).

THEOREM 5.2. (PARAMETRIZATION SETTING OF THE SOLUTIONS)[3]

Primal lifting: *let* (H_a, G_a, H_s, G_s) *be a polynomial quadruplet verifying [5.4] and [5.5]. The set of solutions* $\left(H_a^N, G_a, H_s, G_s^N\right)$ *of [5.4] and [5.5] is provided by:*

$$\left|\begin{array}{l} H_a^N(z) = H_a(z) + G_a(z)S(z^2) \\ G_s^N(z) = G_s(z) - H_s(z)S(z^2) \end{array}\right. \quad \textit{where } S \textit{ is an unspecified Laurent polynomial.}$$

Dual lifting*: let* (H_a, G_a, H_s, G_s) *be a polynomial quadruplet verifying [5.4] and [5.5]. The set of quadruplets* $\left(H_a, G_a^N, H_s^N, G_s\right)$ *solving [5.4] and [5.5] is yielded by:*

$$\left|\begin{array}{l} H_s^N(z) = H_s(z) + G_s(z)T(z^2) \\ G_a^N(z) = G_a(z) - H_a(z)T(z^2) \end{array}\right. \quad \textit{where } T \textit{ is an unspecified Laurent polynomial.}$$

■

Dual lifting is achieved by exchanging the roles of the "a" and "s" indices in the primal lifting.

5.3.2.2. *Lifting and the polyphase method*

Up to now we could construct an infinite number of quadruplets with perfect reconstruction using just one of them thanks to the above parametrization. In fact, the situation is simpler than it appears. On the basis of elementary quadruplets, for example of the "lazy wavelet" or the Haar filters (see examples 5.6 and 5.7) we can obtain an infinite number of solutions by using elementary lifting steps. To specify this result and to define the concept, a detour to introduce the polyphase method, which is interesting in itself, is necessary. We thus first describe it in this section and then examine its consequences for lifting.

In the decomposition part of the filter bank schematized in Figure 5.19 we perform convolution, followed by decimation where only even index terms are preserved for each "branch", also referred to as phase. The odd index terms are not used. We may wonder about the possibility of permuting the decimation and filtering so as to make calculation more efficient. It is the object of the polyphase method.

3 The theorem can also be written using the Fourier transform or directly in the time domain: [STR 96] p. 216.

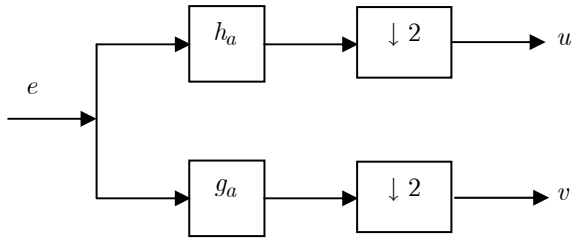


Figure 5.19. *Decomposition part of the filter bank*

Let us consider the calculation of u on the basis of the input sequence $e = \{e_k\}_{k\in\mathbb{Z}}$. The usual form is as follows: we have $u = [\downarrow 2](h_a * e)$, but $(h_a * e)_k = \sum_{n\in\mathbb{Z}} e_{k-n}\,[h_a]_n$ and, thus:

$$u_k = (h_a * e)_{2k} = \sum_{n\in\mathbb{Z}} e_{2k-n}\,[h_a]_n \qquad [5.9]$$

The polyphase method consists simply of reformulating the preceding result by introducing two elementary sequences: the sequence of even index components and the sequence of odd index components both for the input and the filter. For a sequence $r = \{r_k\}_{k\in\mathbb{Z}}$ let us define the "even" and "odd" sub-sequences by $r_P = \{r_{2n}\}_{n\in\mathbb{Z}}$ and $r_I = \{r_{2n+1}\}_{n\in\mathbb{Z}}$[4]. On the basis of [5.9] the new expression becomes:

$$\begin{aligned} u_k &= \sum_{n\in\mathbb{Z}} e_{2k-n}\,[h_a]_n = \sum_{j\in\mathbb{Z}} e_{2k-2j}\,[h_a]_{2j} + \sum_{j\in\mathbb{Z}} e_{2k-(2j+1)}\,[h_a]_{2j+1} \\ &= \sum_{j\in\mathbb{Z}} e_{2(k-j)}\,[h_a]_{2j} + \sum_{j\in\mathbb{Z}} e_{2((k-1)-j)+1}\,[h_a]_{2j+1} \\ &= \sum_{j\in\mathbb{Z}} \left[e_P\right]_{k-j}\left[h_{a,P}\right]_j + \sum_{j\in\mathbb{Z}} \left[e_I\right]_{(k-1)-j}\left[h_{a,I}\right]_j \\ &= \left(e_P * h_{a,P}\right)_k + \left(T_1\left[e_I\right] * h_{a,I}\right)_k \end{aligned}$$

4 If R, R_P and R_I indicate the z-transforms of the sets r, r_P and r_I respectively, then: $R(z) = R_P\left(z^2\right) + z^{-1}R_I\left(z^2\right)$. Consequently:

$$R_P\left(z^2\right) = \frac{R(z) + R(-z)}{2} \quad \text{and} \quad R_I\left(z^2\right) = \frac{R(z) - R(-z)}{2z^{-1}}.$$

Consequently: $u = \left(e_P * h_{a,P}\right) + T_1\left[e_I\right] * h_{a,I}$, where T_1 is the operator of translation of 1 step in time. Similarly we have: $v = \left(e_P * g_{a,P}\right) + T_1\left[e_I\right] * g_{a,I}$. By taking the z-transform of the two previous expressions, we obtain:

$$U(z) = H_{a,P}(z)E_P(z) + z^{-1}H_{a,I}(z)E_I(z)$$
$$V(z) = G_{a,P}(z)E_P(z) + z^{-1}G_{a,I}(z)E_I(z)$$

which is, in matrix form:

$$\begin{pmatrix} U(z) \\ V(z) \end{pmatrix} = \begin{pmatrix} H_{a,P}(z) & H_{a,I}(z) \\ G_{a,P}(z) & G_{a,I}(z) \end{pmatrix} \begin{pmatrix} E_P(z) \\ z^{-1}E_I(z) \end{pmatrix}$$

revealing what is called the polyphase analysis matrix[5]:

$$M_a^{Poly}(z) = \begin{pmatrix} H_{a,P}(z) & H_{a,I}(z) \\ G_{a,P}(z) & G_{a,I}(z) \end{pmatrix}$$

The analysis part of the diagram can then be illustrated as in Figure 5.20.

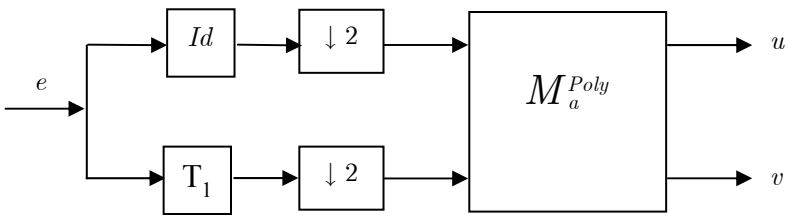


Figure 5.20. *The polyphase form of the decomposition part of the filter bank.* ***Id*** *is the identity transformation and* $\boldsymbol{T_1}$ *is the operator of translation of 1 step in time*

5 Some authors call the polyphase matrix the transposition of the matrix used here.

Thus, we first separate the input *e* components with an even index e_P from those with an odd index e_I; then, the outputs u and v are obtained using convolutions by filters, $\left(h_{a,P}, h_{a,I}\right)$ for example, twice shorter than in Figure 5.19, and a summation. Thus, there are roughly half as many operations than in the initial notation, since the terms $e_P * h_{a,P}$ and $T_1\left[e_I\right] * h_{a,I}$ involved in the calculation of $u = \left(e_P * h_{a,P}\right) + T_1\left[e_I\right] * h_{a,I}$ are each four times shorter than $e * h_a$.

Proceeding in a similar way and defining the polyphase synthesis matrix by:

$$M_s^{Poly}(z) = \begin{pmatrix} H_{s,P}(z) & H_{s,I}(z) \\ G_{s,P}(z) & G_{s,I}(z) \end{pmatrix}$$

the synthesis part can be put in the form represented by Figure 5.21.

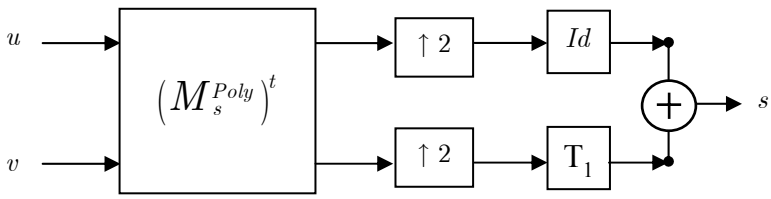


Figure 5.21. *Polyphase form of the reconstruction part of the filter bank*

The perfect reconstruction condition (for $d = 0$) is then rewritten as:

$$\left(M_s^{Poly}(z)\right)^t M_a^{Poly}(z) = \begin{pmatrix} 1 & 0 \\ 0 & z \end{pmatrix}$$

NOTE 5.2.– let us note $\widetilde{H_a}(z) = H_a\left(z^{-1}\right)$ and $\widetilde{M_a}^{Poly}$ the polyphase matrix associated with $\widetilde{H_a}$. This amounts to considering $\widetilde{h_a}$ the inverse filter of the h_a filter. We then find a traditional formulation of the previous perfect reconstruction condition:

$$\left(M_s^{Poly}(z)\right)^t \widetilde{M_a}^{Poly}(z^{-1}) = Id$$

Apart from the advantages concerning the implementation of the decomposition and reconstruction algorithms brought by the polyphase method, we also achieve a particularly important theoretical result.

In the polyphase formulation we may reinterpret the passing from a (H_a, G_a, H_s, G_s) solution of [5.4] and [5.5] to a new solution $\left(H_a^N, G_a, H_s, G_s^N\right)$ or $\left(H_a, G_a^N, H_s^N, G_s\right)$ using theorem [5.2] and multiplying the left side (or the right) of the polyphase analysis matrix (or of the transpose of polyphase synthesis matrix) by the matrix:

$$\begin{pmatrix} 1 & S(z) \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 \\ T(z) & 1 \end{pmatrix} \left(\begin{pmatrix} 1 & -S(z) \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 \\ -T(z) & 1 \end{pmatrix} \text{ respectively).}$$

Such an operation is called an elementary lifting step.

The factorization theorem 5.3 introduced by Daubechies and Sweldens[6] shows that all quadruplet perfect reconstruction filters have a polyphase matrix that can be factorized into a sequence of elementary lifting steps.

THEOREM 5.3. (FACTORIZATION)– *let there be* (H_a, G_a, H_s, G_s) *verifying [5.4] and [5.5], such that the determinant of the polyphase matrix is equal to 1. Then, there exist two families of* m *Laurent polynomials* $\{S_i\}_{i=1}^m$ *and* $\{T_i\}_{i=1}^m$ *and a constant* $K \neq 0$ *such that the polyphase analysis matrix admits the following factorization:*

$$M_a^{Poly}(z) = \left\{ \prod_{i=1}^{m} \begin{pmatrix} 1 & S_i(z) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ T_i(z) & 1 \end{pmatrix} \right\} \begin{pmatrix} K & 0 \\ 0 & 1/K \end{pmatrix}$$

■

This theorem indicates how to construct an infinite number of solutions of [5.4] and [5.5] using only elementary lifting steps starting from a perfect reconstruction quadruplet.

6 This theorem is cited in the article of Daubechies and Sweldens, "Factoring Wavelet Transform into Lifting Steps" [DAU 97]. See also theorem 7.13 on p. 275 of [MAL 98]

Let us stress that the elementary lifting matrices are explicitly invertible and the inverses are also elementary lifting matrices, since:

$$\begin{pmatrix} 1 & F(z) \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -F(z) \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 \\ F(z) & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -F(z) & 1 \end{pmatrix}$$

This result has an important algorithmic consequence when a "polyphase" version of the wavelet decomposition is implemented: calculation of the synthesis step, the inverse of the preceding step, becomes elementary.

5.3.3. *Lifting and biorthogonal wavelets*

We can show[7] that if $(h_a\ g_a\ h_s, g_s)$ is a perfect reconstruction filter bank and if the modulus of the Fourier transforms are bounded, then the families $\{h_a(\cdot - 2k), g_a(\cdot - 2k)\}_{k\in\mathbb{Z}}$ and $\{h_s(\cdot - 2k), g_s(\cdot - 2k)\}_{k\in\mathbb{Z}}$ are biorthogonal bases of $l^2(\mathbb{Z})$. In the particular case where $(h_a, g_a) = (h_s, g_s)$ we obtain an orthogonal basis of $l^2(\mathbb{Z})$.

Let us consider a perfect reconstruction filter bank h_a, h_s, g_a and g_s. If h_a and h_s have finite length, then:

$$\widehat{\varphi_a}(\omega) = \prod_{n=1}^{\infty} \left(\frac{1}{\sqrt{2}} \overline{\widehat{h_a}(\frac{\omega}{2^n})} \right) \text{ and } \widehat{\varphi_s}(\omega) = \prod_{n=1}^{\infty} \left(\frac{1}{\sqrt{2}} \overline{\widehat{h_s}(\frac{\omega}{2^n})} \right)$$

converge in $S'(\mathbb{R})$, the set of tempered distributions.

Consequently, φ_a and φ_s are also defined in $S'(\mathbb{R})$. However, *a priori* there is no reason for φ_a and φ_s to be elements of $L^2(\mathbb{R})$. For that to be true, it is necessary to impose additional conditions. Unfortunately, they are technical and in general difficult to verify. However, we state the theorem introduced by Cohen ([COH 90] and [MAL 98] p. 265).

7 See theorem 7.9 in [MAL 98] p. 262.

THEOREM 5.4.– *let* (h_a, g_a, h_s, g_s) *be a perfect reconstruction filter bank. We suppose that there exist two unique strictly positive trigonometric polynomials* $P_a\left(e^{2i\pi\omega}\right)$ *and* $P_s\left(e^{2i\pi\omega}\right)$ *(up to a normalization) such that:*

$$\left|\widehat{h_a}\left(\frac{\omega}{2}\right)\right|^2 P_a\left(e^{i\pi\omega}\right) + \left|\widehat{h_a}\left(\frac{\omega+1}{2}\right)\right|^2 P_a\left(e^{i\pi(\omega+1)}\right) = 2P_a\left(e^{2i\pi\omega}\right)$$

$$\left|\widehat{h_s}\left(\frac{\omega}{2}\right)\right|^2 P_s\left(e^{i\pi\omega}\right) + \left|\widehat{h_s}\left(\frac{\omega+1}{2}\right)\right|^2 P_s\left(e^{i\pi(\omega+1)}\right) = 2P_s\left(e^{2i\pi\omega}\right)$$

We suppose, moreover, that: $\underset{\omega\in\left[-\frac{1}{2},\frac{1}{2}\right]}{Inf}\left|\widehat{h_a}(\omega)\right| > 0$ *and* $\underset{\omega\in\left[-\frac{1}{2},\frac{1}{2}\right]}{Inf}\left|\widehat{h_s}(\omega)\right| > 0$

Then:

– the functions $\widehat{\varphi_a}$ *and* $\widehat{\varphi_s}$*, and, consequently, the functions* φ_a *and* φ_s*, belong to* $L^2(\mathbb{R})$*. Moreover, the latter two functions satisfy the biorthogonality relations:*

$$\left(\varphi_a(x), \varphi_s\left(x-n\right)\right)_{L^2(\mathbb{R})} = \delta_{0,n} \quad , \quad \forall n \in \mathbb{Z}$$

– the two families $\left\{(\psi_a)_{j,k}\right\}_{j,k\in\mathbb{Z}}$ *and* $\left\{(\psi_s)_{j,k}\right\}_{j,k\in\mathbb{Z}}$ *are Riesz biorthogonal bases of* $L^2(\mathbb{R})$*.*

■

NOTE 5.3.– if $P_a = P_s = 1$ are appropriate, the first two conditions of the theorem and the perfect reconstruction condition are then written, by noting $m_0(\omega) = \frac{1}{\sqrt{2}}\overline{\widehat{h_a}(\omega)}$ and $\widetilde{m_0}(\omega) = \frac{1}{\sqrt{2}}\overline{\widehat{h_s}(\omega)}$:

$$\left|m_0(\omega)\right|^2 + \left|m_0(\omega+\tfrac{1}{2})\right|^2 = 1 \text{ and } \left|\widetilde{m_0}(\omega)\right|^2 + \left|\widetilde{m_0}(\omega+\tfrac{1}{2})\right|^2 = 1$$

$$\overline{m_0(\omega)}\,\widetilde{m_0}(\omega) + \overline{m_0(\omega+\tfrac{1}{2})}\,\widetilde{m_0}(\omega+\tfrac{1}{2}) = 1$$

These conditions are well known (see Chapter 2) and simpler to verify. ■

We can then use the results of lifting in the context of biorthogonal wavelets. Let us consider the quadruplet $(\varphi_a, \psi_a, \varphi_s, \psi_s)$ of scaling functions and biorthogonal wavelets. The process below shows that by using the results of lifting we can construct others quadruplets of scaling functions and biorthogonal wavelets. This is "formal", in the sense that the functions obtained do not necessarily belong to L^2.

Lifting process[8]. Let $\left\{\sigma_k\right\}_{k\in\mathbb{Z}}$ be a family of real numbers containing a finite number of non-zero elements and $(\varphi_a, \psi_a, \varphi_s, \psi_s)$ be a quadruplet of scaling functions and biorthogonal wavelets. For a fixed φ_s the family $\left(\varphi_a^N, \psi_a^N, \varphi_s, \psi_s^N\right)$, defined below, is also "formally" a quadruplet of scaling functions and biorthogonal wavelets.

$$\begin{cases} \psi_s^N(x) = \psi_s(x) - \sum\limits_{k\in\mathbb{Z}} \sigma_k \varphi_s(x-k) \\ \varphi_a^N(x) = 2\sum\limits_{k\in\mathbb{Z}} (h_a)_k\, \varphi_a^N(2x-k) + \sum\limits_{k\in\mathbb{Z}} \sigma_{-k} \psi_a^N(x-k) \\ \psi_a^N(x) = 2\sum\limits_{k\in\mathbb{Z}} (g_a)_k\, \varphi_a^N(2x-k) \end{cases} \qquad [5.10]$$

If we put in parallel the above result and the lifting theorem [5.2], the modifications to the scaling functions and the wavelets for the associated filters correspond to a primal lifting of the Laurent polynomial $S(z) = \sum\limits_k \sigma_k z^{-k}$.

Let us also note that the new wavelet ψ_a^N verifies the same two-scale relation as the wavelet ψ_a. The functions φ_a^N and ψ_s^N satisfy new two-scale relations taking into account the contribution of the Laurent polynomial.

8 See Sweldens, "The lifting scheme: a custom-design construction of biorthogonal wavelets", *Appl. Comput. Harmon. Anal*., vol. 3, no. 2, p. 186-200, 1996. See also theorem 7.12 in [MAL 98], p. 273.

Similarly, permuting the roles of a and s in [5.10] on the basis of the same initial quadruplet $(\varphi_a, \psi_a, \varphi_s, \psi_s)$, we also construct quadruplets in the form of $\left(\varphi_a, \psi_a^N, \varphi_s^N, \psi_s^N\right)$.

5.3.4. *Construction examples*

The end of this chapter presents applications of lifting. First we treat examples that enable us to illustrate the effects of one or more lifting steps and of a parameterized lifting. Then we outline a lifting solution for two problems: construction of a wavelet admitting more vanishing moments that the initial wavelet and then the approximation of a pattern.

The scaling functions and wavelets are constructed using the cascade algorithm described in Chapter 4.

5.3.4.1. *Illustrations of lifting*

The four following examples illustrate lifting.

EXAMPLE 5.9. (THE *BIOR1.3* WAVELET)

In this example we show how to obtain the *bior1.3* wavelet (see Chapter 4) using lifting.

We start with Haar filters (h_a, g_a) and (h_s, g_s) whose **z**-transforms are given by:

$$H_a(z) = \frac{1}{\sqrt{2}}(1+z),\ H_s(z) = \frac{1}{\sqrt{2}}\left(1+z^{-1}\right),$$

$$G_a(z) = \frac{1}{\sqrt{2}}(-1+z) \text{ and } G_s(z) = \frac{1}{\sqrt{2}}\left(-1+z^{-1}\right).$$

The two pairs (scaling function, wavelet) (φ_a, ψ_a) and (φ_s, ψ_s) are represented in Figure 5.22.

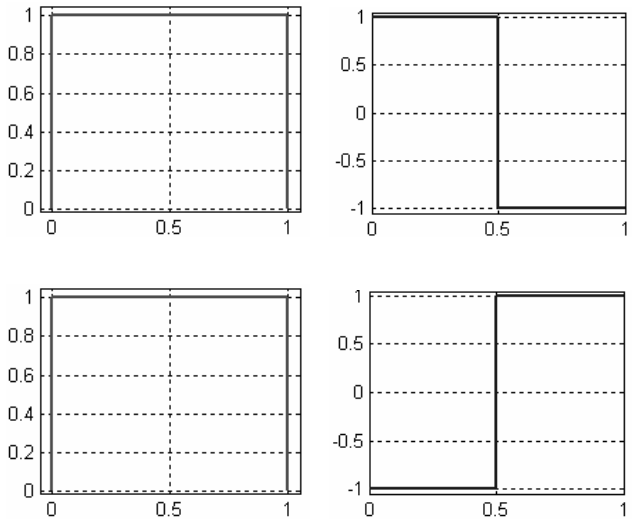


Figure 5.22. *The two pairs (scaling function, wavelet) associated with Haar filters*

Two steps of lifting are performed in order to obtain this.

The first step is a dual lifting with the polynomial $T(z) = \frac{1}{8}\left(1 - z^{-1}\right)$. The filters H_a and G_s are unchanged, the two others become:

$$H_s(z) = \frac{1}{\sqrt{2}}\left(\frac{7}{8} + \frac{9}{8}z + \frac{5}{4}z^{-2} - \frac{5}{4}z^{-3}\right) \text{ and } G_a(z) = z\,H_s(-z).$$

The scaling function φ_a associated with the pair (H_a, G_a) is not modified. The three other functions are transformed as shown in Figure 5.23.

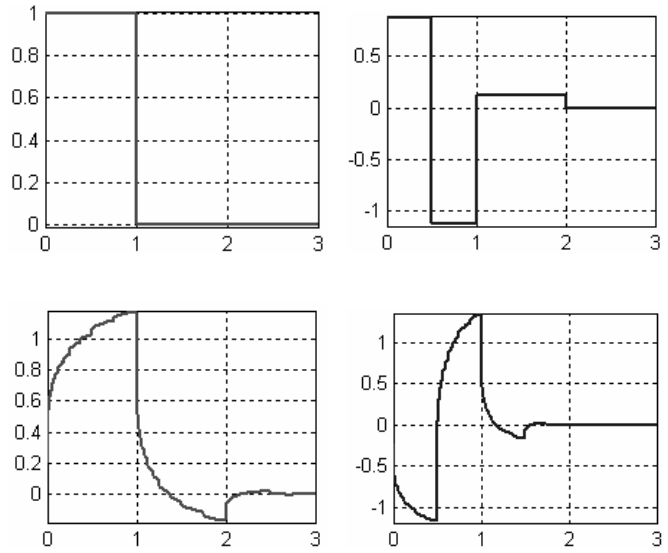


Figure 5.23. *The two pairs (scaling function, wavelet) after the first lifting*

Thus, in passing, we obtain a new biorthogonal wavelet that is non-indexed, but usable.

The second step is also a dual lifting. This time the polynomial used is $T(z) = \frac{1}{8}(z-1)$. Again, the filters H_a and G_s are unchanged. The two others become:

$$H_s(z) = \frac{1}{\sqrt{2}}\left(-\frac{5}{4}z^2 + \frac{5}{4}z + 1 + z^{-1} + \frac{5}{4}z^{-2} - \frac{5}{4}z^{-3}\right)$$

and $G_a(z) = z\,H_s(-z)$.

As in the first step, φ_a is not modified and the three other functions are transformed. The pairs (φ_s, ψ_s) for the analysis and (φ_a, ψ_a) for the synthesis are the two pairs (scaling function, wavelet) corresponding to the *bior1.3* case.

In Figure 5.24 we represent the pairs (scaling function, wavelet) obtained after the second step of lifting.

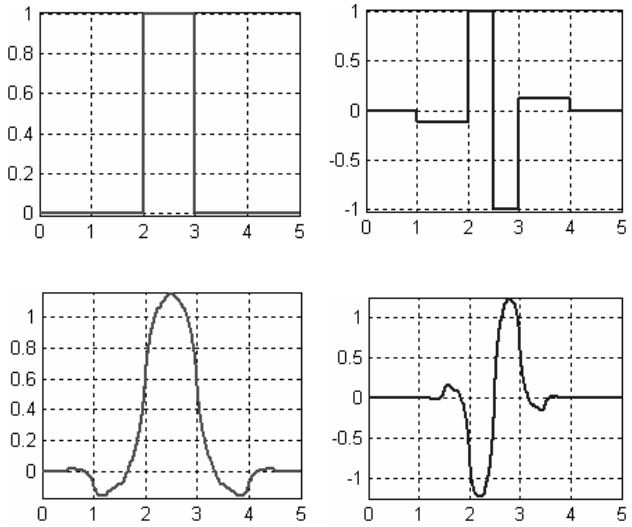


Figure 5.24. *The two pairs (scaling function, wavelet) after the second lifting*

Let us note that there is no unique manner to construct this pair. Indeed, two successive steps of lifting of the same nature (primal or dual) of the respective polynomial P_1 and P_2 can be replaced by only one step (of identical nature) of the polynomial $P_1 + P_2$. For the considered example, the single step of dual lifting of the polynomial $T(z) = \frac{1}{8}\left(z - z^{-1}\right)$ yields the same final result.

Conversely, we can "divide" a step of this type into several successive steps. The intermediate results may prove interesting. Let us again take the example considered. We can obtain the same final result by starting again from the Haar filters and carrying out four steps of lifting. The first two filters consist of applying the first step of lifting above using $T(z) = \frac{1}{8}\left(1 - z^{-1}\right)$.

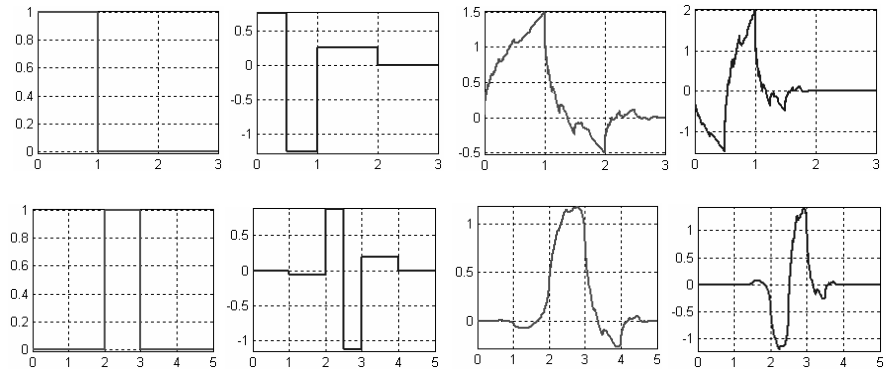


Figure 5.25. *The two pairs (scaling function, wavelet) after the 2nd lifting (top), and the two pairs (scaling function, wavelet) after the 3rd lifting (bottom)*

We then carry out two identical steps consisting of a dual lifting of polynomial $T(z) = \frac{1}{16}\left(z - 2 + z^{-1}\right)$. In Figure 5.25 we represent the pairs (scaling function, wavelet) corresponding to the two intermediate steps of lifting.

There again we obtain two new pairs of biorthogonal wavelets. Let us note that the two synthesis pairs consist of irregular functions, which finally yield a pair of regular and symmetrical functions.

EXAMPLE 5.10. (LIFTINGS PARAMETERIZED ON THE BASIS OF THE HAAR WAVELET)

The purpose of this example is to illustrate the extreme variety of results that can be obtained by applying the same lifting step up to a constant, to a fixed initial quadruplet.

Let us again start with Haar filters and carry out a single step of dual lifting with the polynomial: $T(z) = \frac{1}{K}\left(1 - z^{-1}\right)$. Let us successively choose values 1, 2, 4, 6, 32 and 128 for the constant K. The results are collected in Figures 5.26, 5.27 and 5.28. The first two cases yield highly irregular functions. When the parameter K is small it becomes necessary to calculate the Fourier transforms of the filters precisely in order to know if we still remain within the finite energy functions framework. For $K = 4$ we obtain a pair whose synthesis couple resembles *db2*. For $K = 6$ we have a smoothed version thereof. The last two cases illustrate that as K grows we approach the Haar wavelet, which is discontinuous.

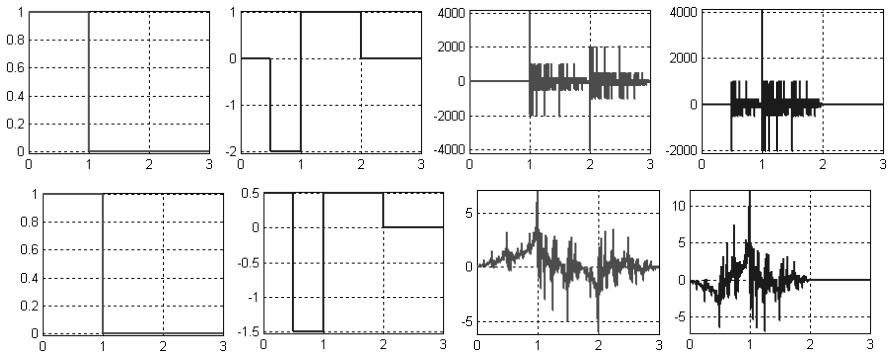


Figure 5.26. *The two pairs (scaling function, wavelet) after one lifting step. Top for K = 1, bottom for K = 2*

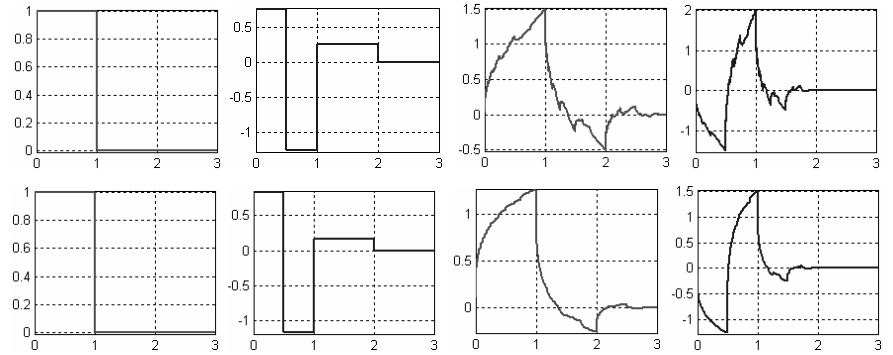


Figure 5.27. *The two pairs (scaling function, wavelet) after one lifting step. Top for K = 4, bottom for K = 6*

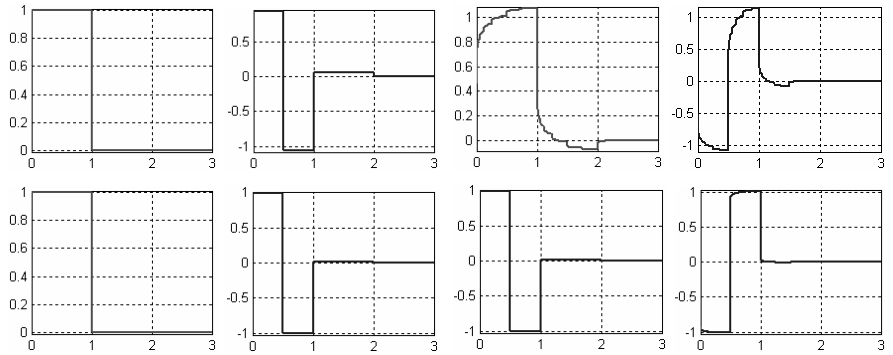


Figure 5.28. *The two pairs (scaling function, wavelet) after one lifting step. Top for K = 32, bottom for K = 128*

EXAMPLE 5.11. (A DUAL LIFTING + A PRIMAL LIFTING)

In all the previous examples we have only used dual liftings. The scaling function φ_a then remains unchanged. In this example we carry out a dual lifting and then a primal lifting. Under these conditions the four functions (φ_a, ψ_a), (φ_s, ψ_s) are modified. The first step is a dual lifting of polynomial: $T(z) = \frac{1}{K}\left(1 - z^{-1}\right)$. The second step is a primal lifting of polynomial: $S(z) = \frac{1}{K}\left(-1 + z^{-1}\right)$. Figure 5.29 illustrates this example for $K = 5$.

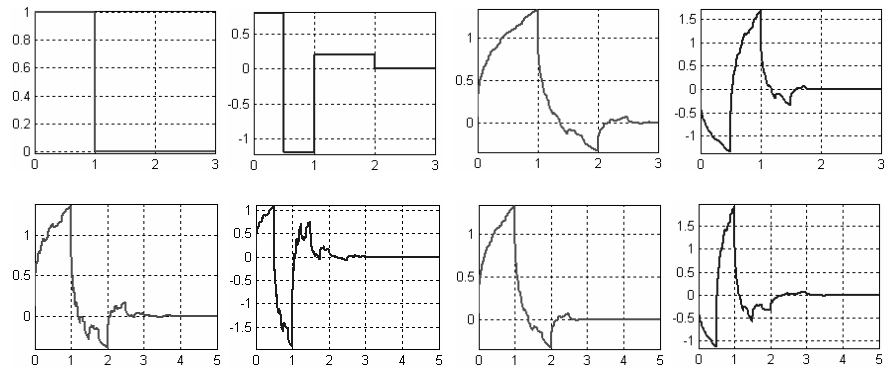


Figure 5.29. *The two pairs (scaling function, wavelet) after two lifting steps*

EXAMPLE 5.12. ("LAZY WAVELET" AND *BIOR1.3*)

Example 5.9 shows how to obtain the *bior1.3* wavelet starting from the Haar wavelet. This example aims at illustrating an alternative construction, on the basis of an object which is not a wavelet, the "lazy wavelet", and by applying a succession of elementary liftings.

The most direct construction uses two elementary lifting steps and a normalization constant $K = \frac{1}{\sqrt{2}}$ (see the factorization theorem 5.3). The first step is a primal lifting of polynomial $S(z) = 1$ which does not lead to a pair of biorthogonal wavelets. The second lifting step is a dual lifting with $T(z) = \frac{1}{16}\left(-z + 8 + z^{-1}\right)$. We can also "divide" this last step into several dual lifting steps. Splitting $T(z) = \left\{\frac{1}{2}\right\} + \left\{\frac{1}{16}\left(-z + z^{-1}\right)\right\}$ yields the Haar wavelet as

an intermediate result. We can also use the split $T(z) = \left\{\frac{1}{32}\left(-z + 16 + z^{-1}\right)\right\} + \left\{\frac{1}{32}\left(-z + z^{-1}\right)\right\}$ leading to a new pair of biorthogonal wavelets. In Figure 5.30, we illustrate this last construction.

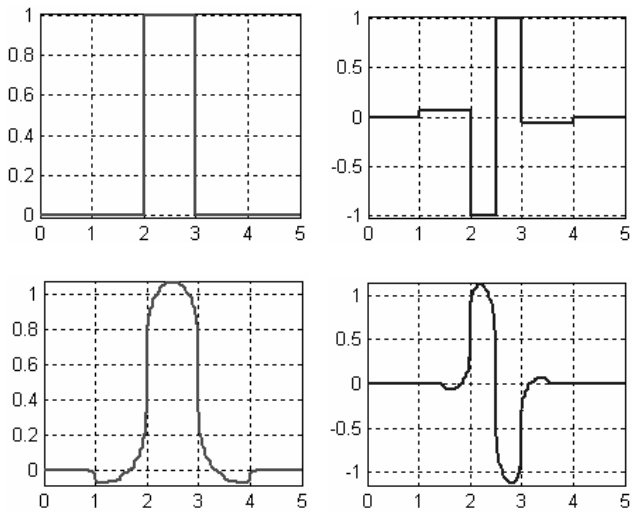


Figure 5.30. *The two pairs (scaling function, wavelet) after the 2nd lifting step*

5.3.4.2. *Construction of wavelets with more vanishing moments*

On the basis of a quadruplet $(\varphi_a, \psi_a, \varphi_s, \psi_s)$ we wish to construct a wavelet ψ_s^N that has more vanishing moments than ψ_s. Let us again take the first of the relations [5.10]: $\psi_s^N = \psi_s - \sum_{k\in\mathbb{Z}} \sigma_k \varphi_s(x-k)$. We multiply it by x^m, for $m = 0,\cdots,M$ and integrate over $\mathbb{R}$. The finite family of coefficients $\{\sigma_k\}_{k\in\mathbb{Z}}$ satisfies a linear system of $(M+1)$ equations:

$$\int_{\mathbb{R}} x^m \psi_s^N(x)\,dx = \int_{\mathbb{R}} x^m \psi_s(x)\,dx - \sum_{k\in\mathbb{Z}} \sigma_k \int_{\mathbb{R}} x^m \varphi_s(x-k)\,dx\,, m = 0,\cdots,M$$

It is therefore enough to find a family of $(M+1)$ coefficients $\{\sigma_k\}_K^{K+M}$ leading to a linear system of $(M+1)$ equations with $(M+1)$ unknowns.

Let us consider an elementary example. We start with the quadruplet $(\varphi_a, \psi_a, \varphi_s, \psi_s)$, consisting of the scaling function and the Haar wavelet for analysis and synthesis and we wish to construct a new wavelet ψ_s^N with two vanishing moments. The new wavelet ψ_s^N has the form:

$$\psi_s^N(x) = \psi_s(x) - \sigma_0\varphi_s(x) - \sigma_1\varphi_s(x-1)$$

The numbers σ_0 and σ_1 are solutions of a linear system of two equations with two unknowns. Since ψ_s has one vanishing moment and φ_s integrates 1, the first equation is reduced to: $0 = \sigma_0 + \sigma_1$. The calculation of the moments of order 1 yields the second equation: $\frac{1}{2} = \sigma_0 + 3\sigma_1$. Consequently, $\sigma_1 = \frac{1}{4}$ and $\sigma_0 = -\frac{1}{4}$.

The obtained ψ_s^N function is:

$$\psi_s^N(x) = \begin{cases} 3/4 & \text{over } [0,1/2] \\ -5/4 & \text{over } [1/2,1] \\ 1/4 & \text{over } [1,2] \\ 0 & \text{otherwise} \end{cases}$$

The functions of the new quadruplet are those obtained in the first row of Figure 5.27, from left to right: φ_s, ψ_s^N, φ_a^N and ψ_a^N.

5.3.4.3. *Approximation of a form by lifting*

In this section we outline the search for a wavelet approximating a form f by lifting. We use the first relation of [5.10] to approximate f by $\psi^N(x) = \psi(x) - \sum_{k\in\mathbb{Z}} \sigma_k\varphi(x-k)$, where the choice of the family $(\sigma_k)_{k\in\mathbb{Z}}$ is to be made for given ψ and φ.

Let us, firstly, make two observations:

– ψ^N is constructed by "correcting" $\psi(x)$, which is a fixed function of W_0, using a chosen function belonging to V_0;

– for ψ^N to be a wavelet, its integral has to be zero and, therefore, that $\sum_{k\in\mathbb{Z}} \sigma_k = 0$.

The approximation of the form f amounts to minimizing $\left\| f - \psi^N \right\|_{L^2}$, i.e. to minimizing the sum of the two terms:

$$\begin{aligned}\left\| f - \psi^N \right\|_{L^2} &= \left\| f - P_{V_0}(f) + P_{V_0}(f) - \left(\psi - \sum_k \sigma_k \varphi(x-k) \right) \right\|^2 \\ &= \left\| f - P_{V_0}(f) - \psi \right\|^2 + \left\| P_{V_0}(f) + \sum_k \sigma_k \varphi(x-k) \right\|^2\end{aligned}$$

because $f - P_{V_0}(f)$ is orthogonal to V_0, $\psi \in W_0 \perp V_0$ therefore, $f - P_{W_0}(f) - \psi \perp V_0$ and $P_{V_0}(f) + \sum \sigma_k \varphi(\cdot - k) \in V_0$.

The first term is fixed when ψ is fixed and the lifting only makes it possible to minimize the second term.

It is important to note that the corrections are on the same scale as f. An approach consists of dilating f so that the scale of the corrections becomes small compared to that of the approximated form. The technique would then consist of approximating a dilated version of f, $F_{-j}(x) = 2^{-j/2} f\left(2^{-j}x\right)$ for $j > 0$ by its projection onto V_0 and then contracting the obtained function. The result is obviously better because this amounts to approximating f by its projection onto V_{-j}, which is a space providing finer approximations. Indeed, for the example of the Haar wavelet, V_0 consists of functions constant in the intervals $[k, k+1[$, $k \in \mathbb{Z}$, whereas V_{-j} is the space of functions constant in 2^j times smaller intervals $[k2^{-j}, (k+1)\,2^{-j}[$. This approach has been deeply investigated by Mesa [MES 05].

Chapter 6

A Short 1D Illustrated Handbook

6.1. Introduction

Interesting signals generally exhibit numerous non-stationary characteristics that constitute a considerable part of the information contained in a series: drifts, trends, breakdowns, beginnings and ends of events, transitory phenomena. Most classical mathematical approaches widely used are well adapted to the study of stationary processes. Among others we may cite the spectral approach linked to the Fourier transform and that resulting from the ARMA processes [AZE 84].

This chapter takes you through the wavelet decomposition of signals and compiles "illustrated handbooks" associated with classical situations frequently encountered in statistics or signal processing. Its application is pedagogical as well as practical, and is based on the analysis of several elementary examples to learn how to recognize forms. The signals are presented followed by tests and trials, and we comment on the results. This chapter consists of three parts.

The first part[1], which is also the longest, concerns discrete signal analysis. It is intentionally over-detailed and exploits the capacity of wavelets to decompose a signal into a sum of approximation and detail signals. The topics dealt with relate to the identification of trends, periodic signals, noises, breakdowns and discontinuities. The commented analysis of 10 simulated signals constitutes the first "illustrated handbook". It is supplemented by an examination of two real signals: an electrical load curve and an electroencephalogram.

---

1 This part has borrowed much from the article [MIS 93] published in *Revue de Statistique Appliquée*, which the authors thank for the kind authorization.

The second part is short and dedicated to the discrete wavelet packet analysis. Wavelet packets are a generalization of wavelets which make it possible, in particular, to improve the frequency resolution of the wavelet analysis. This aspect is illustrated by three examples.

The third part deals with continuous analysis, which is focused on detail coefficients and enables a better detection and a better description of singularities. In particular, we deal with the cone of influence of a singularity, the frequency resolution of the analysis and the precise analysis of the Hölderian regularity of the singularities.

6.2. Discrete 1D illustrated handbook

In this section we are interested exclusively in reconstructed approximation and detail signals (see section 6.2.2). The wavelet coefficients are, in turn, the subject of comments on continuous analysis (see section 6.4).

6.2.1. *The analyzed signals*

The signals are organized by topics illustrating one or more possibilities of wavelet decomposition:

– *identifying periodic signals.* This involves comparing wavelet and Fourier transforms. We examine a superposition of sine p and then two sines of different frequencies side by side;

– *recognizing noises.* We examine the wavelet decomposition of white and colored noise;

– *locating breakdowns and discontinuities.* Three signals contain breakdowns or discontinuities of the derivative of the underlying functions. They make it possible to illustrate the capacity of wavelets to highlight them: “a step”, “two proximal discontinuities”, “a discontinuity of the second derivative”;

– *decomposing into trend + seasonal component + noise.* We are interested in the decomposition of three signals:

- the first two, which are simulated, lend themselves to a classical additive decomposition into trend, seasonal component and noise,

- the third, a real signal, is a piece of the electrical load curve for June 1990 of the French electrical company EDF. The sampling period is one minute and the time series is, up to a transform, the consumption of electricity in France. The description goal led to performing a multiscale analysis of this signal;

– *locating events*. This involves detecting a typical frequency signature in an electroencephalogram.

6.2.2. *Processing carried out*

For the notations and concepts relating to the mathematical framework see Chapter 2.

In the analyses in this section we use wavelets of the *dbN* family and *sym8* (see Chapter 4). Practically speaking, we handle two types of objects:

– *wavelet coefficients* (i.e. detail coefficients). They are especially useful to recognize breakdowns or discontinuities. They are the co-ordinates in the orthonormal bases of W_k repeated 2^k times in order to allow a reading synchronous with the signal. Wavelet coefficients will not be used here because from our point of view they bring less information than reconstructed signals;

– *details and trends (or approximations) reconstructed* in the original time. The simultaneous examination of approximations and details proves very useful for the multiscale aspects and we will cover this point in detail, particularly from the practical point of view. Processing techniques are based on this idea, which we will now specify.

In order to decompose a discrete signal we usually choose to assimilate it to its co-ordinates in the basis of V_0. We will do that noting, however, that other choices are possible, such as, for example, interpolation followed by projection. In addition, the signal is available on a finite time interval. This induces edge effects that are local effects becoming weaker as the support of the analyzing wavelet decreases in size and the resolution increases. Let us recall that, if the discrete time available sample is of length L, the wavelet coefficients, which are the co-ordinates in the basis of W_j (space of details at level j), are of approximately $\frac{L}{2^j}$. The same holds true for the coefficients of the corresponding approximation. These sequences of coefficients are reconstructed in V_0, i.e. by changing basis we get back in the same space as the basic signal (by using inclusions of V_j and of W_j in V_0, for $j > 0$), and thus have a length of L.

Thus, we decompose the processed signal S into a sum of orthogonal signals corresponding to different time scales, i.e.:

$$S = A_j + D_j + \cdots + D_2 + D_1$$

where, if the choice of the wavelet is neglected, we have:

– A_j, the approximation or trend at level j containing the components of S of "period" larger than the 2^j;

– D_k, the detail at level k containing the components of the signal of "period" included between 2^{k-1} and 2^k.

For each signal we simultaneously examine: details D_1 to D_j in this ascending order and with respect to the approximations A_1 to A_j.

Each signal is studied according to a common pattern including the following headings:

– the analyzed signal;

– the analyzing wavelet;

– the number of decomposition levels;

– comments on the graphs;

– possible additions to be tested by the reader.

6.2.3. *Commented examples*

6.2.3.1. *A sum of sines*

Analyzed signal (Figure 6.1): $s_1(t) = \sin(3t) + \sin(0.3t) + \sin(0.03t)$.

Analyzing wavelet: *db3*, 5 levels of decomposition.

Comments

The signal consists of the sum of three sines: "slow", "medium" and "fast" with respect to the sampling period of 1, the corresponding periods are approximately 200, 20 and 2 respectively. We should thus essentially see the "fast" sine in D_1, the "medium" sine in D_4 and, finally, the "slow" sine in A_4. The slight differences noted in the decompositions are due, in particular, to the sampling period and the wavelet choice.

The detail D_1 contains mainly the components of s_1 for a period between 1 and 2, i.e. the "fast" sine, but its period is not visible on the graph. Zooming in on D_1 reveals that each antinode consists of 10 oscillations and makes it possible to estimate the period. We find a period close to 2. The detail D_2 has a very small amplitude. This is clear from the approximations: the first two resemble each other since $A_1 = A_2 + D_2$. The details D_3 and, especially, D_4 contain the "medium" sine. Indeed, we observe a breakdown between approximations 3 and 4.

Approximations 1 to 3 make it possible to estimate the period of the medium sine and thus, there only remains the "slow" sine appearing in A_4. The distance between two successive maxima is 200, which is the period of the slow sine. The latter, still visible in A_5, disappears from the approximation and move into the detail at level 8.

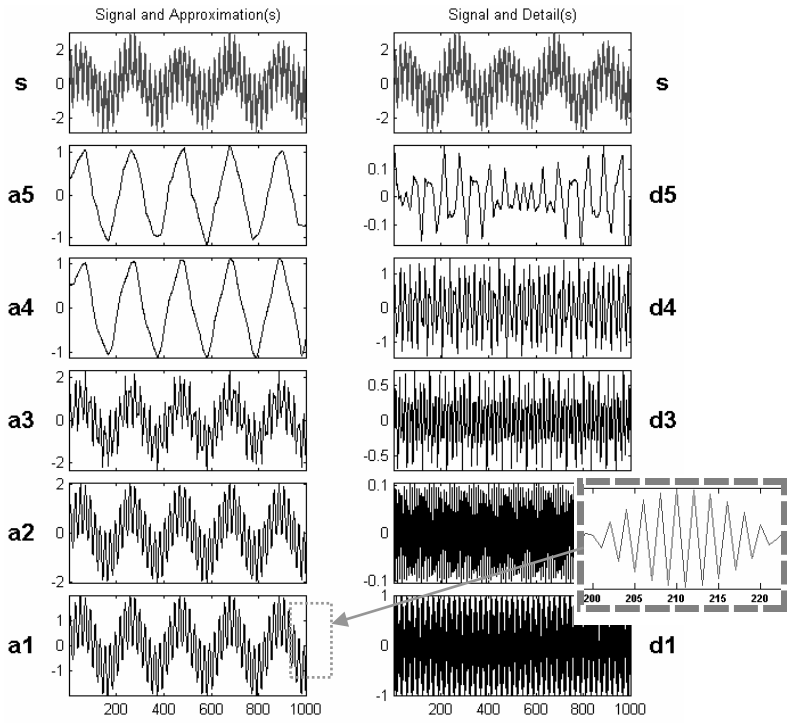


Figure 6.1. *A sum of sines*

Supplements

– Test other wavelets (few changes are notable).

– Test other linear combinations of periodic signals: dyadic period sines, periodized blocks with dyadic period.

6.2.3.2. *A frequency breakdown*

Analyzed signal (see Figure 6.2): two sines side by side.

$$s_2(t) = \begin{cases} \sin(0.03t) \ \textit{if} \ \ t \leq 500 \\ \sin(0.3t) \ \ \textit{if} \ \ t > 500 \end{cases}$$

Analyzing wavelet: *db5*, 5 levels of decomposition.

Comments

The signal consists of a "slow" sine and a "medium" sine, on either side of position 500. These two sines are not continuously connected.

Details D_1 and D_2 make it possible to detect this discontinuity. It is localized very precisely: only a small zone around 500 contains large details. This stems from the fact that the rest of the signal does not have such high frequencies. Let us note that if we were interested exclusively in the location of discontinuity, *db1* or *db2* would be more useful than *db5*.

Details D_3 and D_4 contain the "medium" sine as in the previous analysis.

The "slow" sine appears in A_4. It is more regular than in the previous analysis since the regularity of *db5* is greater than that of *db3*.

The same signal analyzed by the Fourier transform would not have made it possible to detect the moment of change of signal frequency, clearly highlighted here.

Supplements

– Test other wavelets (few changes are notable).

– Analyze a similar signal, but continuous in 500.

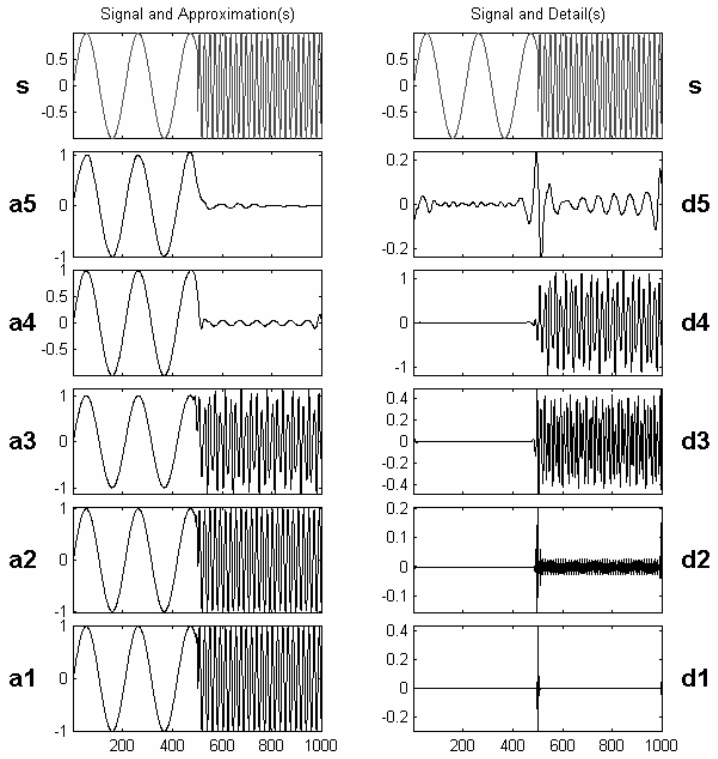


Figure 6.2. *Two sines side by side*

6.2.3.3. *White noise*

Analyzed signal (see Figure 6.3): uniform white noise over $\left[-\frac{1}{2},\frac{1}{2}\right]$.

Analyzing wavelet: *db3*, 5 levels of decomposition.

Comments

At all levels we find noise-type signals which, as we will see, are irregular. This stems from the fact that for a white noise all the frequencies carry the same energy.

The variances, on the other hand, decrease regularly from one level to the next, as can be seen by an ascending reading of the graphs of details and approximations. The reduction in variance is by a factor 2 between two successive levels, i.e. $\mathrm{var}(D_j) = 2\,\mathrm{var}(D_{j+1})$. The effect of such a noise is thus blurred at low resolutions.

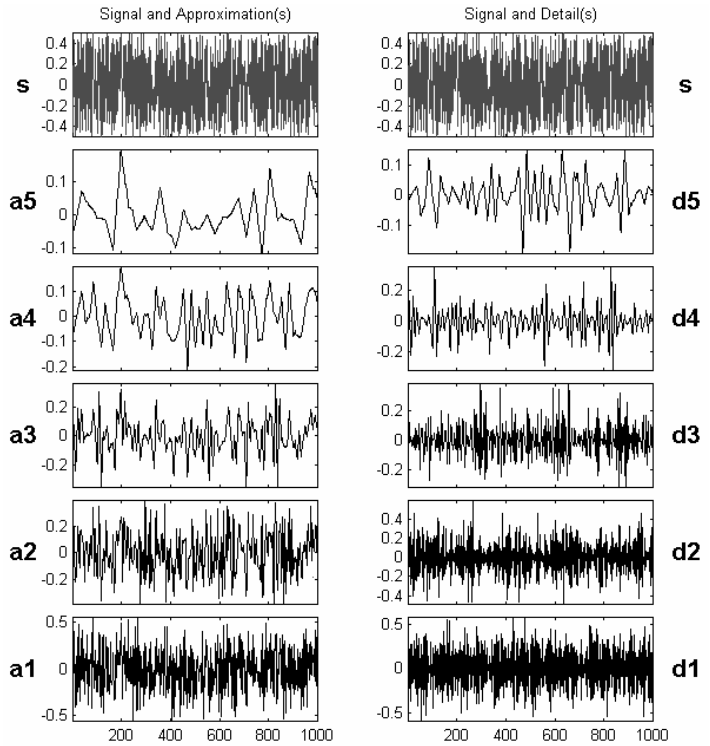


Figure 6.3. *Uniform white noise*

Let us note, finally, that the details and approximations are not white noises and that these signals are increasingly dependent as the resolution decreases. We are dealing with moving averages, whose order depends on the length of the filters associated with the wavelet and on the level j. On the other hand, wavelet coefficients are non-correlated random variables. This property cannot be read in the reconstructed signals examined here.

Supplements

Test a Gaussian white noise and wavelets associated with longer filters.

6.2.3.4. *Colored noise*

Analyzed signal (see Figure 6.4): autoregressive noise AR (3).

$$b_2(t) = -1.5\,b_2(t-1) - 0.75\,b_2(t-2) - 0.125\,b_2(t-3) + b_1(t) + 0.5$$

where $b_1(t)$ is the uniform noise studied in the preceding section.

Analyzing wavelet: *db3*, 5 levels of decomposition.

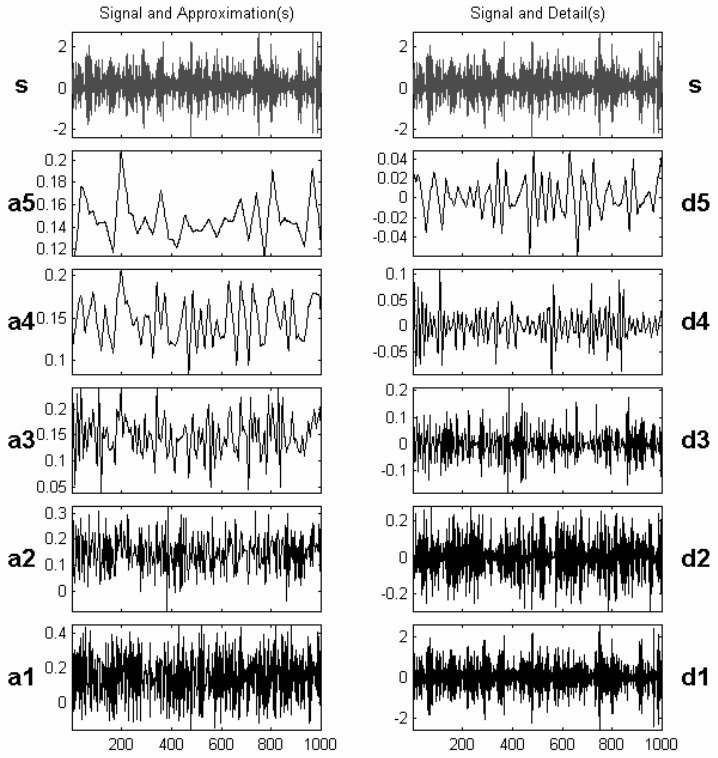


Figure 6.4. *Noise AR (3)*

Comments

This example has to be examined with respect to the previous one, since, contrary to the case of white noise, here we are dealing with a colored noise whose spectrum primarily loads high frequencies.

It is therefore mainly found in D_1, which contains the major portion of the signal. In the graphs the details are on a larger scale than the trends.

In this situation, frequently encountered in practice, the effects of the colored noise in the analysis grow blurred much more quickly than those of white noise.

In A_3, A_4 and A_5 we find the same scheme as in the analysis of white noise, from which this colored noise is constructed using linear filtering.

Supplements

– Test other ARMA, for example:

$b_3(t) = -1.5\,b_3(t-1) - 0.75\,b_3(t-2) - 0.125\,b_3(t-3) + b_1(t) - 0.7\,b_1(t-1)$

– Test an integrated ARMA (ARIMA), for example: $b_4(t) = b_4(t-1) + b_3(t)$.

6.2.3.5. *A breakdown*

Analyzed signal (see Figure 6.5): a step signal.

Analyzing wavelet: *db2*, 5 levels of decomposition.

Comments

Here we deal with the simplest example of a breakdown: a step function. The jump occurs at time instant 500.

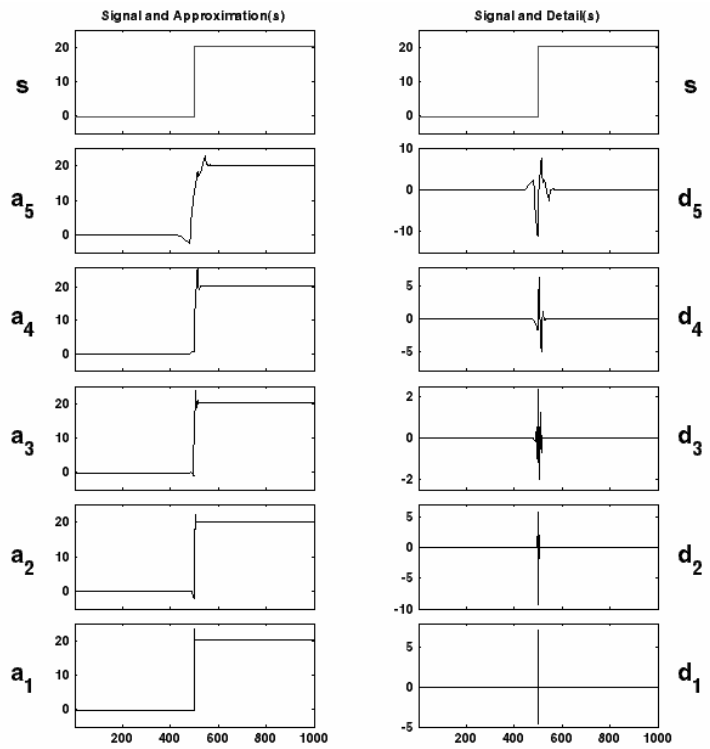


Figure 6.5. *A step signal*

The breakdown is detected at all levels but, of course, with greater precision at high resolutions (levels 1 and 2) than at low ones (levels 4 and 5). It is very precisely localized at level 1 where only a very small zone around the jump time is visible.

Let us note that in fact the reconstructed details each essentially contain a basic wavelet in V_0.

Let us note, finally, that the localization of the breakdown is better the shorter the wavelet associated with a filter is.

Supplements

– Test other wavelets.

– Replace the step function by an impulse.

– Add a noise to the signal.

6.2.3.6. *Two breakdowns of the derivative*

Analyzed signal (see Figure 6.6): two proximal discontinuities of the derivative.

Analyzing wavelets: *db2* and *db7*, 4 levels of decomposition.

Comments

The signal consists of two straight lines with the same slope, on both sides of a very short platform starting at 500. Besides, the platform is not very visible over the original signal with the naked eye.

Two analyses are performed: one with the highly localized wavelet *db2*, which has a short filter, is presented in the left-hand column, and the other, with the wavelet *db7* and a longer filter, is shown in the right-hand column.

In both analyses the platform is very well detected: D_1 is zero outside a limited zone. The presence of two singularities is clear for *db2*, for which we distinguish the initial and final time instants of the platform. On the contrary, for *db7* two discontinuities of the first derivative are fused and only the whole platform is "visible".

This example suggests choosing wavelets associated with short filters to distinguish proximal discontinuities. The examination of other levels of detail again illustrates the lack of precision of detection at low resolutions. Both wavelets *db2* and *db7* filter segments of the straight line and analyze discontinuities because they are orthogonal to the polynomials of a degree lower or equal to one.

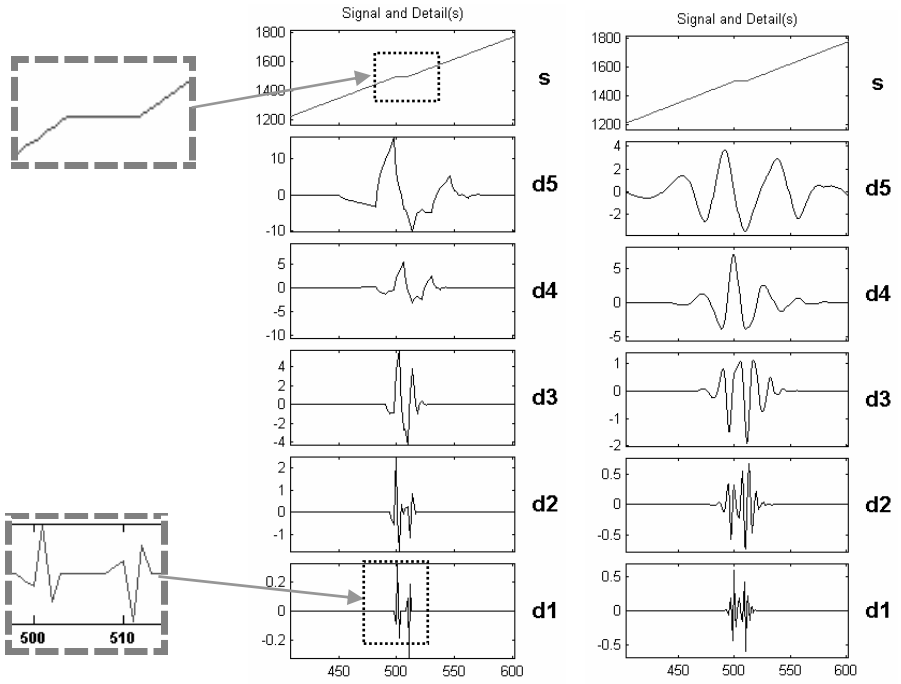


Figure 6.6. *Two proximal discontinuities of the derivative*

Supplements

– Test other wavelets.

– Vary the size of the platform.

– Add a noise.

6.2.3.7. *A breakdown of the second derivative*

Analyzed signal (see Figure 6.7): a discontinuity of the second derivative

$$s_3(t) = \begin{cases} \exp(-4t^2) \ \textit{if} \ \ t < 0 \\ \exp(-t^2) \quad \textit{if} \quad t \geq 0 \end{cases}$$

Analyzing wavelets: *db1* and *db4*, 2 levels of decomposition.

Comments

This example shows that the number of vanishing moments can be an important criterion in the choice of the wavelet. The function consists of two exponentials connected at 0 and the analyzed signal is the sampling of the continuous function with steps of 10^{-3}. It is analyzed by two wavelets: *db1*, wavelet with low regularity, on the left and *db4*, a more regular wavelet, on the right.

For *db4*, on the right, the discontinuity is well detected: the details are large only in 0 and zero everywhere else. It is the only element arising from the analysis.

On the left, for *db1*, the situation is completely different: the black zones correspond to very fast oscillations of the details. These values are equal to the difference between the function and an approximation by a piecewise constant function. The value of details is very small, about 10^{-3} for *db1*, because s_3 is very smooth and does not contain high frequencies. This order of magnitude is even smaller for *db4* (about 10^{-6}) because the associated basis functions are more regular than for *db1*. The slow decrease of the amplitude of details as *t* varies from -0.5 to 0 and the slow growth from 0 are explained by the fact that the derivative of s_3 is zero in 0 and continuous.

In this example we note that it is necessary to take a wavelet with sufficient vanishing moments in order to detect only the singularity. Indeed, the function s_3 can be written in the form:

$$s_3(t) = \begin{cases} 1 - 8t^2 + t^4 R_g(t) & \text{if } t < 0 \\ 1 - 2t^2 + t^4 R_d(t) & \text{if } t \geq 0 \end{cases}$$

We will denote $R_g(t)$ and $R_d(t)$ simply $R(t)$.

The wavelet *db4* has four vanishing moments and, consequently, the details reflect the fluctuations of $t^4 R(t)$ that are small around 0.

The wavelet *db1* has only one vanishing moment and therefore the details contain variations of $bt^2 + t^4 R(t)$. Those of the term bt^2 dominate, which explains the existence of oscillations and their amplitude, different on either side of 0, since the coefficient of t^2 changes.

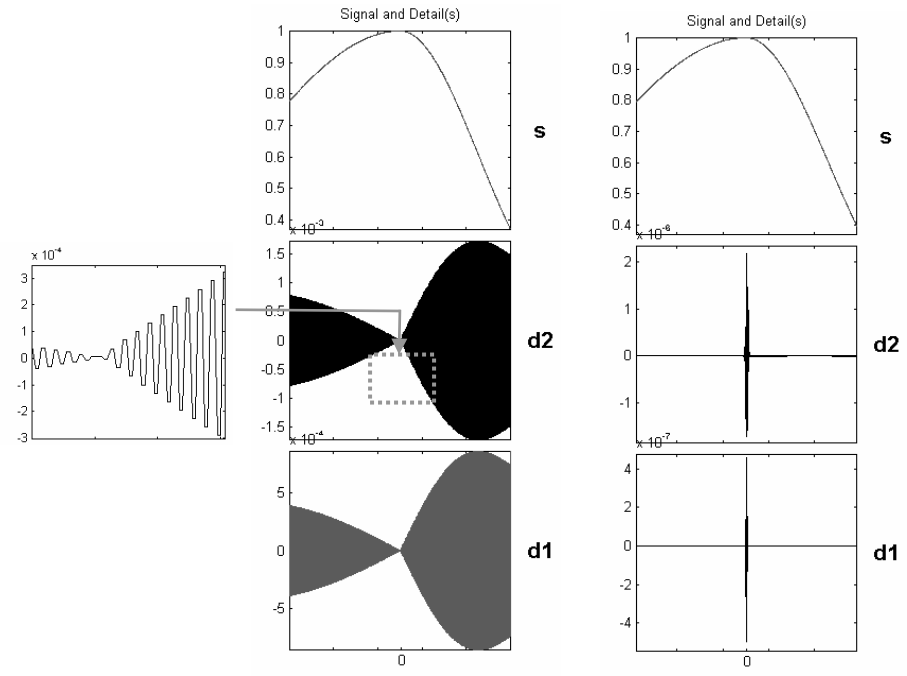


Figure 6.7. *A discontinuity of the second derivative*

Supplements

– Add a noise. If it is sufficiently large, it masks the discontinuity and the detection fails.

– Sample less frequently; the analysis is sensitive to the sampling period.

6.2.3.8. *A superposition of signals*

Analyzed signal (see Figure 6.8): a triangle + a sine + a noise.

Analyzing wavelet: *db5*, 7 levels of decomposition.

Comments

The signal examined here is the sum of three components: a triangle, a sine and a noise. The sine $\sin(0.3t)$ is the one already seen in signals s_1 and s_2 studied previously. The noise is the uniform white noise presented in section 6.2.3.3, but divided by 3.

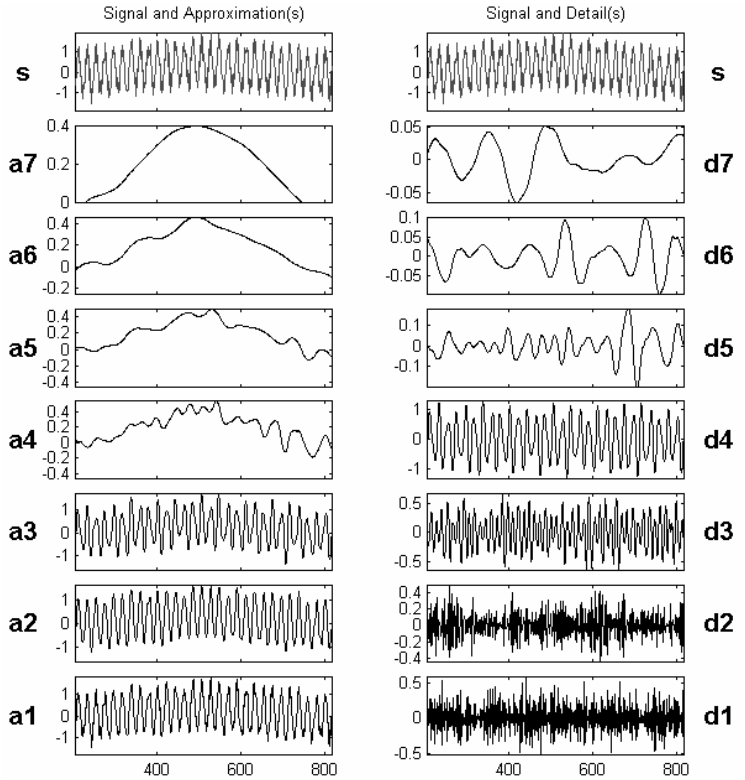


Figure 6.8. *A triangle + a sine + a noise*

Details D_1 and D_2 are due to the noise. We will be able to compare them with those of noise analysis. Details D_3 and, especially, D_4 are due to the sine. There again we will be able to compare with the details D_3 and D_4 of the analyses of s_1 and s_2. Higher level details are increasingly small and come from the noise. The approximation A_7 contains the triangle, generally well recovered. Its "round" form is partly due to the regularity of the wavelet *db5*.

Supplements

– Replace the triangle by a polynomial.

– Replace the white noise by an ARMA noise.

6.2.3.9. *A ramp with colored noise*

Analyzed signal (see Figure 6.9): a ramp + an ARMA noise.

Analyzing wavelet: *db3*, 6 levels of decomposition.

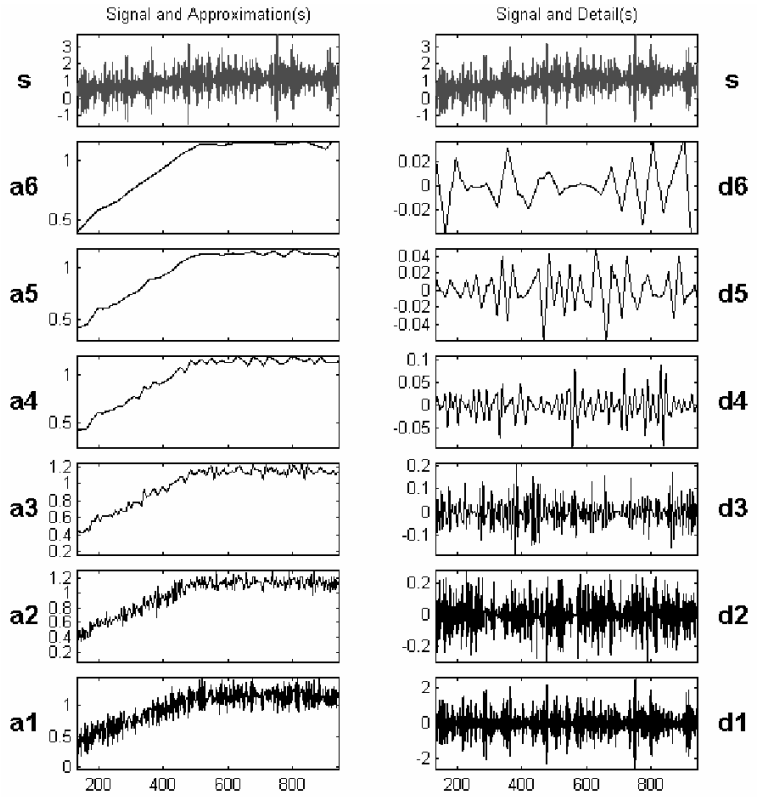


Figure 6.9. *A ramp + an ARMA noise*

Comments

The signal is built using the "trend" plus noise scheme. The "trend" is a slow linear rise from 0 to 1 until the time instant $t = 500$ and is constant afterwards. The noise is the centered AR(3) noise varying roughly between –3 and 3, examined in section 6.2.3.4.

A priori the situation seems difficult from the point of view of the separation of the ramp and the noise. Indeed, the order of magnitude of the noise is six times larger than that of the ramp. In fact, that is not important: the two components of the signal being well separated in frequency. The approximations are quite acceptable from level 3, with satisfactory reconstructions of the ramp.

Wavelet analysis being additive, the details are those of the previously analyzed colored noise, except for a small area around 500. In the same way, the approximations are equal to those of the colored noise, plus smoothed versions of the ramp.

Supplements

Test others wavelets.

6.2.3.10. *A first real signal*

Analyzed signal (see Figure 6.10): piece of an electrical load curve.

Analyzing wavelet: *db3*, 5 levels of decomposition.

Comments

The shape of the series is a central peak preceded and followed by a hollow, and then a second definitely less pronounced peak.

Level 1 and 2 details have the same order of magnitude and account for local irregularities due to the noises. The detail at level 3 presents large values at the beginning and the end of the main "peak" and makes it possible to locate the corresponding hollows. Level 4 detail causes the coarser morphological aspects of the series to appear (three successive peaks) and conforms remarkably well to the shape of the curve. They are the essential components of the signal with periods shorter than half an hour.

This is also seen very clearly in the approximations: A_1 and A_2 look very much alike and A_3 still constitutes a reasonably accurate approximation of the original signal. By examining A_4, however, we note a considerable loss of information. This information consists of detail at level 4.

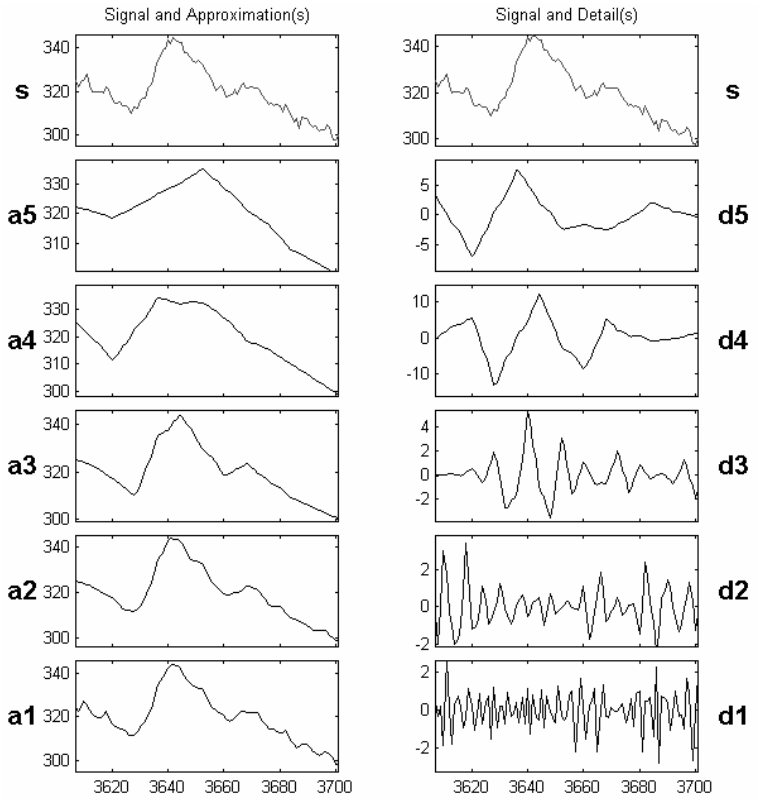


Figure 6.10. *An electrical load curve*

In the case of this record, it is the multiscale aspect that is the most interesting and the one carrying most information: the essential components of the electrical signal to supplement a bi-hourly description (homogenous with the level 5 approximation) are the components with a "period" ranging between 8 and 16 minutes.

6.2.3.11. *A second real signal*

Analyzed signal (see Figure 6.11): piece of an EEG.

Analyzing wavelet: *sym8*, 5 levels of decomposition.

Comments

It is well-known that electroencephalograms (EEG) which reflect the electrical activity of the brain, are non-stationary signals: their spectral characteristics (i.e. in the frequency domain) change with time.

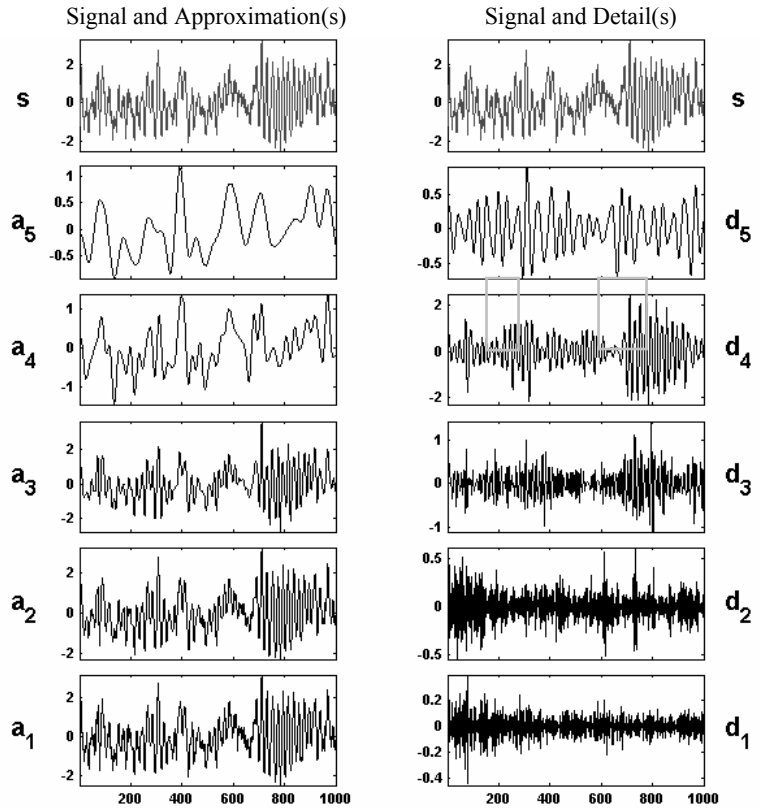


Figure 6.11. *An EEG*

In order to interpret such signals, it is interesting to identify the time instants of change and to locate the relevant frequencies (or scales). For the signal analyzed here we seek to locate the areas associated with the position from 700 to 900, for the main one, and from 200 to 350 for the secondary one.

The examination of the orders of magnitude of the fluctuations of details shows that the level 4 detail is the most important.

The examination of approximations, from the finest A_1 to the coarsest A_5, leads to observing that passing from level 3 to level 4 causes the clearest break. This confirms the importance of the detail at level 4 since $A_4 - A_3 = D_4$.

Details at levels 1 and 2 primarily capture the local irregularities caused by noise and the abrupt changes of the signal at the positions 80, 620 and 730.

Detail D_4 isolates the two areas sought in this analysis (see boxes appearing in detail 4). They correspond to zones locally homogenous in frequency and of considerable amplitude.

Wavelet analysis makes it possible to easily capture useful information in level 4 detail that we could then process and analyze.

6.3. The contribution of analysis by wavelet packets

Wavelet packets are a generalization of orthogonal wavelets. They enable a finer analysis by decomposing not only the approximation but also the detail at each level. Introduced to deal with the lack of frequency resolution in the wavelet analysis, their principle is to some extent to cut the details up in frequency bands (see Chapter 2). Consequently, we construct representations of coefficients in the time-frequency plane that are adequate to highlight their contribution.

Here we illustrate the improvement brought by wavelet packets for the frequency resolution of the analysis with three examples: two chirps (linear and quadratic) followed by a periodic signal and finally, in the last example, we consider a more composite signal.

6.3.1. *Example 1: linear and quadratic chirp*

The first signal presented is a "linear chirp", i.e. an oscillating signal whose frequency $250\pi t$ increases linearly with time.

We consider the function $y = \sin\left(250\pi t^2\right)$, $t \in [0,1]$ and take a sampling of length 512. In Figure 6.12 we present an analysis at level 4 by wavelets (on the left) and by wavelet packets (on the right) using the wavelet *db2*.

Decomposition by wavelets (stemming from the analysis by packets) and complete decomposition by wavelet packets are represented in the same way. The horizontal bands associated with the packets are arranged in order of ascending frequency, lower frequencies being at the bottom. The coefficients are colored according to their absolute value.

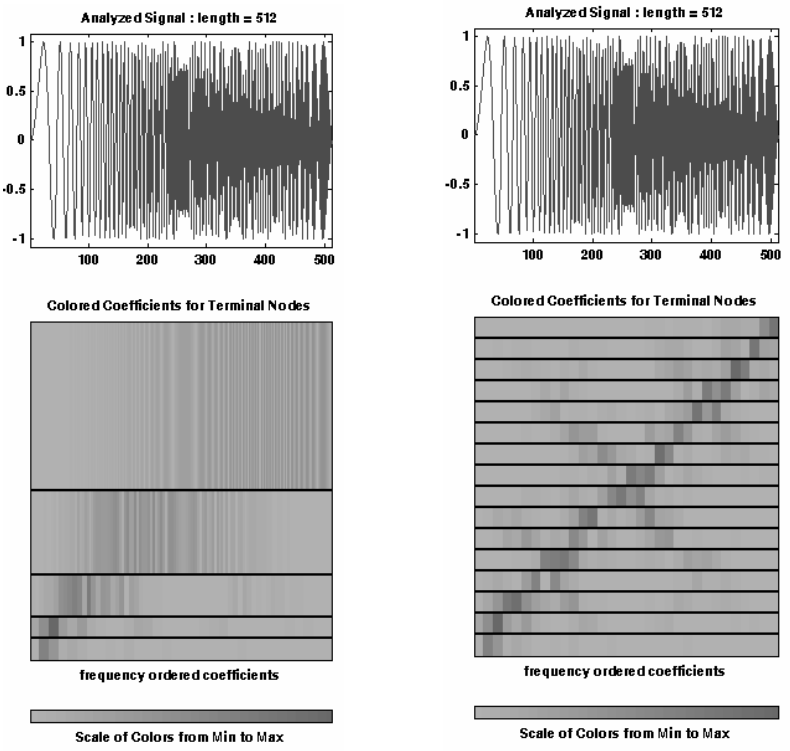


Figure 6.12. *Decomposition of a linear chirp: on the left using wavelets and on the right using wavelet packets*

The height of the colored bands is inversely proportional to the scale of the corresponding node in the underlying tree, which represents the organization of the wavelet packets (see section 2.5.3). Consequently, this height is proportional to the associated frequency bandwidth. Thus, for wavelet decomposition, on the left, we have (from top to bottom) bandwidths which are respectively proportional to 1/2, 1/4, 1/8 and 1/16 for details at level 1 to 4, and 1/16 again for approximation at level 4. On the right, for wavelet packets decomposition, each node of the tree is on the same scale 2^4 and, thus, all the frequency bands have the same width of 1/16.

For the wavelet analysis it is very difficult to identify the property of linearity of frequency. On the other hand, on the representation of the packet coefficients, it is obvious that those with the largest absolute value are organized following a line.

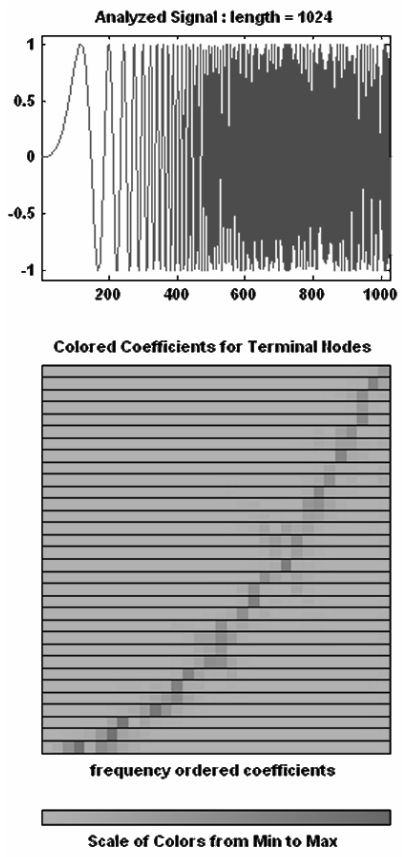


Figure 6.13. *Decomposition of a quadratic chirp into wavelet packets with sym12*

We can perform the same experiment with a "quadratic chirp", i.e. with a function of the type $y = \sin\left(kt^3\right)$ whose frequency is kt^2. In Figure 6.13 we find the decomposition at level 5 using wavelet packets of such a signal using the wavelet *sym12*.

We then clearly identify the quadratic time-frequency dependence on the representation of the packet coefficients: the largest in absolute value are distributed in the time-frequency plane following a parabola.

6.3.2. *Example 2: a sine*

The signal is a sampling with a length of 256 of a sinusoid with a period 8. We use the Haar wavelet and carry out a wavelet analysis at level 7 (on the left in Figure 6.14) and a same level analysis by wavelet packets (on the right).

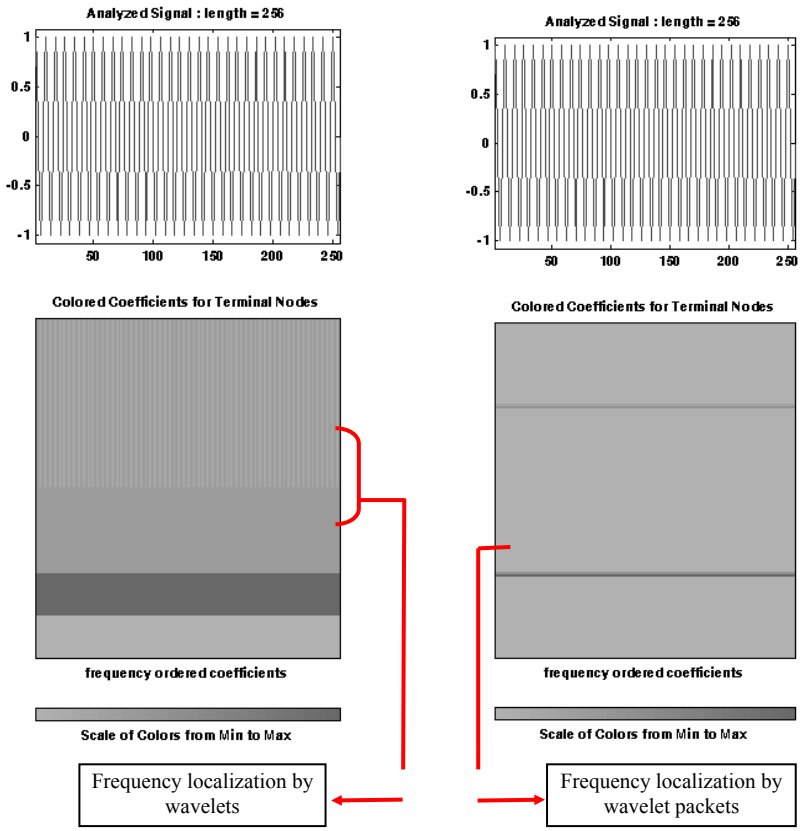


Figure 6.14. *Decomposition of a sine: on the left using wavelets and on the right using wavelet packets*

The frequency representation and the mode of coloring of the coefficients are the same as in the previous example. The x-axis represents time and the y-axis represents the growing frequencies, from the bottom to the top.

The wavelet decomposition localizes the period of the sine while only expressing that it belongs to the interval [8, 16], which illustrates the lack of frequency

resolution. On the other hand, estimating the period is much easier by exploiting the precise information provided by the wavelet packet decomposition: instead of the interval [8, 16] on the left, we find, on the right, the period 8 associated with the darkest line. We also note the presence of another rather dark line, corresponding to period 2. It will not be retained because its presence depends on the choice of the wavelet.

6.3.3. *Example 3: a composite signal*

Let us now consider a more composite signal produced by summing two chirps, linear and quadratic, and a sine.

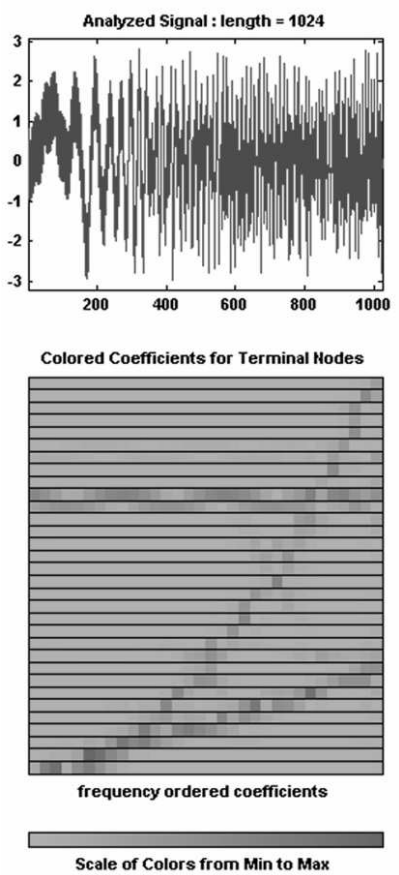


Figure 6.15. *Decomposition with sym12 in wavelet packets of the sum of two chirps and a sine*

In Figure 6.15 we find the decomposition at level 5 using wavelet packets of such a signal using the wavelet *sym12*.

We easily detect the three major signal components on the representation of the wavelet packet coefficients in the time-frequency plane. The straight line with positive slope is associated with the linear chirp, the parabola to the quadratic chirp and the straight line parallel to the horizontal axis is associated with the sine. This, in particular, illustrates the additivity of the wavelet packet analysis.

6.4. "Continuous" 1D illustrated handbook

To illustrate the usefulness of continuous analysis we will concentrate on the two following aspects:

– detection of singularities;

– improved resolution in time and scale compared to discrete analysis.

The principal difference with discrete analysis lies in the absence of the notion of approximation: there are only detail coefficients. Moreover, we do not consider reconstructed signals.

In this section, we examine three aspects:

– illustration of the time resolution and description of the cone of influence in the analysis of a singularity;

– illustration frequency resolution of the analysis;

– detailed study of the Hölderian regularity of singularities.

6.4.1. *Time resolution*

6.4.1.1. *Locating a discontinuity in the signal*

Let us take as a first example the characteristic function of an interval, which is equal to 1 on the interval and 0 elsewhere, and perform an analysis using the continuous wavelet transform. First, we use a wavelet with a small support: the Haar wavelet.

The signal (at the top of Figure 6.16) is of length 1,024 and has two discontinuities. We perform an analysis for the scales from 2 to 128. The coefficients (at the bottom of the figure) are colored according to their absolute value: the smallest appear light, while the largest are darker. We note that

discontinuities are easily located on small scales. The more the scale increases, the broader the support of the analyzing wavelet and more coefficients are thus influenced by the singularity. This explains the visual "cone" effect.

Now let us use (Figure 6.17) a wavelet with larger support, for example the "Mexican hat" (the second derivative of a Gaussian).

The coefficients are colored as previously. We note that discontinuities are perfectly located on small scales. However, the support of the analyzing wavelet grows quicker. The cone effect is still visible and, thus, much more important; on the larger scales the singularities overlap in the sense that the cones of influence are less clearly separated.

In Figure 6.18 we perform a 3D plot of the results (*db1* on the left and the Mexican hat on the right) and we then obtain a more "spectacular" effect.

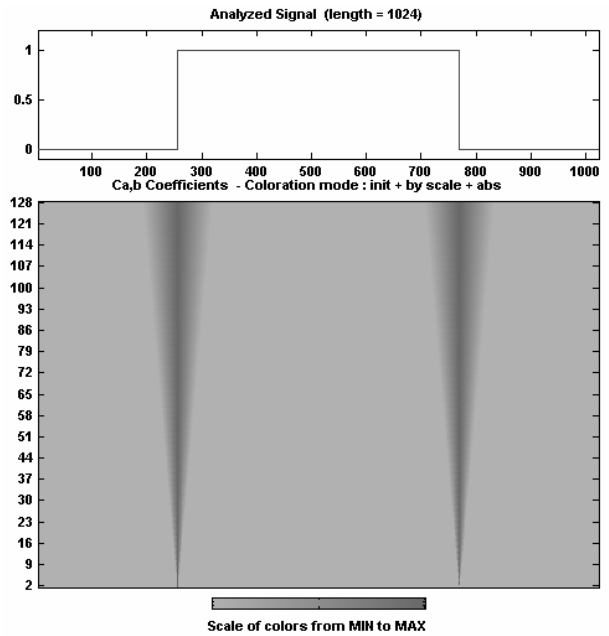


Figure 6.16. *Continuous analysis with the Haar wavelet*

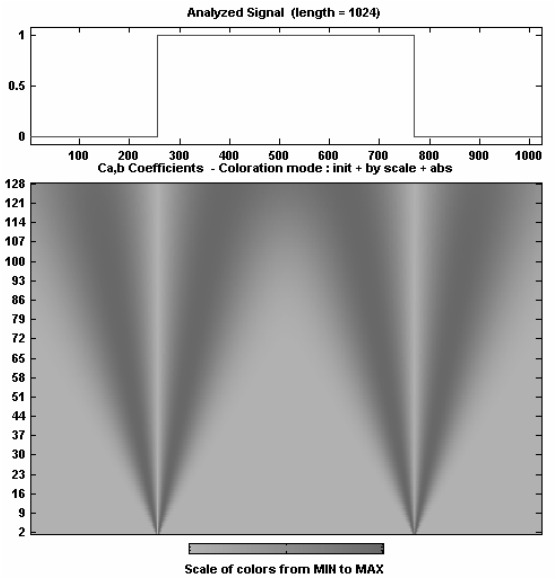


Figure 6.17. *Continuous analysis with the "Mexican hat"*

The large scales are represented in the foreground in order to highlight the cone effect. We clearly see the difference in behavior between the two wavelets for the large scales; small scales are less dissimilar.

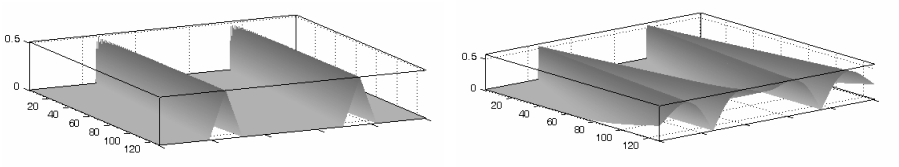


Figure 6.18. *3D representations of continuous analyses: with Haar on the left and the "Mexican hat" on the right*

6.4.1.2. *Locating a discontinuity in the derivative of the signal*

Let us take as a second example a "triangle" function, which has discontinuities in the first derivative (see Figure 6.19).

One of these discontinuities is clearly visible in the middle of the signal, whereas the others are invisible to the naked eye. Indeed, in the middle of each slope, with positions 250 and 750, there is a small invisible platform. Let us perform an analysis using a very regular wavelet: the 8th derivative of a Gaussian.

The signal of length 1,024 is analyzed for the scales from 1 to 64. The coefficients are colored from the smaller, in lighter shades, to the largest, in dark shades. All discontinuities are located perfectly, in particular on the small scales.

The cone effect is visible three times: for the platforms and for the central point. The amplitude of the breakdowns in the derivative is maximal for the central point; this is visible on the coefficients for all the scales. Also, let us note that a half-cone appears at each edge. It is due to the discontinuity introduced to the derivative by the signal extension using by zero padding.

In the last two graphs at the bottom of Figure 6.19 we find a section of the surface of the wavelet coefficients for two scales (32 and 1). For the scale a = 1 discontinuities are precisely localized in time, despite the size of the wavelet support. On the other hand, for a large scale (a = 32), the localization is much worse: for example, the singularity in t = 500 influences coefficients for the entire interval [420, 580].

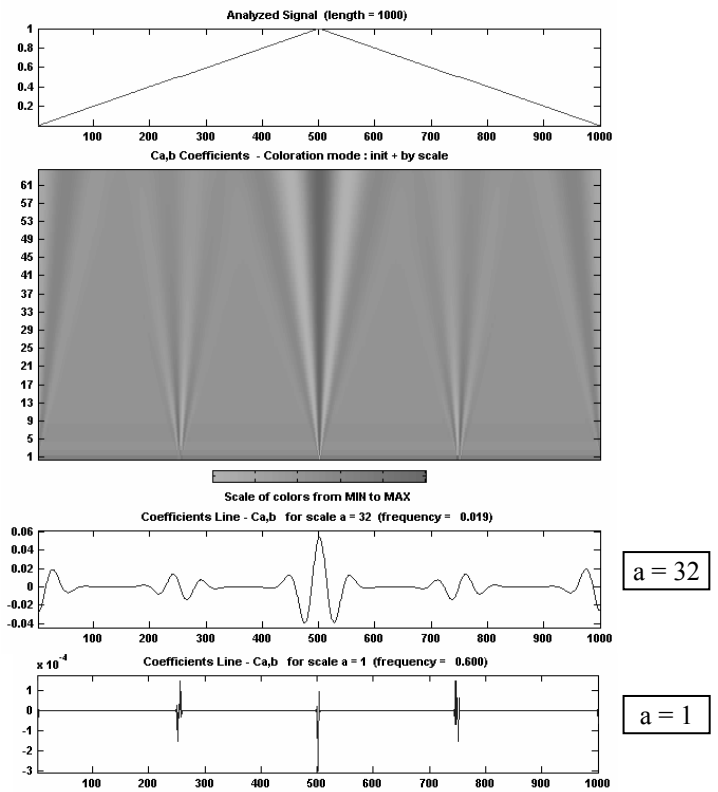


Figure 6.19. *Continuous analysis with the 8th derivative of the Gaussian*

6.4.2. *Regularity analysis*

6.4.2.1. *Locating a Hölderian singularity*

The analyzed signal is the sum of a Gaussian and a signal presenting a reflection point in $x_0 = 0.7$ (see Figure 6.20). This Hölderian singularity[2] is of exponent 0.4. More specifically:

$$f(x) = e^{-128(x-0.3)^2} - |x - 0.7|^{0.4} - Ax - B$$

The values A and B are such that $f(0) = f(1) = 0$. The function is thus very regular except at the reflection point x_0.

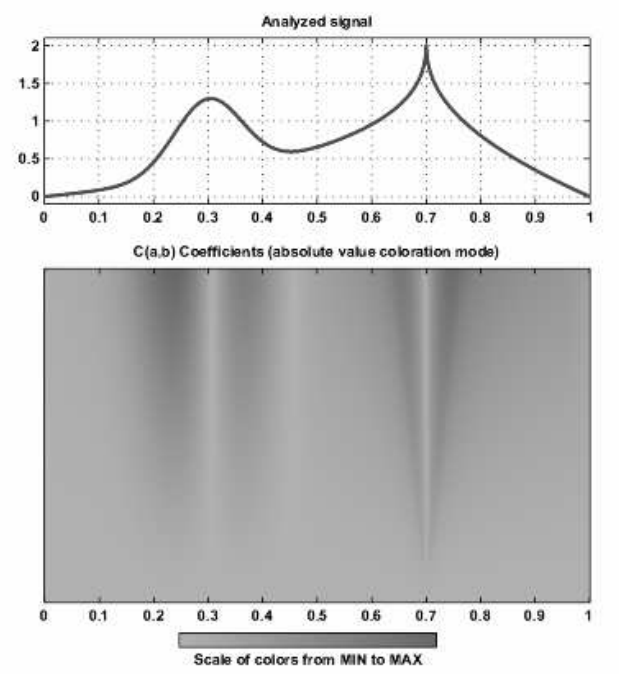


Figure 6.20. *Continuous analysis of a Hölderian singularity*

2 The function g is Hölderian of exponent α in x_0, if there is a constant C such that: $|g(x) - g(x_0)| \leq C|x - x_0|^{\alpha}$.

We perform an analysis for the scales varying from 1 to 128 with the Haar wavelet. The examination of the coefficients colored by absolute value, highlights two points:

– to the left of 0.4 a phenomenon appears as soon as the scale is sufficiently large, becoming very visible starting from scale 16. However, nothing is detectable for the small scales in this zone, which therefore does not contain singularities. The visible phenomenon is the reflection of the changes of inflection of the curve;

– to the right the singularity appears starting with the small scales and is propagated following a cone of influence of the coefficients with vertex $x_0 = 0.7$;

In fact, a more precise analysis of the Hölderian regularity of the singularity can be carried out by suitably standardizing the coefficients.

6.4.2.2. *Analysis of the Hölderian regularity of a singularity*

Let us come back to the preceding analysis, but this time representing the normalized coefficients injecting the exact Hölder exponent of the singularity, here $\alpha = 0.4$:

$$R(a,b) = \frac{\left|C_f(a,b)\right|}{|a|^{0.5+0.4}}$$

The exponent 0.5 comes from the expression of the coefficients $C_f(a,b)$. Indeed:

$$C_f(a,b) = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} f(t)\,\psi\left(\frac{t-b}{a}\right) dt$$

By a changing of variable $y = \dfrac{t-b}{a}$ we then obtain:

$$C_f(a,b) = \sqrt{a}\int_{\mathbb{R}} f(ay+b)\,\psi(y)\,dy = \sqrt{a}\int_{\mathbb{R}} [f(ay+b) - f(b)]\,\psi(y)\,dy$$

The second equality comes from $\int \psi = 0$. In this form we can take the Hölderian nature of f into account and the exponent 0.5 stems from $\sqrt{a}$ in the last equality.

The coefficients $R(a,b)$ are colored so that the smallest ones appear light and the largest ones are dark. In the last diagram of Figure 6.21 the curves $a \rightarrow R(a,b)$ are represented for scales 2, 4, 8, 16, 32, 64 and 128.

We note that for the maxima, the minima and the inflection points, as well as for all other points, except for the point $x_0 = 0.7$, the ratio $R(a,b)$ tends to 0 when $a \rightarrow 0$. Is this behavior stable when we change the wavelet?

Let us illustrate this point (Figure 6.22) by replacing the Haar wavelet that has a "small" support by the "Mexican hat" (2nd derivative of a Gaussian) whose effective support is "larger".

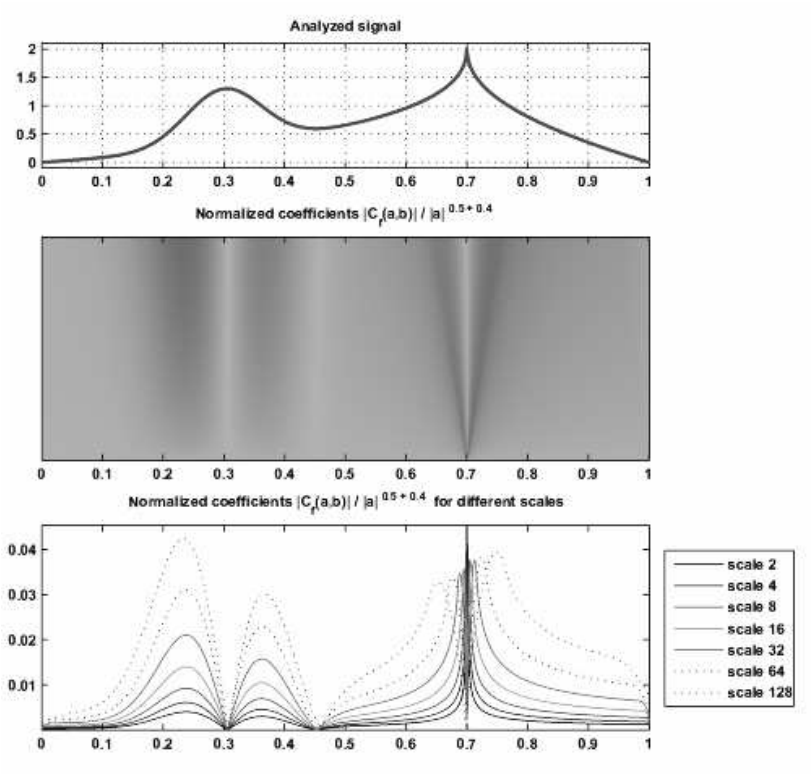


Figure 6.21. *Continuous analysis with the Haar wavelet*

The curves obtained for both wavelets are very different but the property "$R(a,b) \xrightarrow[a \rightarrow 0]{} 0$, except in $b = x_0$" remains valid. Moreover, as we can note in

the two preceding analyses, the ratio $R(a, x_0)$ is almost constant for all the scales. The local Hölderian regularity of a function is linked to the local behavior of the continuous wavelet transform. Thus, in the neighborhood of x_0 we have:

$$\frac{|f(x) - f(x_0)|}{|x - x_0|^{\alpha}} \approx Cte \quad \text{is almost equivalent to:} \quad a \to \frac{|C_f(a, x_0)|}{|a|^{0.5+\alpha}} \approx Cte$$

We can illustrate (Figure 6.23) this theoretical result by analyzing the same function using the 6th derivative of a Gaussian, on a more restricted scales band (here a varies from 1 to 64).

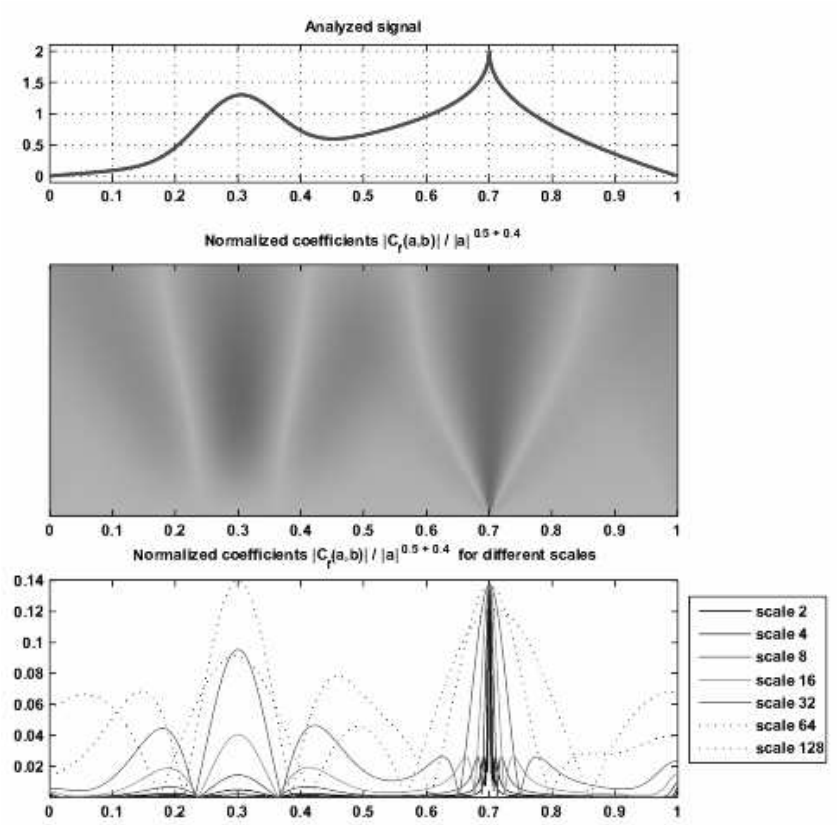


Figure 6.22. *Continuous analysis with the "Mexican hat"*

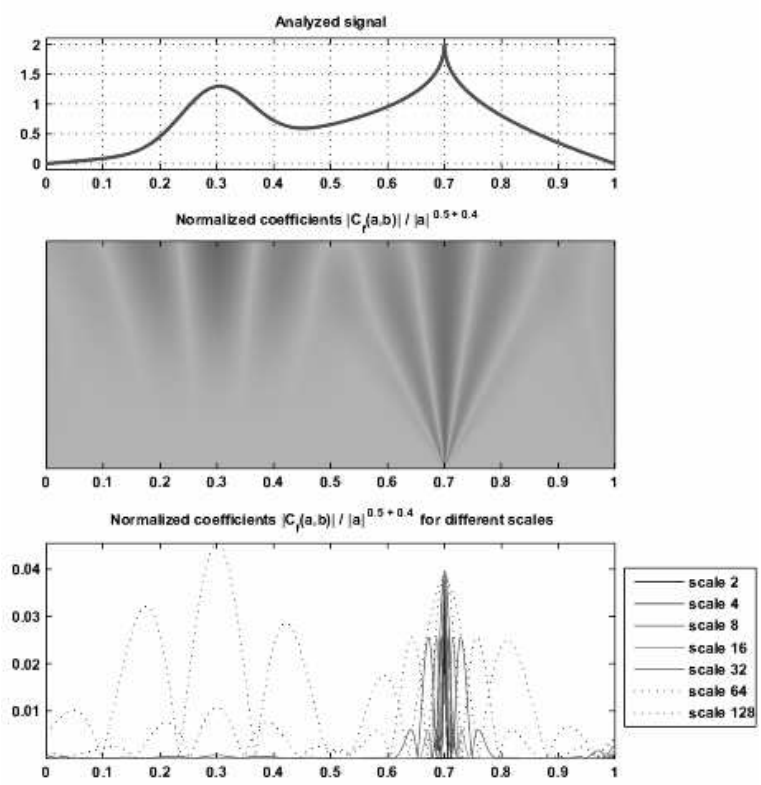


Figure 6.23. *Continuous analysis with the 6th derivative of a Gaussian*

On the surface (Figure 6.24) obtained by building a 3D representation of the coefficients we note that for the position $x_0 = 0.7$ the watershed following the scales, i.e. the curve $a \rightarrow R(a, x_0)$, is quite parallel to the dotted line $z = 0.02$.

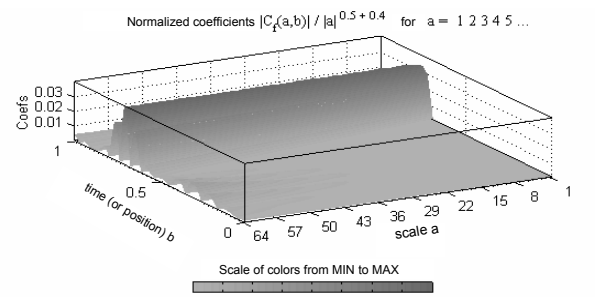


Figure 6.24. *3D representation of the continuous analysis with the 6th derivative of a Gaussian*

6.4.2.3. *Study of two Hölderian singularities*

What happens when two singularities of different Hölder exponents are present?

The analyzed signal is the sum a Gaussian and a signal presenting two Hölderian singularities (see Figure 6.25). The first one is of exponent 0.2 and the second one of exponent 0.6.

More precisely:

$$f(x) = e^{-128(x-0.5)^2} - 2|x-0.8|^{0.6} - 2|x-0.2|^{0.2} - Ax - B\,.$$

The values A and B are such that $f(0) = f(1) = 0$.

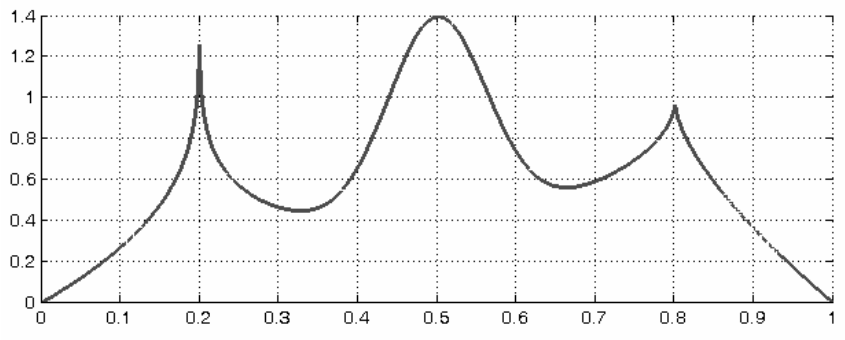


Figure 6.25. *Sum of a Gaussian and a signal with two Hölderian singularities*

We visualize the coefficients obtained by normalizing them in two different ways:

– the first (Figure 6.26), adapted to the singularity of exponent 0.2, suppresses the coefficients related to the other singularity;

– the second (Figure 6.27), adapted to the singularity of exponent 0.6, strongly amplifies the coefficients related to the first singularity. In order to improve visualization, the coefficients in the neighborhood of the first singularity were removed.

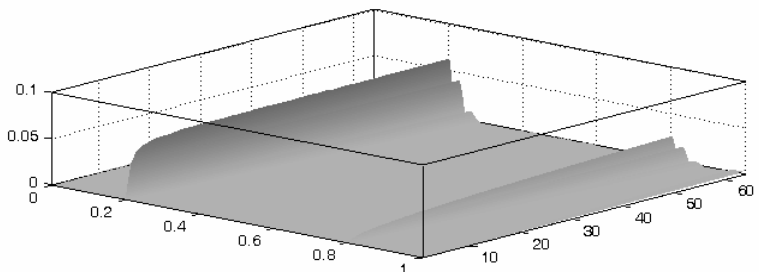


Figure 6.26. *Visualization of the normalized continuous transform:* $\frac{\left|C_f(a,b)\right|}{|a|^{0.5+\mathbf{0.2}}}$

The positions and the exponents of the two Hölderian singularities are well detected. Let us note that the position of the singularities may be located for the majority of wavelets. On the other hand, the value of the Hölder exponent requires more efforts and is not immediately readable. For more details see [DAU 92] p. 45 and [HOL 90].

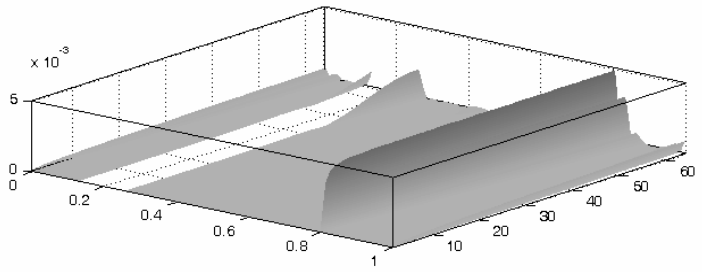


Figure 6.27. *Visualization of the normalized continuous transform:* $\frac{\left|C_f(a,b)\right|}{|a|^{0.5+\mathbf{0.6}}}$

6.4.3. *Analysis of a self-similar signal*

The analyzed signal is the "devil's staircase" associated with the triadic Cantor set (see the top part of Figure 6.28). It is an interesting signal because it is exactly self-similar.

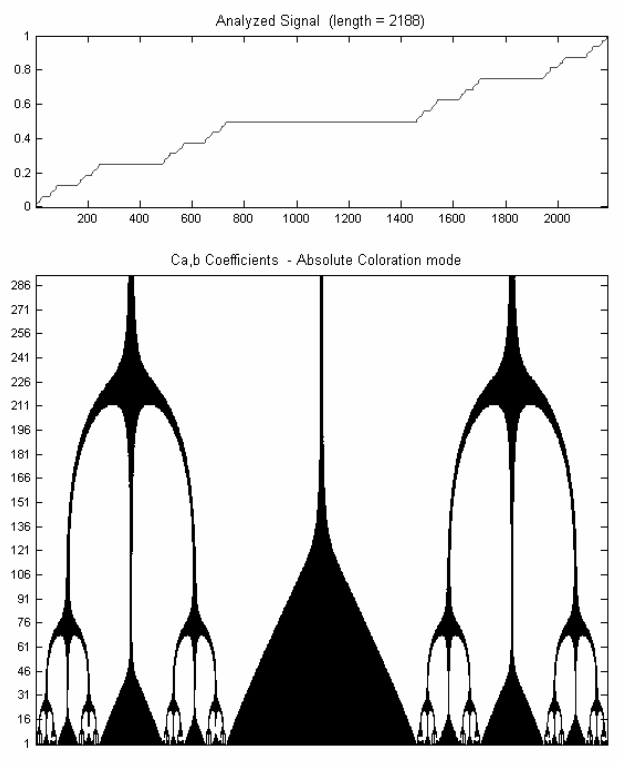


Figure 6.28. *Wavelet analysis of "devil's staircase"*

The construction principle is as follows we start with the interval $[0,1]$, which we cut into three equal intervals and we then associate the value of $\frac{1}{2}$ to the points of the medium segment $\left[\frac{1}{3},\frac{2}{3}\right]$. Once again we apply the split to the two extreme segments, and we then associate values $\frac{1}{2^2}$ and $\frac{3}{2^2}$ to the two new medium segments: $\left[\frac{1}{9},\frac{2}{9}\right]$ and $\left[\frac{7}{9},\frac{8}{9}\right]$ respectively. We apply the same process recursively to the four remaining segments.

We thus obtain a function defined everywhere on $[0,1]$ except for a set of points of zero measure, called the triadic Cantor set, to which we extend by continuity. The result is a continuous function in $[0,1]$ called "devil's staircase", which is derivable everywhere apart from the Cantor's set. Then, in order to obtain the signal to be

analyzed, we take a number in the form of $3^K + 1$ with a sufficiently large K and we sample the function obtained previously.

We perform the wavelet decomposition of this signal by the second derivative of the Gaussian, which is symmetric, for the scales between 1 and 300. In Figure 6.28 we find the wavelet coefficients that are colored by associating two colors with the absolute values of the coefficients: black for a zero value and white for other values.

In this case it is the zero values of the wavelet coefficients that carry information.

We can note an astonishing repetition of identical patterns: the self-similar structure of the signal is easily read for wavelet coefficients. The junction patter with three "legs" is repeated along the scales.

The slight difference in symmetry, visible for the position 1,800 and scales above 180, is due to the edge effects.

If we zoom in on the left part of the preceding figure (see Figure 6.29), we can specify this first visual impression.

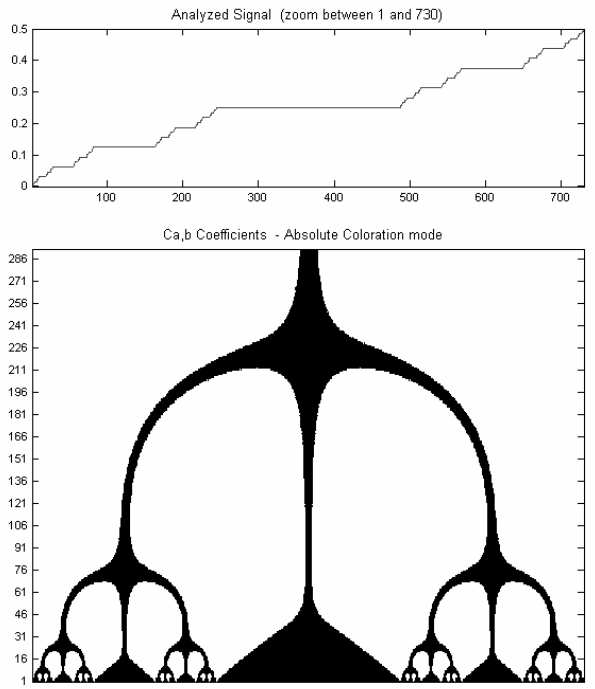


Figure 6.29. *Wavelet analysis of the "devil's staircase" (zoom)*

If one looks at the scales corresponding to the junctions of the minima lines, which are in black in the figure, we find the following scales: $a = 25$, then $a = 75$, and $a = 225$. They result from each other by multiplication by 3, which reflects the rule of signal construction in a spectacular way.

Many works deal with the analysis of deterministic or random generalizations of such situations. The former describe geometrical self-similarity as here and, more generally, what we will call the fractals. The latter relate to the self-similarity in distribution, such as for example the fractional Brownian processes. The analysis of such situations is made by means of the processing of maxima lines of the wavelet decompositions (see for example the method known as MMTO for wavelet transform modulus maxima in [ARN 95]).

Chapter 7

Signal Denoising and Compression

7.1. Introduction

Two of the great successes of wavelets are signal and image denoising and compression, which are often regarded as particularly difficult problems. This chapter tries to explain the reasons for this, by focusing on one-dimensional signals.

Denoising and the estimation of functions based on wavelet representations lead to simple and powerful algorithms that are often easier to fine-tune than the traditional methods of functional estimation. Signal compression constitutes a field where wavelet methods also appear very competitive for reasons fundamentally close to those that make wavelet-band denoising work. Indeed, the signals that we are interested in have in many cases very sparse wavelet representations and are very well represented using few coefficients. Here we deal with denoising followed by signal compression, exposing the first topic in greater detail.

First of all, we tackle the principle of denoising by wavelets, then a statistical introduction to the thresholding methods and finally some examples. We initially focus on a model of form $Y_t = f(t) + \varepsilon_t$. This model is very simple and the assumptions about noise are very strong. Fortunately, the "attraction basin" of denoising methods is much broader. Thus, we examine two extensions of this model through examples: in the case of noise with multiple rupture charges of the variance and in that of a real signal where the noise structure is unknown.

We then go further on the use of the translation invariant transform (introduced in Chapter 3) which in certain situations makes it possible to improve the results obtained by means of the preceding methods.

Moreover, we are interested in two traditional questions in statistics: estimating the probability density function and estimating the regression function in the random design case. For these two problems we outline a way of constructing estimators on the basis of wavelets enabling the use of the DWT.

Lastly, we briefly touch upon the main ideas concerning compression. The problem of compression is often synonymous with image compression, whose applications are particularly interesting and developed in Chapter 8. For example, wavelets are used in the new image compression standard JPEG 2000. Nevertheless, applications for signals are numerous, in particular the compression of audio signals.

The chapter finishes with some bibliographical comments.

7.2. Principle of denoising by wavelets

7.2.1. *The model*

Denoising consists of restoring a useful signal from observations corrupted by an additive noise.

The simplest statistical model of denoising is as follows:

$$Y_t = f(t) + \varepsilon_t \quad , \quad t = 1, \cdots, n \qquad [7.1]$$

where f is an unknown function, the variables $(Y_t)_{1 \leq t \leq n}$ are observed and $(\varepsilon_t)_{1 \leq t \leq n}$ is a centered Gaussian white noise with an unknown variance σ^2. We then have to reconstruct the signal $(f(t))_{1 \leq t \leq n}$ or to estimate the function f solely on the basis of observations.

7.2.2. *Denoising: before and after*

The problem of denoising is easy to understand by starting with an example. At the top of Figure 7.1 we see a real noisy signal.

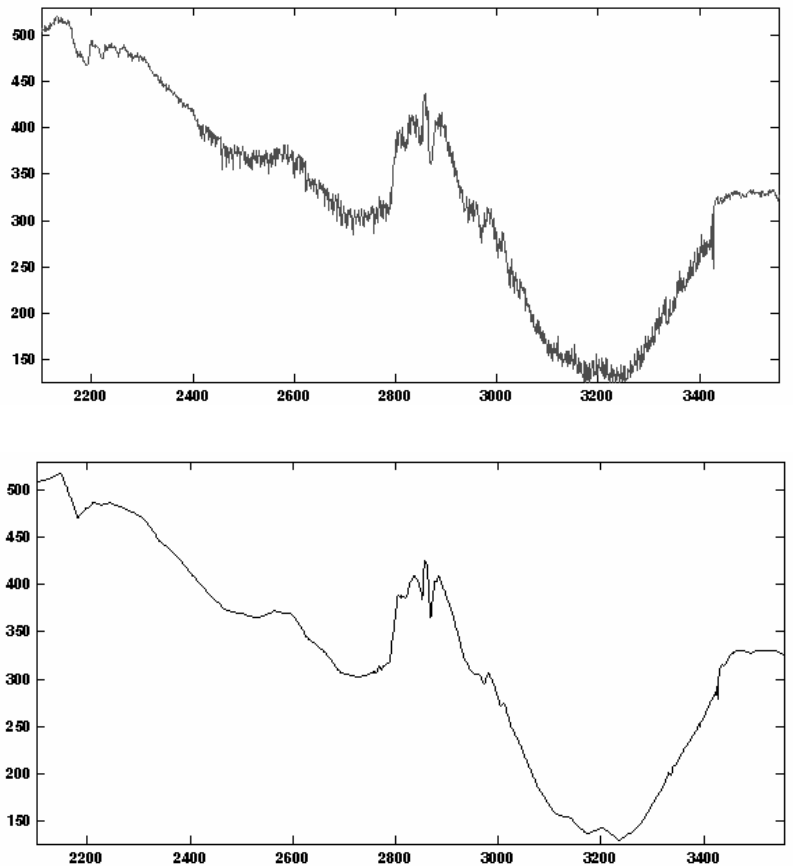


Figure 7.1. *Denoising by wavelets: before and after*

The denoised signal by wavelets is in the lower part. It is obviously well denoised, both in the zones where the underlying signal is smooth, and in the zone around the position 2,900 where the signal presents irregularities, abrupt variations. The changes in the level of noise around the positions 2,400 and 3,450 are well absorbed by the process. Traditional methods are incapable of such an adaptation in time.

7.2.3. *The algorithm*

The basic algorithm of denoising is very simple and contains three steps:

– decomposition;

– selection or thresholding of the coefficients;

– reconstruction.

Starting with the signal, we first decompose it over an orthogonal wavelet basis using the discrete transform. Then we select a part of the coefficients through thresholding and we keep the coefficients of approximation of a suitably chosen level intact. Lastly, using thresholded coefficients we reconstruct a signal by applying the inverse discrete transform to them. The signal obtained in this manner is the denoised signal.

7.2.4. *Why does it work?*

Let us quote some arguments which help us understand the effectiveness of these methods; they will be supported throughout the chapter:

– wavelet decomposition is additive, consequently, the analysis of Y is equal to the sum of the analyses of the signal f and of the noise ε;

– if noise ε is supposed to be white, then the wavelet coefficients on all the scales are white noises with the same variance;

– useful signals are, in many cases, smooth enough, except in rare locations like the beginning and the end of transitory phenomena, or ruptures, for example. This renders the decomposition of f by wavelets very sparse and is then very well represented by the coefficients of a rather rough approximation to which some large detail coefficients are added;

– if the irregularities of the signal generate larger coefficients than the scale of noise, the process of selection retains the coefficients related to the signal, provided that we can suitably estimate this scale;

– the analysis is local and, consequently, thresholding leads to a local regularization of the original signal.

7.3. Wavelets and statistics

Numerous papers cover in detail the vaster topic of function estimation. Wavelets have given a boost to the functional estimation techniques. They are particularly effective when the functions that we wish to be estimated are coded parsimoniously on the chosen wavelet basis, i.e. when few coefficients are enough to describe each function to be estimated.

These techniques are used to estimate functions of different natures: probability density, spectral density, regression function, hazard function. Moreover, the

estimators benefit from the extraordinary algorithmic speed (of linear complexity) of the discrete decomposition.

7.3.1. *Kernel estimators and estimators by orthogonal projection*

Denoising is a particularly interesting specialization of a more general problem known in statistics as function estimation. Traditional methods of estimation of the f function in the model [7.1] are of two types.

On the one hand, we have the local methods based on the kernel leading to estimating $f(x)$ by a local average of the values of Y_t for t close to x using a regularizing kernel K, for example, the Gaussian density function, and a temporal window δ:

$$\hat{f}(x) = \sum_t \left[K\left(\frac{x-t}{\delta}\right) / \sum_s K\left(\frac{x-s}{\delta}\right) \right] Y_t$$

On the other hand, we find the orthogonal projection methods, which suppose that the function f belongs to a space F, for which we have an orthogonal basis $(e_m)_m$, so that f is written as $f = \sum_m \gamma_m e_m$. The estimator is deduced from it by estimating the γ_m by $\hat{\gamma}_m$ using Y_t:

$$\hat{f} = \sum_m \hat{\gamma}_m e_m$$

In the formulae above, $\hat{f}$ indicates the estimator of f. Hereafter, all the estimated quantities will be noted in the same way (not to be confused with the Fourier transform).

7.3.2. *Estimators by wavelets*

A first idea, present very early in works using wavelets, consists of proposing estimators of the form $\hat{f} = \sum_k \hat{\beta}_{J,k} \varphi_{J,k}$ for finite energy functions f, using only the scaling function φ. They are at the same time kernel estimators (taking for K the wavelets kernel given by $K(x,y) = \sum_k \varphi(x-k)\varphi(y-k)$) and estimators by

orthogonal projection (taking $\varphi_{j,k}$ for e_m). They have the aim of estimating f by the estimate of the approximation of f of level *J*.

A second idea consists of supplementing this estimate by additional details and considering estimators of the following form:

$$\hat{f} = \sum_k \hat{\beta}_{J,k}\varphi_{J,k} + \sum_{J\leq j\leq J_0}\sum_k \hat{\alpha}_{j,k}\,\psi_{j,k}$$

by carefully choosing $\hat{\alpha}_{j,k}$ we threshold the wavelet coefficients of the signal Y (see the following section). That enables the desired adaptivity.

7.4. Denoising methods

In this section, we provide some theoretical elements concerning the methods of denoising while dwelling on the underlying ideas rather than on the detail of the impressive list of available strategies adapted to neighboring contexts. A long article of Antoniadis, Bigot and Sapatinas [ANT 01] provides a very complete assessment of these techniques and presents an extensive comparative study of these many strategies; we can refer to it to get some very beneficial information.

Let us consider an orthonormal wavelet basis on the interval obtained by gathering timescale atoms without distinguishing them by the usual double indexing. It is noted $(g_i)_{1\leq i\leq n}$. We decompose using this basis:

$$Y_t = f(t) + \varepsilon_t,\ \ t = 1,\cdots,n$$

Wavelet decomposition being additive, the analysis of Y is equal to the sum of the analyses of the signal f and of the noise ε, the coefficients are then written:

$$d_i = \theta_i + w_i,\quad i = 1,\cdots,n \qquad [7.2]$$

with a link which can be expressed by means of an orthogonal matrix W associated with the discrete wavelet transform: $d = WY$, $\theta = Wf$ and $w = W\varepsilon$.

NOTE 7.1.– the DWT transforms n given elements in the temporal field into n coefficients in the wavelet domain. The transform is linear and can be expressed by means o an n by n matrix, noted here W. Depending on the hypotheses concerning

the treatment of edge problems, this matrix can be orthogonal or "nearly orthogonal" (see [WAL 02] p. 232). Here it is presumed orthogonal and we note the functions of the underlying wavelet basis as g_i.

Since the basis to which we are projecting is orthonormal, the noises ε and w have the same probabilistic properties. Indeed, as $w = W\varepsilon$, w inherits the Gaussian and centered nature of ε and:

$$\text{cov}(w,w) = Ew^T w = E\varepsilon^T W^T W\varepsilon = E\varepsilon^T\varepsilon = \text{cov}(\varepsilon,\varepsilon) = \sigma^2 I$$

We seek an estimator $\hat{f}$ of f that minimizes the risk defined by the mean square error:

$$R(f,\hat{f}) = E\left(\left\|f-\hat{f}\right\|^2\right) = E\left(\sum_{t=1}^{n}\left(f(t)-\hat{f}(t)\right)^2\right) \qquad [7.3]$$

Due to the orthonormality of the transformation we also have $R(f,\hat{f}) = R(\theta,\hat{\theta})$.

7.4.1. *A first estimator*

The first idea, traditional in signal processing, is diagonal attenuation [MAL 98]. The estimator of the corresponding f consists of modifying the decomposition of $Y = \sum_{t=1}^{n} d_t g_t$ by attenuating the coefficients d_i using a deterministic sequence a_i. The estimator $\hat{f}$ then has the following form:

$$\hat{f} = \sum_{i=1}^{n} a_i\, d_i g_i \qquad [7.4]$$

We choose a_i by minimizing $R(f,\hat{f})$, a criterion to which we will assign a form useful to the discussion. We have:

$$E\left(\left\|f-\hat{f}\right\|^2\right) = E\left(\sum_{i=1}^{n}\left(\theta_i g_i - d_i a_i g_i\right)^2\right) = \sum_{i=1}^{n} E(\theta_i - d_i a_i)^2$$

however:

$$\begin{aligned} E(\theta_i - d_i a_i)^2 &= E(\theta_i - (\theta_i + w_i)\, a_i)^2 \\ &= \theta_i^2 (1 - a_i)^2 + a_i{}^2 E w_i{}^2 - 2\theta_i (1 - a_i)\, a_i E w_i \end{aligned}$$

The noise w_i being centered ($Ew_i = 0$) and of variance σ^2 ($Ew_i{}^2 = \sigma^2$), we have:

$$E\left(\left\|f - \hat{f}\right\|^2\right) = \sum_{i=1}^{n} \theta_i^2 \left(1 - a_i\right)^2 + \sigma^2 \sum_{i=1}^{n} a_i{}^2 \qquad [7.5]$$

This form makes it possible to calculate the optimal solution given by $a_i = \theta_i^2 / (\theta_i^2 + \sigma^2)$ and leads to $\hat{f}_a$ such that:

$$E\left(\left\|f - \hat{f}_a\right\|^2\right) = \sigma^2 \sum_{i=1}^{n} \theta_i^2 / \left(\theta_i^2 + \sigma^2\right)$$

Obviously, this solution is impossible to implement since a_i depends on θ_i, which are the coefficients of the unknown function f.

7.4.2. *From coefficient selection to thresholding coefficients*

The second idea consists of selecting certain coefficients and eliminating the others, i.e. adding the constraint $a_i \in \{0,1\}$ in the preceding section.

Let us denote again the associated estimator by $\hat{f}$. If we denote by $I = \left\{ i \in \{1,2,\ldots,n\} \mid a_i = 1 \right\}$ the set of selected indices, expression [7.5] makes it possible to write the objective criterion in the following form:

$$E\left(\left\|f - \hat{f}\right\|^2\right) = \sum_{i \notin I} \theta_i^2 + \sum_{i \in I} \sigma^2 = \sum_{i \notin I} \theta_i^2 + card(I)\sigma^2 \qquad [7.6]$$

In other words, selecting i in I costs σ^2 and excluding it costs θ_i^2, therefore, $E\left(\left\|f - \hat{f}\right\|^2\right)$ is minimal if only those indices for which $\left|\theta_i\right| \geq \sigma$ appear in I. This means that the optimal selection strategy is a thresholding of the coefficients.

Thresholding a coefficient u means that it is kept if $\left|u\right| \geq threshold$ and that it is replaced by 0 otherwise. For this strategy, the cost is:

$$E\left(\left\|f - \hat{f}_s\right\|^2\right) = \sum_{i=1}^{n} \min\left(\theta_i^2, \sigma^2\right) \qquad [7.7]$$

where $\hat{f}_s$ is the thresholded estimator.

NOTE 7.2.

– Nothing in the above section depends on the fact that the basis $(g_i)_{1\leq i\leq n}$ is a wavelet basis. The interest of wavelets, within this framework, stems from their capacity to represent f using very few significantly non-zero coefficients. However, if the basis is orthonormal, the energy of f is equal to the sum of the squared wavelet coefficients. Consequently, the energy of f concentrates in the few non-zero coefficients of f, which thus have large absolute values. The idea of thresholding consists of preserving the coefficients of Y associated with the coefficients of f that are larger than the characteristic scale of the noise.

– The key point is the least squares criterion. Here we find the strategies of selection of variables in linear regression by means of penalized least squares.

– From equation [7.6] we deduce that the error committed by the estimator $\hat{f}_M$, obtained by preserving M coefficients of the largest absolute value, has the following form: $E\left(\left\|f - \hat{f}_M\right\|^2\right) = \left\|f - \hat{f}_M\right\|^2 + M\sigma^2$; that is, the sum of the two terms, an approximation error plus a term related to random fluctuations. Starting with this penalized criterion, traditionally used in statistics, also makes it possible to obtain thresholding as the solution.

– We demonstrate that $E\left(\left\|f - \hat{f}_a\right\|^2\right) \leq E\left(\left\|f - \hat{f}_s\right\|^2\right) \leq 2E\left(\left\|f - \hat{f}_a\right\|^2\right)$, meaning that the optimal errors of the estimators $\hat{f}_a$ and $\hat{f}_s$ are of the same order. ■

For the same reasons as at the end of the preceding section, the strategy of thresholding is not applicable. We have here, in fact, an ideal estimator, which is called an "oracle" with the ability to tell us, knowing θ_i, the coefficients to be preserved. The performance of this procedure is a useful comparison base for the estimators which we will implement.

7.4.3. *Universal thresholding*

The strategy finally adopted consists of thresholding the coefficients of Y not according to the coefficients of f with the threshold σ, but directly with a threshold T suitably determined according to the noise model.

In the case of Gaussian white noise, the estimator $\hat{f}$ obtained by preserving only such coefficients of Y that $|d_i| > \sigma\sqrt{2\log(n)}$ and by replacing the others with 0 is satisfactory since it leads to almost as good results (up to a logarithmic factor) as the "oracle". Indeed, it is shown that:

$$R\left(f,\hat{f}_s\right) \leq R\left(f,\hat{f}\right) \leq (2\log(n)+1)\left(\sigma^2 + R\left(f,\hat{f}_s\right)\right)$$

where the name of the universal threshold given by Donoho and Johnstone to $T = \sigma\sqrt{2\log(n)}$ is derived from. But where does its value come from?

A centered Gaussian noise ε with a variance σ^2 is not bounded, but we know that $\lim_{n\to\infty} P\left\{\max_{1\leq i\leq n} |\varepsilon_i| > \sigma\sqrt{2\log(n)}\right\} = 0$, in other words we can consider that essentially $|\varepsilon_i| < \sigma\sqrt{2\log(n)}$.

Thus, the procedure is interpreted as the setting of all the coefficients, which could be ascribable to noise, to zero. Sometimes it also proves to be too selective in practice and other thresholds have been developed that do not exhibit this drawback.

7.4.4. *Estimating the noise standard deviation*

The threshold depends on σ, which gives the scale of the noise. How can σ, which is unknown in practice be estimated?

In the context of model [7.1] we have many estimators, but we may benefit from wavelet decomposition, which provides an elegant estimator based on the following idea.

The decomposition of Y with $Y_t = f(t) + \varepsilon_t$ is equal to the sum of the analyses of the signal f and of the noise ε. Noise ε is white, therefore its coefficients on all the scales are white noises of the same variance since the wavelets are orthonormal. Lastly, in general, the decomposition of f is very sparse and only some large level 1 detail coefficients (denoted cd^1) are not ascribable to ε. Consequently, the level 1 detail coefficients are primarily a white noise with

variance σ^2. It is then enough to use a robust estimator, i.e. not sensitive to large values of f, of the standard deviation of the level 1 detail coefficients to estimate σ. In the Gaussian case the estimator has the form:

$$\hat{\sigma} = \underset{}{median}\ \left(\left|cd^1(i)\right|\right)/0.6745$$

Here we exploit the relation between the standard deviation and the median for this distribution of errors. Indeed, for l independent Gaussian centered random variables $X_1, X_2, ..., X_l$ with a standard deviation σ, we have $E(median(|X_i|,\ 1 \le i \le l)) \approx 0.6745\sigma$.

7.4.5. *Minimax risk*

More light is shed on the performances of this type of procedure by the traditional minimax theory in the context of the function estimation in statistics. We seek estimators with good properties for very broad classes of functions. Let us note $\mathbb{F}$ the set of functions of a space *F* with a norm smaller than *M*, typically balls of Sobolev, Hölder or Besov spaces [TRI 92].

Within this framework we measure the quality of the estimator by the worst performance reached over the functions of $\mathbb{F}$: $\sup_{f\in\mathbb{F}} R\left(f,\hat{f}\right)$. We then seek estimators whose risk $\sup_{f\in\mathbb{F}} R\left(f,\hat{f}\right)$ is close to the minimum risk defined naturally by $\inf_{\hat{f}} \sup_{f\in\mathbb{F}} R\left(f,\hat{f}\right)$ and called minimax risk.

We have results showing that the estimator by thresholding introduced previously is minimax up to a logarithmic factor:

$$\inf_{\hat{f}} \sup_{f\in\mathbb{F}} R\left(f,\hat{f}\right) \le \sup_{f\in\mathbb{F}} R\left(f,\hat{f}\right) \le C.\log(n).\inf_{\hat{f}} \sup_{f\in\mathbb{F}} R\left(f,\hat{f}\right)$$

for very broad classes of functions $\mathbb{F}$ including Sobolev, Hölder and Besov spaces [HAR 98].

This type of result is valid for many variants of estimators by wavelets and for various loss functions that are not necessarily quadratic. It is all the more remarkable since it requires only a limited *a priori* knowledge of the regularity of the estimated

function, contrary to the traditional methods mentioned in the preceding section. Functional spaces, such as Besov spaces, are explored systematically and play a particular role. They make it possible to tackle functions that are rather irregular locally and may present patterns contrasting with regular zones, irregular zones, fast local variations and ruptures. The estimation can be carried out without knowing the global regularity of the function and, in this sense, it adapts to the unknown regularity.

7.4.6. *Further information on thresholding rules*

There are many variants of these estimators based on wavelet coefficient thresholding (see the references quoted in the introduction and the summary paper of [ANT 01]). They differ primarily by:

– strategy of thresholding: global, level by level, local in time, by blocks;

– type of thresholding: hard, soft and other regularizations. An ingredient used by the basic alternatives is the use of "soft" thresholding instead of "hard" thresholding used previously (see Figure 7.2). The former is defined by: for a real y, $\delta_{hard,\lambda}(y) = y$, if $|y| > \lambda$, and 0, if not. Soft thresholding makes this function continuous by shrinking the kept coefficients. It is defined by: $\delta_{soft,\lambda}(y) = (y - sign(y)\,\lambda)$, if $|y| > \lambda$, and 0, if not. All else being equal, it leads to smoother estimators, less sensitive to the noise;

– choice of the value of thresholds;

– strategy of construction of the estimator:

- it can be direct or Bayesian. In this latter case we formulate hypotheses *a priori* on the distribution of coefficients. Typical modeling consists of considering a mixture of populations reflecting the idea that some coefficients are due to the noise and others to the signal. We deduce the threshold on the basis of the average loss *a posteriori*. It can be shown that soft and hard thresholdings may be interpreted from a Bayesian point of view by suitably selected loss functions. Of course, other thresholding procedures stem from Bayesian strategies,

- it may be based on the search for (less robust) minimax estimators or just minimax with a logarithmic loss,

- it may be adapted to various loss functions and a particular choice of $\mathbb{F}$.

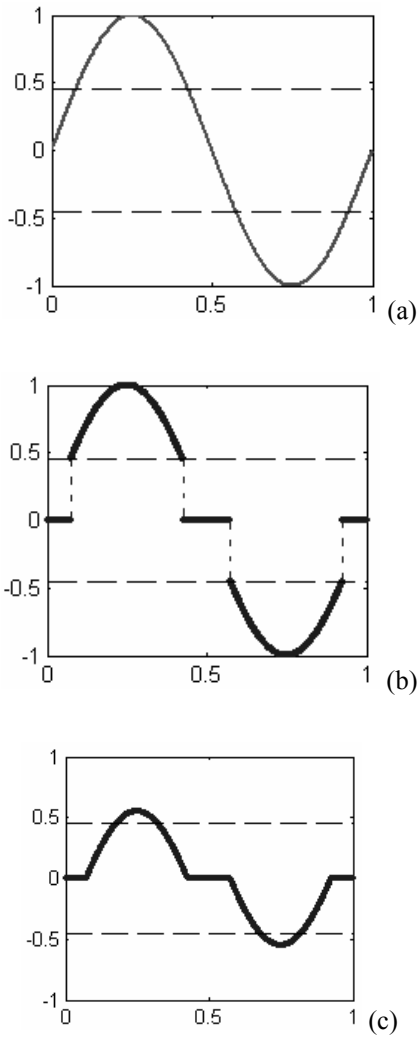


Figure 7.2. *(a) original signal, (b) hard thresholding and (c) soft thresholding*

7.5. Example of denoising with stationary noise

Let us take the simple example of the Doppler signal making it possible to clearly capture the spirit of the technique of denoising by wavelets. In Figure 7.3 we find the signal used for estimation and, in Figure 7.4, the analysis of the signal observed which is a noisy version of the previous one.

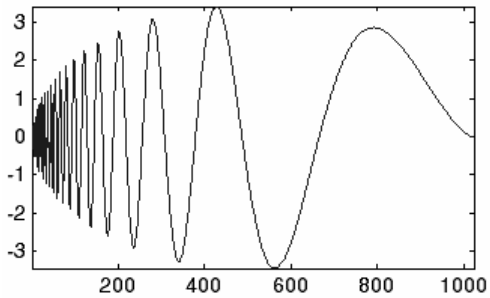


Figure 7.3. *A Doppler signal*

Figure 7.4 is organized in two columns:

– in the first column we see the noisy signal (**s**) and then, below, approximations from level 5 (the coarsest, $\mathbf{a_5}$) to level 1 (the finest, $\mathbf{a_1}$);

– in the second column, at the top we see a colored version of the wavelet coefficients from level 1 to 5 (**cfs**), followed by the noisy signal and then, in the bottom part, the details from level 5 (the coarsest, $\mathbf{d_5}$) to level 1 (the finest, $\mathbf{d_1}$).

The wavelet used is an almost symmetric compactly supported wavelet of order 4, noted *sym4*. The denoising strategy consists of:

– preserving such an approximation that the noise is absent or very attenuated ($\mathbf{a_4}$ or $\mathbf{a_5}$);

– supplementing this approximation by the pieces of finer details clearly ascribable to the useful signal, rejecting the parts regarded as originated by the noise.

It is precisely what the methods of denoising by wavelets do automatically, in particular the universal thresholding, which is the most widespread method.

In the first column of Figure 7.5 we see the wavelet coefficients from level 5 to level 1. To make them more readable they are repeated 2^k times at the level k (which explains the gaps particularly visible for $k > 3$). In each of these graphs we note the presence of two horizontal dotted lines: the coefficients inside the band are zeroed by denoising.

In the second column the noisy signal *s* is superimposed on the denoised signal. Below there are two graphs: a colored version of the wavelet coefficients from level 1 to 5 of the original disturbed signal, followed by the counterpart for the thresholded wavelet coefficients, from which the denoised signal is reconstructed.

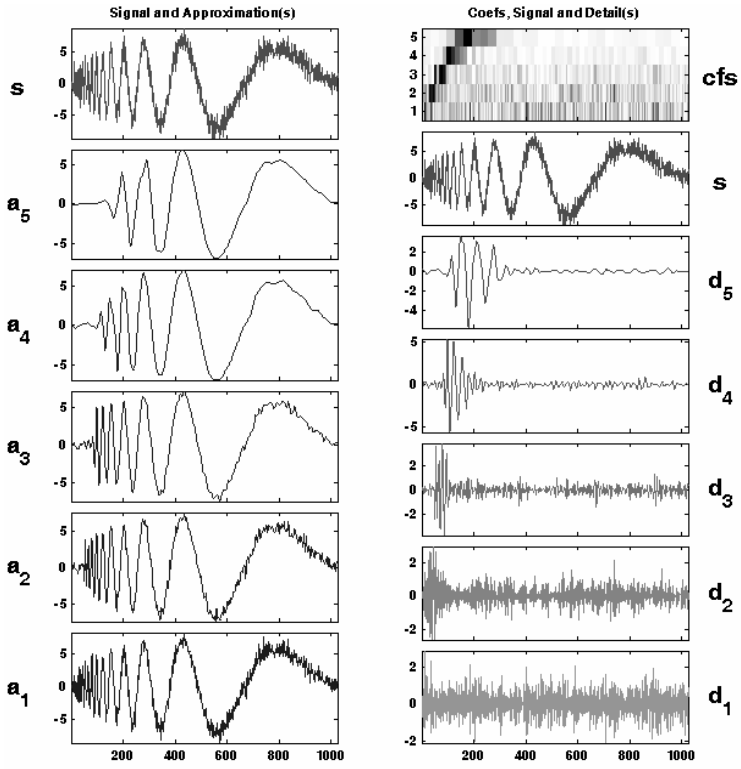


Figure 7.4. *Wavelet analysis of a noisy Doppler*

Let us see through this example how denoising works. The coefficients of approximation are completely preserved, which initially leads to introducing the first component of the denoised signal, which has the form: $\hat{A}_d = \sum_k \hat{\beta}_{5,k}\phi_{5,k}$. The estimated coefficients $\hat{\beta}_{5,k}$ are simply the coefficients obtained by decomposition of the initial signal. The approximation $\hat{A}_d$, which is not represented directly in Figure 7.5, corresponds to the approximation $\mathbf{a_5}$ of Figure 7.4.

Only several larger wavelet coefficients (i.e. detail coefficients) are preserved; the others are replaced by zeros. We may look at the coefficients before and after this operation of thresholding in the second column to note the sparsity of preserved coefficients. We thus obtain the second component of the denoised signal, which has the form $\hat{D}_d = \sum_{1\le j\le 5} \sum_k \hat{\alpha}_{j,k}\,\psi_{j,k}$.

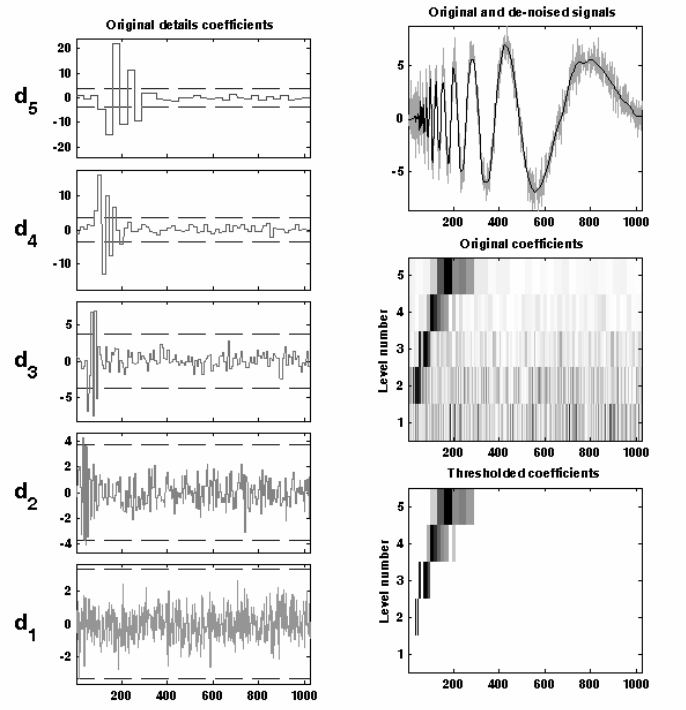


Figure 7.5. *Denoising by wavelets of a Doppler signal*

The estimated coefficients $\hat{\alpha}_{j,k}$ are simply the thresholded version of the detail coefficients obtained by decomposition of the initial signal. The denoised signal is then $\hat{s}_d = \hat{A}_d + \hat{D}_d$.

It is visible that the result obtained at the top on the right has a good quality, except for the very beginning of the signal, which oscillates too much on a small scale with small fluctuations compared to those of the noise.

7.6. Example of denoising with non-stationary noise

The model [7.1] is very simple, even elementary, and the assumptions about the noise are very strong. Fortunately, the "attraction basin" of the denoising methods is much broader. We examine two extensions of this model using examples:

– case of noise presenting multiple change-points of the variance;

– case of a real signal where the noise structure is unknown.

7.6.1. *The model with ruptures of variance*

We always suppose that $Y_t = f(t) + \varepsilon_t$, $t = 1,\cdots,n$, but the standard deviation of ε_t is no longer necessarily constant and may depend on t. More precisely, we allow the existence of time instants $1 = t_0 < ... < t_k < t_{k+1} < ... < t_{K+1} = n$ defining $K+1$ time interval of (t_k, t_{k+1}), inside of which $\sqrt{E\varepsilon_t^2} = \sigma_k$. Obviously, the number of ruptures K, the instants of rupture $\{t_k\}$ and the standard deviations $\{\sigma_k\}$ are unknown.

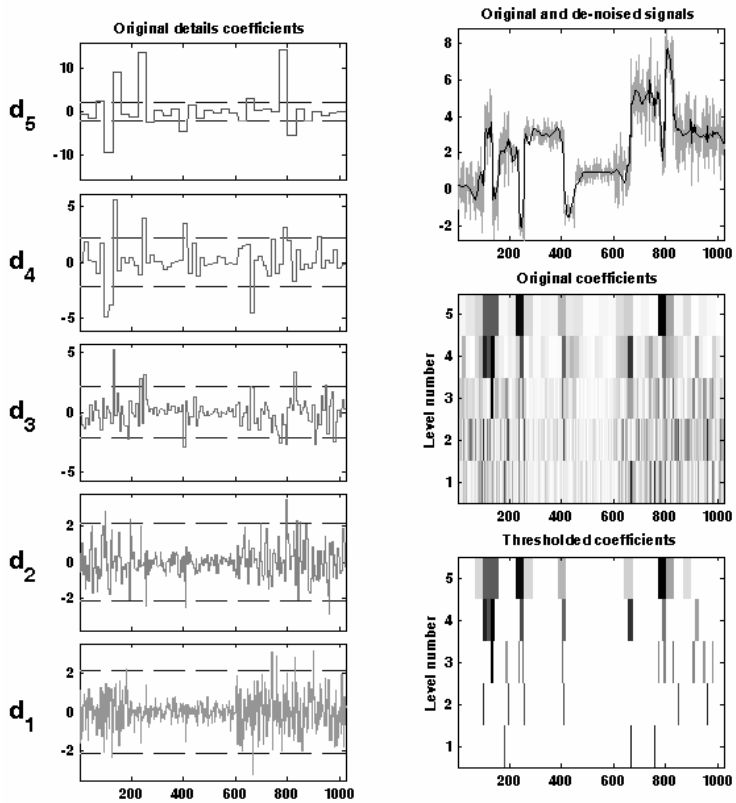


Figure 7.6. *Soft universal thresholding for noisy block signal*

In Figure 7.6 universal thresholding is applied to a signal made up of a succession of blocks disturbed by Gaussian white noise, whose standard deviation

changes at positions 200 and 600. The examination of the denoised signal and the residuals (see Figure 7.7) clearly shows an over-smoothing, in particular, in the interval (200, 600), where the edges of the blocks are badly recovered.

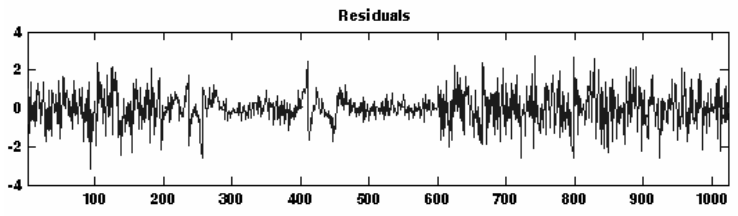


Figure 7.7. *Residuals after soft universal thresholding for noisy blocks signals*

The use of hard thresholding improves the reconstituted edges a little but, basically, the problem persists. It is therefore necessary to seek an adapted strategy making it possible to take into account change-points of the variance and not to settle for the basic procedure resulting in under-smoothing the very noisy parts and over-smoothing those which contain little noise.

7.6.2. *Thresholding adapted to the noise level change-points*

The problem of estimating f with wavelets is tackled by a strategy close to that used in the case of the basic model. However, here the thresholding of the coefficients of Y is carried out interval by interval, after identifying t_k and estimating σ_k needed to calculate the thresholds. In other words, the only important modification is the change detection stage. Here we describe an algorithm implemented by the authors in their wavelets software [MIS 00]. The detection of change-points is performed thanks to an observation already made during the estimation by wavelets of the noise variance: the detail coefficients of level 1 primarily consist of white noise having the same characteristics as ε. It is then enough to apply an algorithm of detection of variance changes to this new signal. We then use the work of Lavielle [LAV 99] which, through dynamic programming, enables us to obtain the best configuration of t_k for any fixed number K of ruptures. Finally, to select the best K we proceed by penalization, i.e. we minimize a criterion which is the sum of two terms: the first of adjustment decreasing with K and the second of penalty growing with K.

In Figure 7.8 we visualize the change instants found through this procedure on the basis of detail coefficients. The clear break points introduced here are clearly

highlighted by the procedure. Besides, they are easy to detect with the naked eye, but the determination of t_k and the thresholds requires using the algorithm.

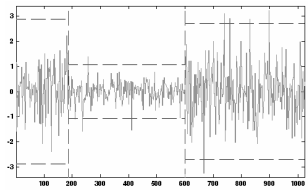


Figure 7.8. *Interval-dependent thresholds according to the changes of variance*

If we apply denoising by interval thresholding, we obtain the result presented in Figure 7.9. This result is much more satisfactory than that obtained by applying global thresholding, as seen, for example, around position 400.

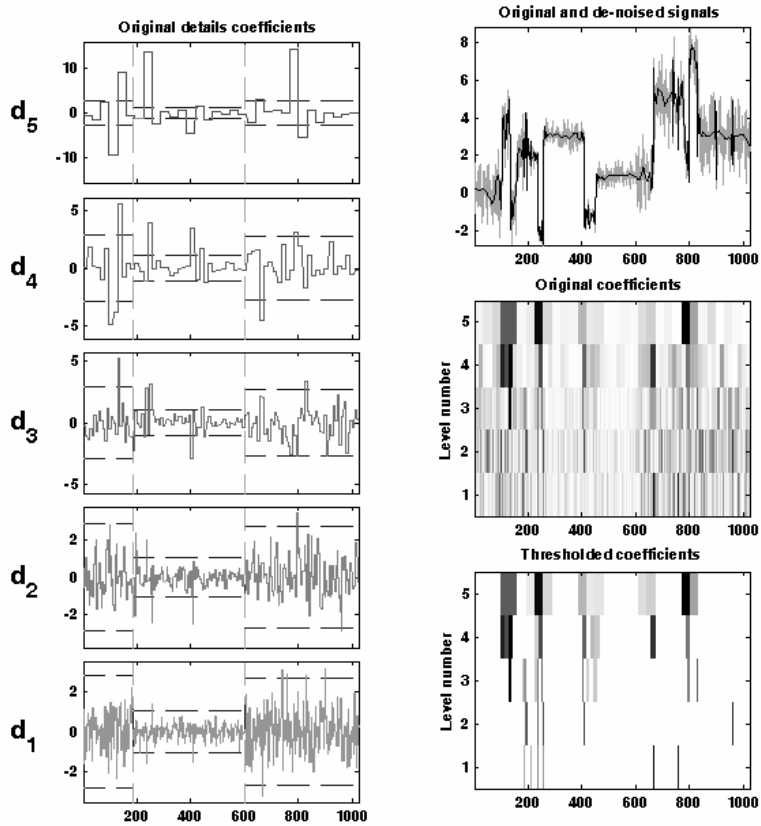


Figure 7.9. *Thresholding adapted to noise variance changes*

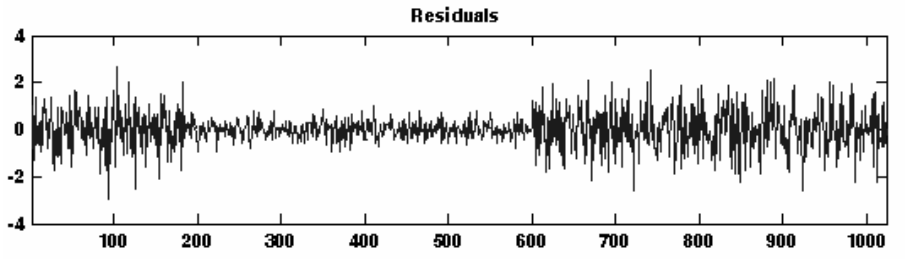


Figure 7.10. *Residuals after thresholding adapted to noise variance changes*

7.7. Example of denoising of a real signal

A second natural extension of the model [7.1] is to make it possible to take into account much less precise noise structure (unknown distribution, time dependence). Two situations are considered:

- noise unknown but homogenous in variance by level of scale;
- noise unknown and non-homogenous in variance at each level.

Examples illustrate the robustness of the denoising algorithms by thresholding and show the capacity of the techniques to be adapted to much more general contexts, in order to deliver results that are undoubtedly not optimal but remain largely acceptable.

7.7.1. *Noise unknown but "homogenous" in variance by level*

It is the denoising by wavelets of an electrical load consumption signal. This problem is difficult since the noise is of an unknown nature but the homogenity of the size of the noise by level is easily distinguished from the detail signals of Figure 7.11.

The method employed here is inspired by the previous situation adapting the thresholding to the level so that we require nothing but the homogenity of the noise by level. Thresholding is carried out as in the basic model, but the coefficients are selected level by level due to a threshold, which does not depend on the global $\hat{\sigma}$, but on $\hat{\sigma}_j$ estimated by level. To estimate σ_j we proceed in the same manner as to estimate σ, but considering the level j detail coefficients.

Whilst examining the first column of Figure 7.11 we see that the method considers all coefficients from levels 1 to 3 as ascribable to noise and they are thus all zeroed.

On the other hand, for levels 4 and 5, the procedure primarily selects wavelet coefficients in the zone around position 400, enabling an excellent restitution of the abrupt signal fluctuations.

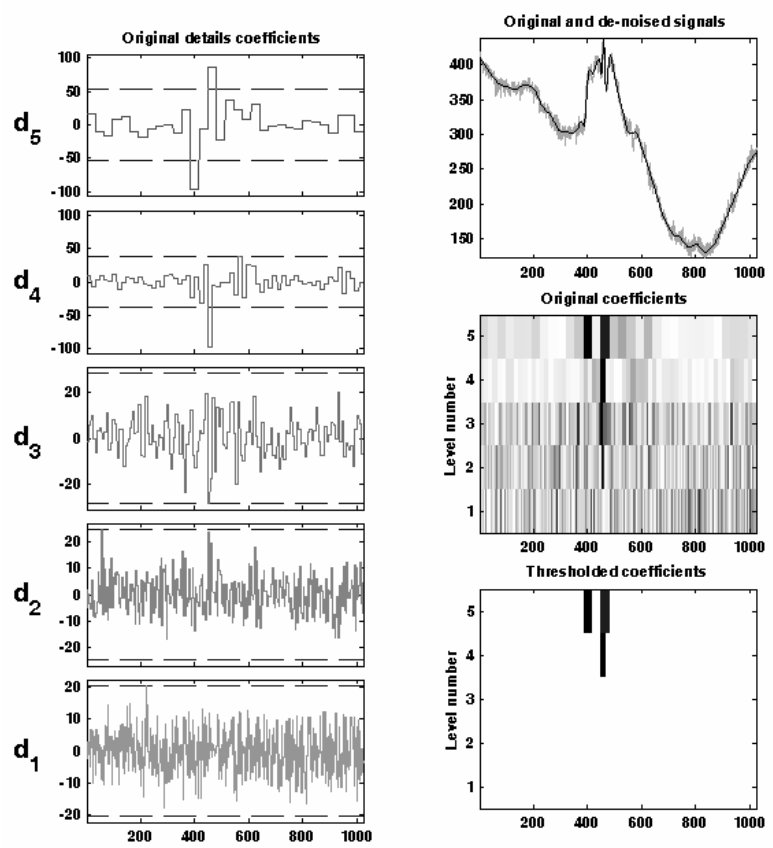


Figure 7.11. *Electrical consumption signal (1) – unknown but homogenous noise*

7.7.2. *Noise unknown and "non-homogenous" in variance by level*

Let us now examine denoising by wavelets of another portion of the same electrical load consumption signal where two difficulties coexist: the noise is of unknown nature and the size of the noise is non-homogenous in time. This is easily seen in Figure 7.12. The strategy to be adopted is, naturally, to apply simultaneously the trick by taking into account the variance changes (seen in the previous section) and the level-dependent thresholding. Here is the very satisfactory result obtained.

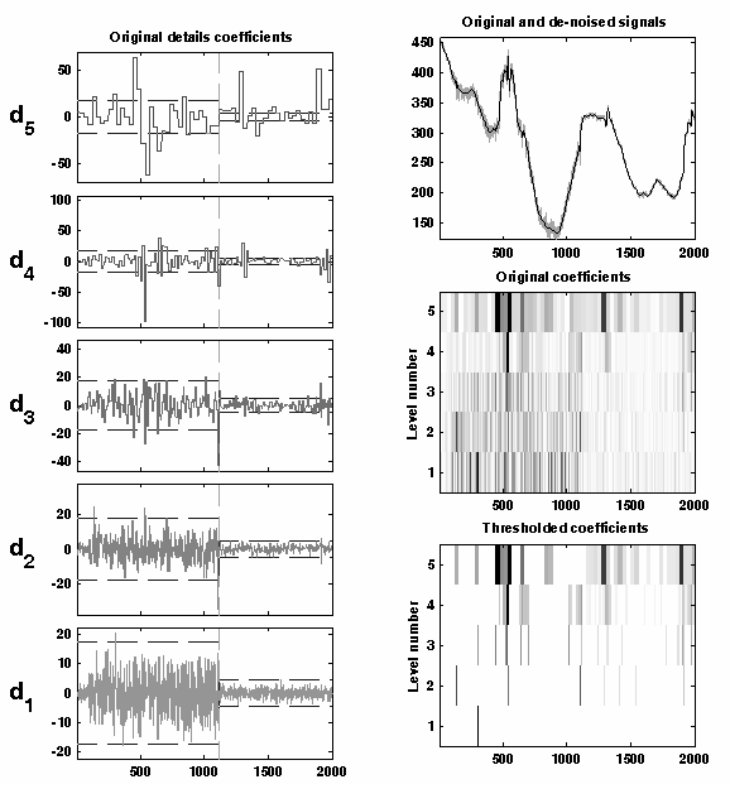


Figure 7.12. *Electrical consumption signal (2) – noise with variance changes*

7.8. Contribution of the translation invariant transform

The discrete wavelet transform (DWT) of a signal of length n is not translation invariant: the analysis of a translated signal is not equal to the translation of the signal analysis. This is true even for a periodized signal (DWT with periodic extension; see Chapter 3). The detection of singularities (alignment, amplitude of the largest coefficient) can depend on the arbitrary choice of the origin.

A manner of restoring the translation invariance is to perform a redundant analysis and then to take averages. For a signal X of length 2^J it consists of calculating not just one decomposition but 2^J different decompositions corresponding to all the circular translations of X. Each of these transformations is invertible but none is translation invariant. On the other hand, if we take the average of the obtained decompositions, we obtain the desired property of invariance.

Moreover, these transforms are easy to obtain; it suffices to reuse the DWT scheme and to remove the decimation phase (for more details see Chapter 3).

The main applications of this transformation are analysis [NAS 95] and, especially, denoising. The idea of translation invariant denoising, introduced by Coifman and Donoho [COI 95], consists of carrying out thresholding on each basis and taking the average of the obtained denoised signals. There are two competing effects present: a better detection of singularities due to taking into account all the shifts in the analysis and an effect of more intense smoothing due to the averaging of the denoised signals on each basis. A good compromise consists of using quite irregular wavelets with a small support and applying hard thresholding. This strategy can improve the restitution of the edges and it is also used for denoising images.

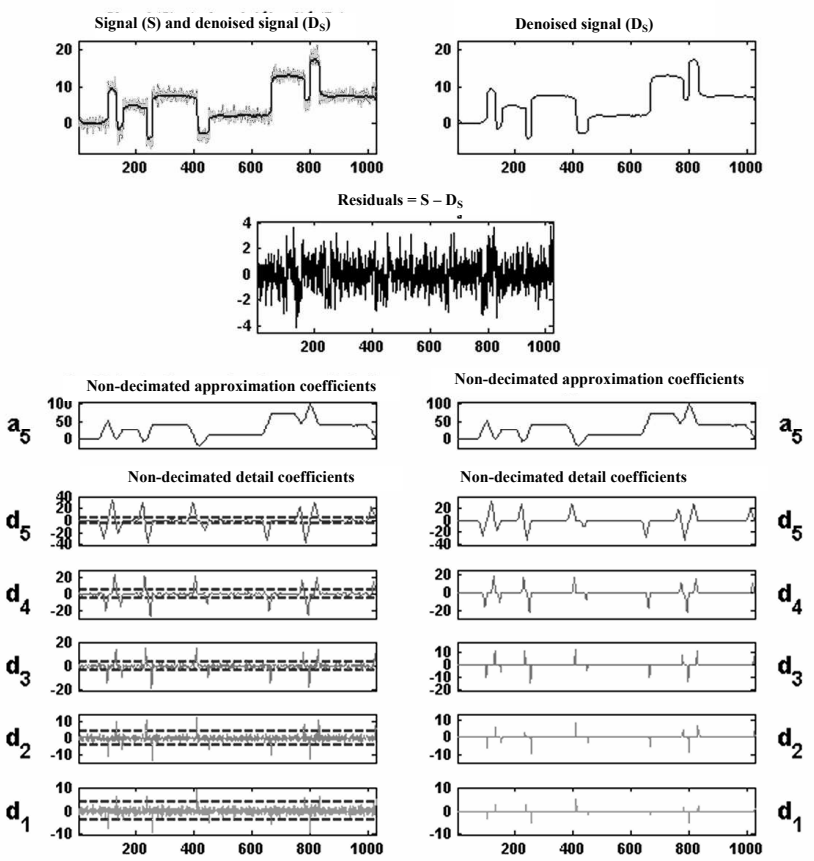


Figure 7.13. *Translation invariant transform: soft thresholding*

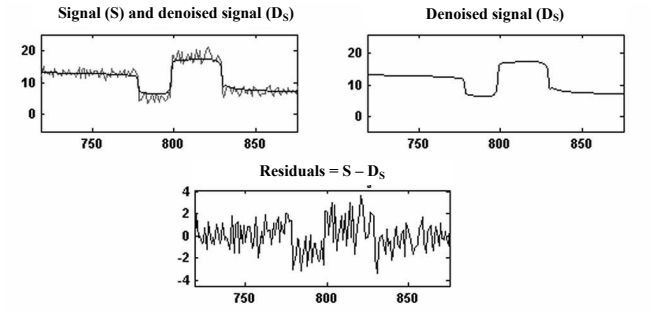


Figure 7.14. *Translation invariant transform: soft thresholding (zoom)*

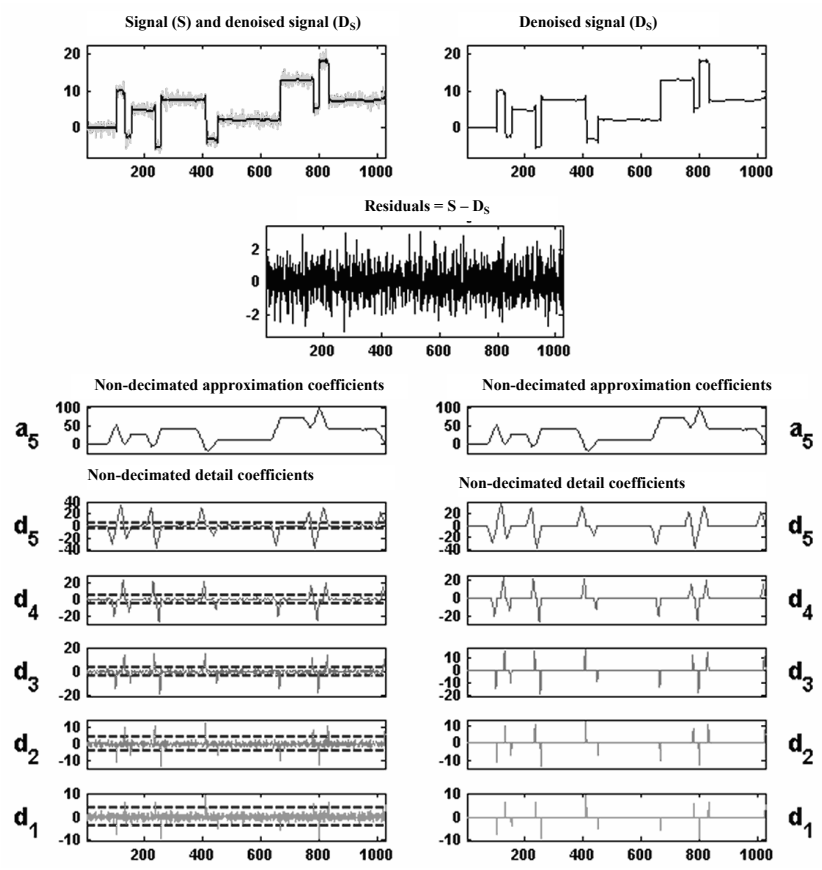


Figure 7.15. *Translation invariant transform: hard thresholding*

Let us illustrate denoising with an example using the translation invariant transform, by applying soft thresholding followed by hard thresholding. We are within the framework of the basic model and the analyzed signal is the superposition of a succession of blocks and a noise. Let us use the Haar wavelet adapted to the shape of the signal. Soft universal thresholding is then applied and we obtain (see Figure 7.13) a denoised signal that is suitable but seems over-smoothed around the signal discontinuities.

We can confirm it by examining more closely the residuals around one of these discontinuities, for example around position 800 (see Figure 7.14). These residuals clearly contain information ascribable to the signal to estimate.

Now let us apply the same denoising strategy and perform hard thresholding. We obtain (see Figure 7.15) a much more satisfactory estimate from the point of view of the estimated function, in particular around discontinuities, as well as from the point of view of the residuals whose compatibility with a Gaussian white noise is very satisfactory.

7.9. Density and regression estimation

In this section, we consider two traditional questions of statistics that are estimating the probability density function and the regression function. We outline a way of constructing estimators using wavelets for these two problems enabling us to use the DWT. The idea is to bring these two situations back to the case of denoising via a suitable preprocessing.

7.9.1. *Density estimation*

We consider a sample $X_1, X_2, ..., X_n$ of n independent and identically distributed random variables of unknown probability density function $h(x)$, supposed to be of finite energy. The goal is to construct an estimator $\hat{h}$ of this density using the data $x_1, x_2, ..., x_n$.

Among the difficulties is the fact that x_i are not equally spaced. We initially present the idea of constructing the estimator, then a "binning" technique that simplifies the estimator by a regrouping of measures in classes whose equidistant centers facilitate calculation.

The function h can be represented on the wavelet basis constructed using the functions φ and ψ in the following way:

$$h = \sum_k a_{J,k}\varphi_{J,k} + \sum_{j=-\infty}^{J}\sum_k d_{j,k}\psi_{j,k} = A_J + \sum_{j=-\infty}^{J} D_j \qquad [7.8]$$

To estimate h it suffices to estimate the coefficients $a_{J,k}$ and $d_{j,k}$. By definition of the coefficients, we have:

$$a_{J,k} = \int \varphi_{J,k}(x)h(x)dx$$

As h is a probability density function, $a_{J,k}$ has a particularly useful interpretation: $\int \varphi_{J,k}(x)h(x)dx = E\left(\varphi_{J,k}(X_i)\right)$ is the expectation of the random variable $\varphi_{J,k}(X_i)$. Such an expectation is estimated quite simply by the empirical mean:

$$\hat{a}_{J,k} = \frac{1}{n}\sum_{i=1}^{n}\varphi_{J,k}(X_i) \qquad [7.9]$$

In the same way we have $d_{j,k} = \int \psi_{j,k}(x)h(x)dx = E\left(\psi_{j,k}(X_i)\right)$ estimated by:

$$\hat{d}_{j,k} = \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(X_i) \qquad [7.10]$$

Using the expression of h given by [7.8] and replacing the coefficients by their estimates given by [7.9] and [7.10], and then by thresholding the estimated detail coefficients we obtain an estimator $\hat{h}$ of the form:

$$\hat{h} = \sum_k \hat{a}_{J,k}\varphi_{j,k} + \sum_{j=J-j_0}^{J}\sum_k \hat{d}_{j,k}\mathbb{1}_{\left\{|\hat{d}_{j,k}|>t\right\}}\psi_{j,k} \qquad [7.11]$$

As in the case of denoising strategies we keep the estimates of the approximation coefficients at the level J and we threshold the detail coefficients estimates. This enables us to avoid the oscillations present when all the $\hat{d}_{j,k}$ are kept.

However, the x_i are not equally spaced and it is impossible to calculate this estimator by using the DWT fast algorithm. We may propose an approximation, based on a regrouping in classes, making it possible to benefit from the DWT's speed of calculation.

We introduce a histogram $\widehat{H}$ of the values of X in m bins of the same width. The bin marks are in a vector denoted X^b and we associate with each bin the corresponding number of observations gathered in a vector denoted Y^b. We then define $\widehat{H}(x) = \dfrac{Y^b(r)}{n}$ for x belonging to the bin r. By using $\widehat{H}$ we may then write:

$$\hat{d}_{j,k} = \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(X_i\right) \approx \frac{1}{n}\sum_{r=1}^{m} Y^b(r)\,\psi_{j,k}\left(X^b(r)\right) \approx c\int \psi_{j,k}(x)\,\widehat{H}(x)\,dx$$

where c is the binwidth.

The signs ≈ express the idea that we lose information when the histogram is used instead of the X_i values and when the integral is approximated. The last ≈ sign is interesting: it means that $\hat{d}_{j,k}$ is, up to a constant, the wavelet coefficient of the function $\widehat{H}$ associated with the level j and the position k. A similar result holds for $\hat{a}_{J,k}$.

Calculations to be performed on the data transformed in this manner then acquire a form, which makes it possible to use the DWT by applying it to the "binned" data seen as a signal.

7.9.2. *Regression estimation*

The problem of estimating the regression function is solved in a similar way. The essential differences, obviously, relate to the model:

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1,\cdots,n$$

where f is the unknown regression function, the $(X_i, Y_i)_{1\le i\le n}$ are observable and $(\varepsilon_i)_{1\le i\le n}$ is an unobservable white noise, centered and of variance σ^2. We have to estimate the function f merely on the basis of the observations $(x_i, y_i)_{1\le i\le n}$.

The fundamental difference with the case of denoising is that here the X_i are no longer deterministic, but are randomly generated according to the unknown probability density function h.

We consider $(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)$ of n independent random pairs sampled from the common distribution h. The function $g = f.h$ is introduced, from where we get $f = \frac{g}{h}$ with the convention that $\frac{0}{0} = 0$.

As previously, we will carry out a "binning" of the values of X into m bins. Let n_r be the number of observations belonging to the r^{th} bin and let $X^b(r)$ be its center. We then define $Y^b(r) = \frac{1}{n_r} \sum_{\{j|X_j \in bin(r)\}} Y_j$.

Let us seek an estimator of f of the form $\hat{f} = \frac{\hat{g}}{\hat{h}}$, where $\hat{h}$ is the estimator provided by a suitably normalized histogram of X. We then use the same technique as that used for density estimation: the coefficients of g are estimated by:

$$\hat{a}_{J,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i \varphi_{J,k}(X_i) \text{ and } \hat{d}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i \psi_{j,k}(X_i)$$

We obtain approximations of these estimates by the wavelet coefficients of the sequence Y^b:

$$\hat{a}_{J,k} \approx \frac{1}{n}\sum_{r=1}^{m} Y^b(r)\varphi_{J,k}\left(X^b(r)\right) \text{ and } \hat{d}_{j,k} \approx \frac{1}{n}\sum_{r=1}^{m} Y^b(r)\psi_{j,k}\left(X^b(r)\right)$$

which in this situation also allows us to use the DWT.

Of course, the quality of the obtained estimators is not completely controlled since it results from the estimation by wavelets, but applied to data transformed by a kind of preliminary estimate (see [HAR 98] pp. 134 and 183).

7.10. Principle of compression by wavelets

7.10.1. *The problem*

The problem of compression is often reduced to images, whose applications are particularly interesting (see Chapter 8). For example, wavelets are used in the new image compression standard JPEG 2000 [JPE 00].

Nonetheless, the compression of one-dimensional signals is also relevant. Thus, the legal obligation for certain public service operators to preserve their customers' consumption curves over long periods necessarily implies their compression. Similarly, it is crucial to compress audio signals for fast transmissions via the international network.

Wavelets, with other techniques, contribute to provide effective solutions to this problem. Of course, the complete compression process includes phases of quantization, coding and decoding in addition to wavelet processing itself. Should we concentrate on this, the objective then consists of obtaining the sparse representation of the signal being compressed. Here we tackle the problem of compression by wavelets only from this viewpoint.

7.10.2. *The basic algorithm*

The basic algorithm is very similar to that of denoising and proceeds in three steps:

– decomposition;

– selection or thresholding of coefficients (by means of hard thresholding);

– reconstruction.

Starting from the signal being compressed we first decompose it on an orthogonal wavelet basis using the discrete transform. Then, we select part of the coefficients by hard thresholding, preserving intact the coefficients of approximation of a suitably selected level. Lastly, starting from the thresholded coefficients, we reconstruct a signal applying to them the inverse discrete transform. The obtained signal is the compressed signal.

7.10.3. *Why does it work?*

The key argument for the understanding of the effectiveness of these methods, resides in the capacity of the wavelets bases, to represent broad classes of functions in a very parsimonious way.

7.11. Compression methods

7.11.1. *Thresholding of the coefficients*

The simplest framework to introduce the heart of compression by wavelets is to take a function f with finite energy and considering an orthonormal wavelet basis of $L^2(\mathbb{R})$ denoted simply by $\{g_m\}_{m\in\mathbb{N}}$.

The representation of f in this basis is $f = \sum_{m\in\mathbb{N}} \langle f, g_m \rangle g_m$. The coefficients of f are $\{\langle f, g_m \rangle\}_{m\in\mathbb{N}}$; we will denote them by $\alpha_m = \langle f, g_m \rangle$.

Performing a compression consists of selecting a fixed number M of coefficients. Thus, in general, such a compression is carried out with loss. Let us denote by *I* the set of indices of the M selected coefficients and $f_M = \sum_{m\in I} \alpha_m g_m$ the approximation obtained.

If we agree to assess the quality of compression by using the quadratic error, we have: $\|f - f_M\|^2 = \sum_{m\notin I} \alpha_m^2$. Since $\|f\|^2 = \sum_{m\in\mathbb{N}} \alpha_m^2$, it is clear that the best choice of I consists of preserving the M largest coefficients, in absolute value. By

denoting $\left\{\alpha_{m_p}\right\}_{p\in\mathbb{N}}$ the coefficients sorted in descending order, the optimal quadratic error is then $\left\|f - f_M\right\|^2 = \sum_{p\geq M} \alpha_{m_p}^2$.

We can construct the "compressibility" curve on a Doppler signal: on the x-axis we find the number M of preserved coefficients and on the y-axis is the ratio of preserved energy $\left\|f_M\right\|^2 / \|f\|^2$, or $1 - (\left\|f - f_M\right\|^2 / \|f\|^2)$ if we express it according to the associated relative quadratic error. Two curves are represented in Figure 7.16: one for the initial signal, the other for the coefficients of this signal analyzed with the wavelet *db1*. For the signal and the coefficients, we preserve the M largest value (in absolute value).

We can say that the signal energy concentrates in a far fewer coefficients for the wavelet representation than for the usual signal representation: the curve of coefficients grows very quickly towards 1 and is always clearly above that of the signal.

The signal has been decomposed at level 8 using the wavelet *db1*. The largest coefficient of decomposition, which is an approximation coefficient of level 8, contains 23.97% of the energy. The greatest signal coefficient, in turn, holds merely 0.28% of the energy. These two extreme coefficients are represented by two dots in Figure 7.16.

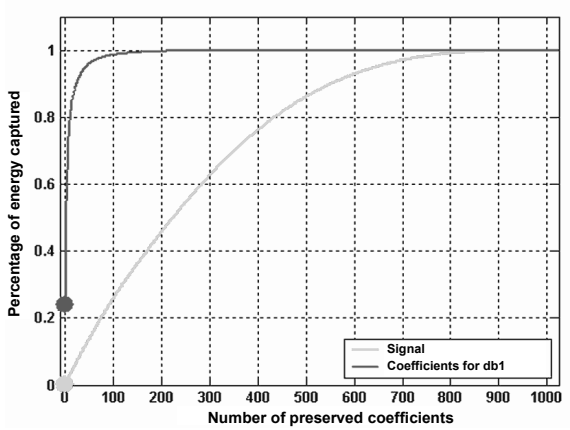


Figure 7.16. *Compressibility curves*

With others wavelets the compressibility curve primarily has the same aspect. For example, in Figure 7.17, we have represented the curves of compressibility of the signal and the decompositions by wavelets *db1* and *sym6* with a logarithmic scale for the number of coefficients. The more vanishing moments the wavelet has, the "better" the capacity of compression of the basis.

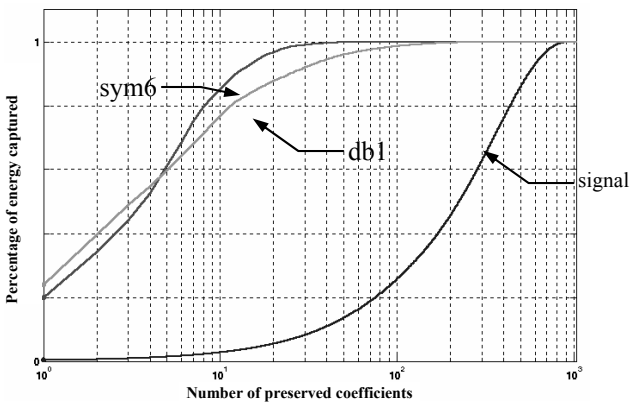


Figure 7.17. *Compressibility curves for the signal, and wavelets db1 and sym6*

7.11.2. *Selection of coefficients*

Other strategies can be used in this context, when the norm chosen to measure compression quality is not the $\| \bullet \|_2$ norm used above. This can be irrelevant for perceptual reasons.

In [BIR 97], Birgé and Massart propose a strategy enabling greater flexibility. They establish a result of the optimal speed convergence for functions f belonging to Besov spaces. This strategy results in selecting large coefficients at each level. More precisely, the approximation coefficients at the level J are kept and for each finer level $j \leq J$, only the n_j largest detail coefficients, by absolute value, are kept. The numbers n_j grow with j as $M/(J+1-j)^{\alpha}$ where M is a positive constant and $\alpha > 1$ is a parameter. The method thus consists of preserving a rough approximation and selecting few coefficients for the levels of fine details, which contain high frequency signal fluctuations and more and more coefficients so for the coarsest levels.

7.12. Examples of compression

Let us illustrate the methods of compression by wavelets of one-dimensional signals, starting by considering a synthetic signal compressed via a global thresholding.

7.12.1. *Global thresholding*

Figure 7.18 is organized into two columns.

In the first column we find only one plot making it possible to adjust the global threshold for the coefficients of the signal to be compressed: all the coefficients which are, in absolute value, lower than the threshold (equal to 8.049 and located by the vertical dotted lines) are zeroed by the compression procedure. The possible values of the threshold are carried by the x-axis. Two curves are drawn: one increasing, giving the percentage of zeros in the representation of the compressed signal, the other decreasing, giving the percentage of the energy preserved by the compressed signal.

In the second column, in the top graph, the signal is superimposed on the compressed signal. At the bottom we find a colored version of the wavelet coefficients (obtained using *sym4*) from levels 1 to 5 of the original signal and in the bottom graph, the counterpart for the thresholded wavelet coefficients, from which the compressed signal is reconstructed.

The percentage of zeros in the wavelet representation is 94.98% for 92.21% of energy preserved. This stems from the fact that the signal is noisy and, consequently, that the very fast fluctuations are uninteresting and therefore removed. The graphs of the coefficients before and after thresholding make it possible to note the sparsity of the preserved coefficients.

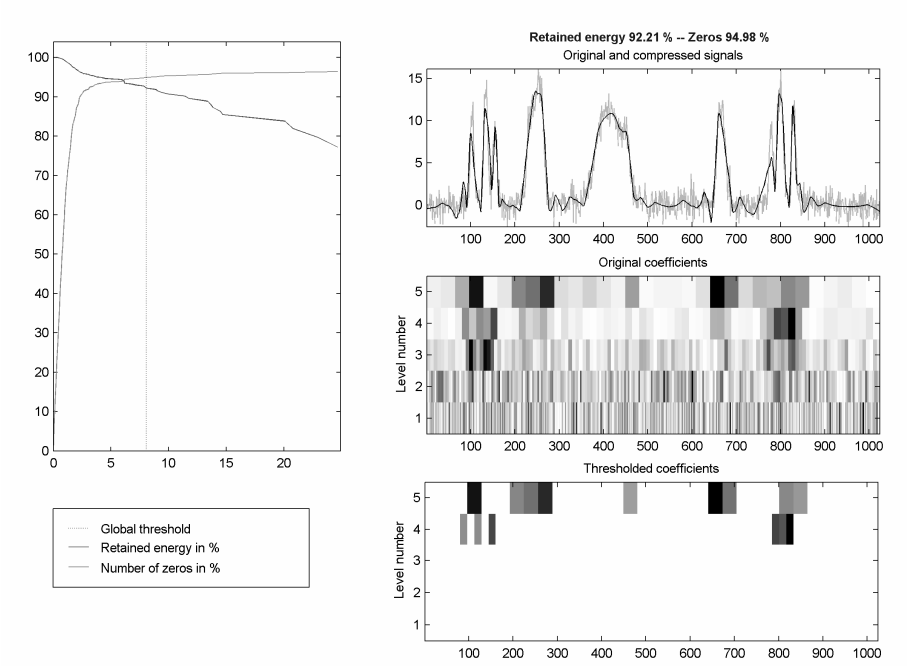


Figure 7.18. *Compression (global thresholding)*

The resemblance of the results produced by denoising and compression is obvious. This is not surprising; it is natural that a good method of compression with loss leads to eliminate the noise. Indeed, if we compress a noisy signal of the type defined by [7.1] by preserving the largest coefficients, we will denoise it, at least partially, since noise coefficients are generally all "small"; this is what occurs in this example. The difference is that the threshold chosen for compression comes from external constraint: frequency bandwidth, available memory, prescribed compression ratios. This is why denoising is not perfect here and fluctuations absent from the original signal f appear in the compressed signal. On the contrary, in the case of denoising the technique selects the threshold automatically by estimating the noise level using the information provided by the model.

7.12.2. *A comparison of the two compression strategies*

The two strategies described above often yield similar results for reasonable choices of the parameters. Let us examine, for example, (see the graph on the right

in Figure 7.19) a signal made up of a portion of a sinusoid to which we have added a jump at position 310, a second jump at position 730 and a white noise. Let us compare the method of selection and the associated global thresholding.

First of all, let us apply the strategy of selection of the largest coefficients by level. We then obtain Figure 7.19.

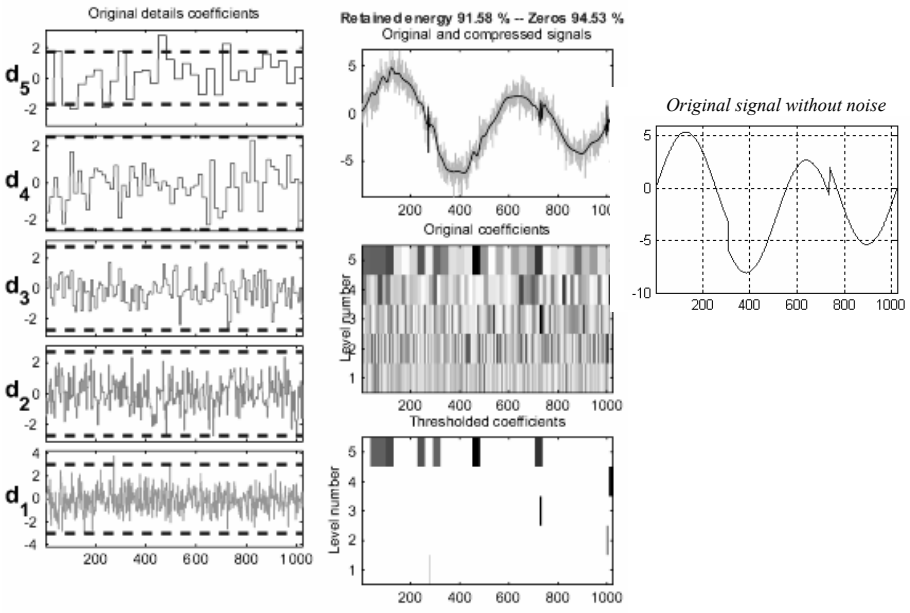


Figure 7.19. *Compression (selection by level)*

This strategy leads to saving approximately 5% of the coefficients and makes it possible to capture 92% of energy as indicated above the right-hand column. The result obtained is satisfactory since the jump and the discontinuity are preserved, the noise is eliminated except around three positions.

If one applies the corresponding strategy of global thresholding, i.e. leading to the same percentage of zeros in the representation after thresholding, we obtain the result displayed in Figure 7.20.

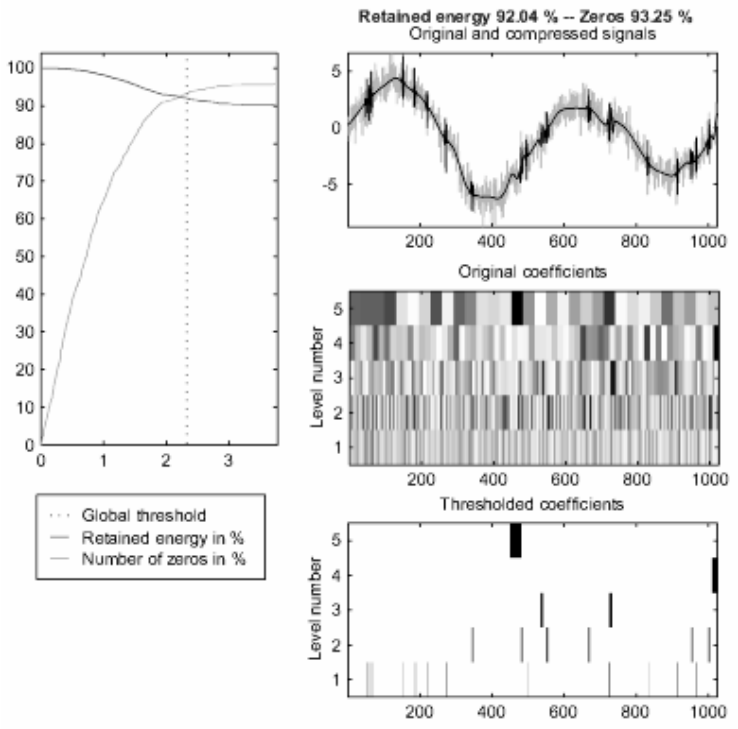


Figure 7.20. *Compression (global thresholding)*

The result is naturally better from the point of view of the percentage of preserved energy, since for this criterion and for a fixed percentage of zeros it is the best possible selection. Let us note, however, that the gain is relatively small: we pass from 91.67% to 92.03%. On the other hand, more coefficients linked to the noise are preserved, so the signal compressed by this method is of less satisfactory aspect than the previous one.

This toy example clearly illustrates the difference between the two strategies. To this end it is enough to compare the graph of the preserved coefficients at the bottom on the right of the two figures: global thresholding tends to preserve coefficients at small scales corresponding to levels 1, 2 or 3, whereas the selection strategy tends towards large scales (levels 4 or 5).

7.13. Denoising and compression by wavelet packets

We can easily extend the algorithms of denoising and compression to the wavelet packet bases. This extension is especially useful in compression; for a fixed orthogonal wavelet it exploits the increased capacity of compressibility of the wavelet packet bases compared to the corresponding wavelet bases.

Let us illustrate this point by examining the example of a noisy chirp. For the wavelet *sym8* and level 6 decomposition, let us fix the threshold so that the percentage of preserved energy is 90%. The effect of this global thresholding applied to the wavelet decomposition is provided in Figure 7.21. This decomposition is performed using the periodic signal extension, in order to be able to directly compare the percentages of zeros in the compressed signals.

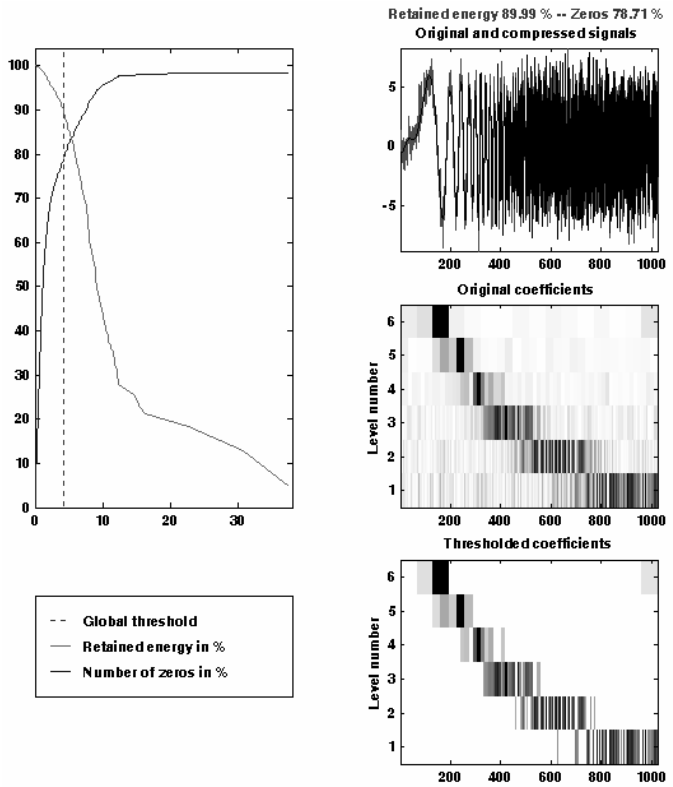


Figure 7.21. *Compression by wavelets – sym8, level 6, global thresholding*

The percentage of zeros obtained is 78.71%. Let us then apply the threshold leading as previously to 90% of energy preserved in the wavelet packet decomposition of the signal. The percentage of zeros obtained is now shifted to 88.77%, that is a gain of 10%.

7.14. Bibliographical comments

Let us conclude this chapter with some historical and bibliographical observations (see Chapter 12 of [MEY 93]). The first papers appeared at the beginning of the 1990s and powerful ideas emerged very early, for example, in the work of Donoho, Johnstone, Kerkyacharian, Picard, Doukhan and Antoniadis for denoising and the work of Coifman, Cohen and De Vore for compression. These ideas were then developed by many authors. In order to grasp the extent of the stimulation exerted by these ideas in the statistics and signal community, see [ANT 95].

We find typical results in the rather numerous books dedicated, especially since 1998, to the question of the use of wavelets in statistics. Ogden [OGD 97] offers a basic presentation thereof. The books by Vidakovic [VID 99] and Percival and Walden [PER 00] constitute at present the two most complete references in the field. In the extremely rich book by Mallat [MAL 98], two chapters tackle these questions from a point of view of signal processing. Several points of our book take this point of view as a starting point. For a more technical vision see Härdle *et al.* [HAR 98]. Moreover, see the papers on synthesis by Antoniadis [ANT 97], Antoniadis *et al.* [ANT 01] and Abramovic *et al.* [ABRA 00].

Chapter 8

Image Processing with Wavelets

8.1. Introduction

The applications of wavelets in image processing are often synonymous with compression. Today, the need for storing large quantities of information and for fast transmission through international communication networks are key issues. This is why two examples of wavelet used in this field have had a particularly pronounced impact:

– for fingerprint storage, the FBI finally chose an algorithm containing wavelets [BRI 95];

– more recently, the compression standard JPEG 2000 [JPE 00] has also been built around algorithms containing wavelets.

These successes had a resounding echo and contributed to the popularization of wavelets. For this reason, the last part of this chapter is devoted to what remains the main 2D-wavelet application: compression.

Before focusing on this point this chapter gives a rapid presentation of the theoretical framework and considers two of the more marginal but still efficient 2D-wavelet applications. We introduce the concepts of decomposition, approximation and detail for an image. They then play a crucial role in:

– the edge detection. In this part we illustrate the capacity of wavelets to locally analyze the fluctuations of image grayscale levels. The processed examples show that, almost without processing, the analysis of images by wavelets makes it possible to extract a new image, from which we can isolate the edges;

– the fusion of images[1]. Here we crucially use the local character of wavelet coefficients in order to mix two decompositions stemming from different images. We first present the problem and the technique used with a simple example. The second processing, which is more realistic, fuses two fuzzy images, while the last, less serious processing, amalgamates a mask of the Japanese Nô theater with the bronze bust of a Roman emperor.

We then use two examples to tackle the techniques of wavelet image denoising, which prove to be powerful and easy to use. The problem and the procedure are presented in an example of an artificially noisy image before considering a real noisy image.

Finally, in the last section we present the principles of compression by wavelets and then concentrate on aspects specifically related to wavelets before tackling the complete compression chain with a particular interest in quantization and coding. This point plays a very important part and has a place in this book for the following reason. In the preceding version of the JPEG standard (1992), the discrete cosine transform (DCT) was employed to transform images before coding. The first tests concerning the use of wavelets in compression, although promising, did not yield results clearly superior to those of the DCT. The creation and implementation of the coding algorithm EZW (*embedded zerotree wavelet encoding*; see [SHA 93]) led to the development of the JPEG 2000 standard. This algorithm fundamentally exploits the arborescent nature of the multi-resolution analyses; it is presented at the end of this chapter and in an appendix at the end of the book.

8.2. Wavelets for the image

Multi-resolution analyses of $L^2(\mathbb{R})$ were defined in Chapter 2 as a family of decreasing subspaces having various properties related to approximation, dilation and translation. Similarly, we can introduce multi-resolution analyses of $L^2\left(\mathbb{R}^2\right)$ (see [DAU 92] Chapter 10, p. 313-315 and [MAD 92]). However, 2D multi-resolution analyses usually come from constructions using tensor products. Hereafter we will restrict ourselves to this case.

Starting with a 1D multi-resolution analysis we note the associated approximation and detail spaces as V_j^{1D} and W_j^{1D}. For every level j the approximation space of the 2D multi-resolution analysis is obtained as a sum of four 1D tensor products:

1 Our presentation is based upon the ideas used by P.M. Zeeuw in [ZEE 98].

$$V_{j-1}^{2D} = \overline{\left(V_j^{1D} \otimes V_j^{1D}\right)} \oplus \overline{\left(V_j^{1D} \otimes W_j^{1D}\right)} \oplus \overline{\left(W_j^{1D} \otimes V_j^{1D}\right)} \oplus \overline{\left(W_j^{1D} \otimes W_j^{1D}\right)}$$

This relation is also written:

$$V_{j-1}^{2D} = V_j^{2D} \oplus \left[W_j^{2D}\right]_h \oplus \left[W_j^{2D}\right]_v \oplus \left[W_j^{2D}\right]_d$$

Indeed, $V_{j-1}^{2D} = V_{j-1}^{1D} \otimes V_{j-1}^{1D}$ with $V_{j-1}^{1D} = V_j^{1D} \oplus W_j^{1D}$, which makes it possible to obtain the first expression for V_{J-1}^{2D}.

Whatever the method of 2D construction, whether using tensor products or not, we have a scaling function, as in 1D, and three wavelets instead of one. In the particular case of tensor construction, if φ and ψ indicate respectively the scaling function and the 1D wavelet, we have:

– the scaling function: $\varphi^{2D}(x,y) = \varphi(x)\varphi(y)$;

– three wavelets:

$$\psi_1^{2D}(x,y) = \varphi(x)\psi(y), \quad \psi_2^{2D}(x,y) = \psi(x)\varphi(y), \quad \psi_3^{2D}(x,y) = \psi(x)\psi(y)$$

The algorithms of decomposition and reconstruction of the 2D-DWT, with preceding tensor construction, are described in Chapter 3.

8.2.1. *2D wavelet decomposition*

As in 1D, two types of objects are handled in 2D:

– *approximation or detail coefficients* corresponding to coordinates in the bases of spaces V_j^{2D} and W_j^{2D};

– *reconstructed approximations and details* corresponding to projections on spaces V_j^{2D} and W_j^{2D}.

Here we are dealing with monochromatic images, i.e. in grayscale. An integer is associated with each pixel of the image, which is an index in an ordered table of "colors". We speak then of indexed images. Within this framework, an image is then a matrix of integers. There are, of course, true color images, often associated with three matrices, for example the RGB coding. We will not cover here the processing of such images (for additional information see [SKO 01]).

For a 2D signal noted X, we affect the values to the coordinates in V_0^{2D}. We break X up into a sum of orthogonal signals corresponding to different visualization or resolution scales. Thus, we have:

$$X = A_1 + D_1 = \cdots = A_j + D_j + \cdots + D_2 + D_1$$

Decomposition along three directions of detail spaces implies that in 2D:

$$A_{j-1} = A_j + D_j = A_j + \left[\left(D_h \right)_j + \left(D_v \right)_j + \left(D_d \right)_j \right]$$

where D_h, D_v and D_d indicate respectively what is usually called horizontal, vertical and diagonal details.

An important point distinguishes the 1D from the 2D, it is the notion of "vision" attached to the 2D. In general, we process an image representing more than one simple function with two variables or a matrix because it implicitly contains the concept of "visual rendering". We could also draw parallels with the 1D if, for example, we attached the concept of "audio rendering" to a signal.

8.2.2. *Approximation and detail coefficients*

8.2.2.1. *Horizontal, vertical and diagonal details*

Let us take an image extracted from an A. Dürer painting, for which we carry out a level 1 analysis, followed by level 2 analysis using the Haar wavelet.

The initial image of size (256 × 256) presents marked geometric aspects. In particular, vertical and horizontal lines clearly stand out in the analyses.

In Figure 8.2 on the left, the coefficients of approximation and the coefficients of detail of level 1, are represented as follows:

– *A*: approximation coefficients;

– *H, V, D*: coefficients of horizontal, vertical and diagonal details respectively.

The size of the "small images" respects the initial image proportions and the number of coefficients: (256 × 256) = 4 × (128 × 128). The intensity of color is proportional to the absolute value of coefficients, from the smallest in dark to the largest in light. Each small image of coefficients is colored independently of the others.

At level 2, the level 1 detail coefficients are preserved and the level 1 approximation coefficients are decomposed.

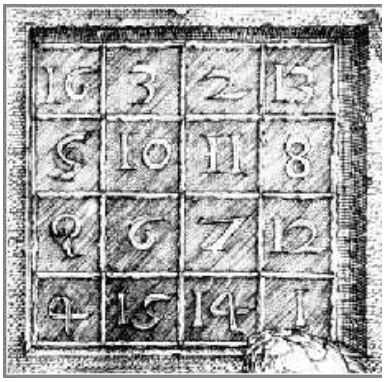
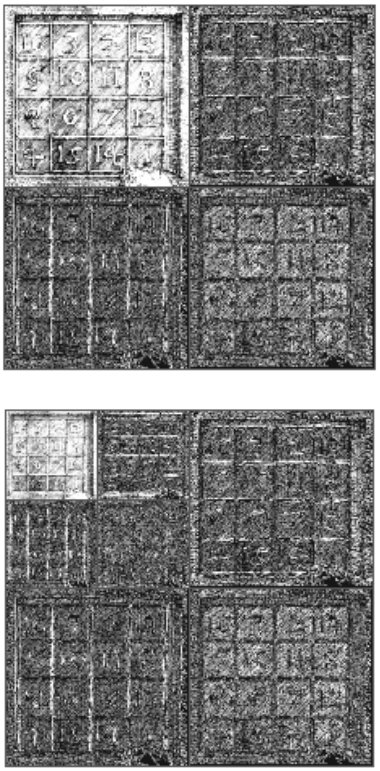
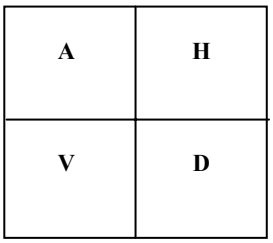


Figure 8.1. *Magic square*





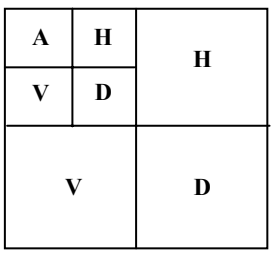


Figure 8.2. *Magic square decomposition at level 1 (top) and at level 2 (bottom). The two graphs on the right show the organization of the coefficients*

8.2.2.2. *Two representations of decomposition*

We commonly use two types of representation for 2D wavelet decomposition. The first, already presented, highlights the proportions between the various components. The second representation highlights the arborescent aspect of wavelet decomposition.

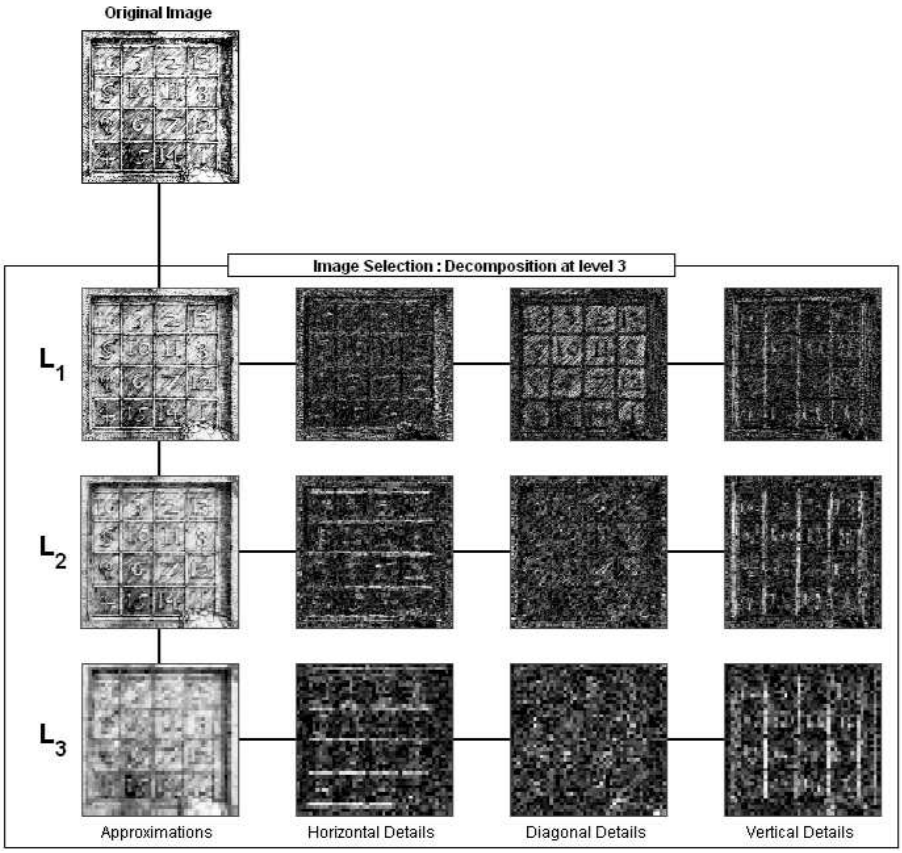


Figure 8.3. *Level 3 decomposition of the magic square*

Let us consider level 3 decomposition of the previous image, represented in Figure 8.3. For each level we keep detail coefficients and decompose the approximation coefficients. Each line L_j, for $j = 1,2,3$, presents the level j approximation coefficients followed by the coefficients of the three details for the same level. At the line j, each coefficient matrix is a 4^j time smaller than the original image. The above figure does not respect this scale difference.

The coefficients obtained using the Haar wavelet make it possible to suitably distinguish the principal geometric aspects of the analyzed image. The use of another wavelet would make the direct reading of the coefficients more difficult, even impossible. To mitigate this drawback, the strategy that has to be adopted in order to interpret the results of the analysis is the same as in 1D: it is enough to consider the associated reconstructed signals (see section 8.2.1). This is what we will do now.

8.2.3. *Approximations and details*

All the images in this section have the same size as the original image and the same resolution.

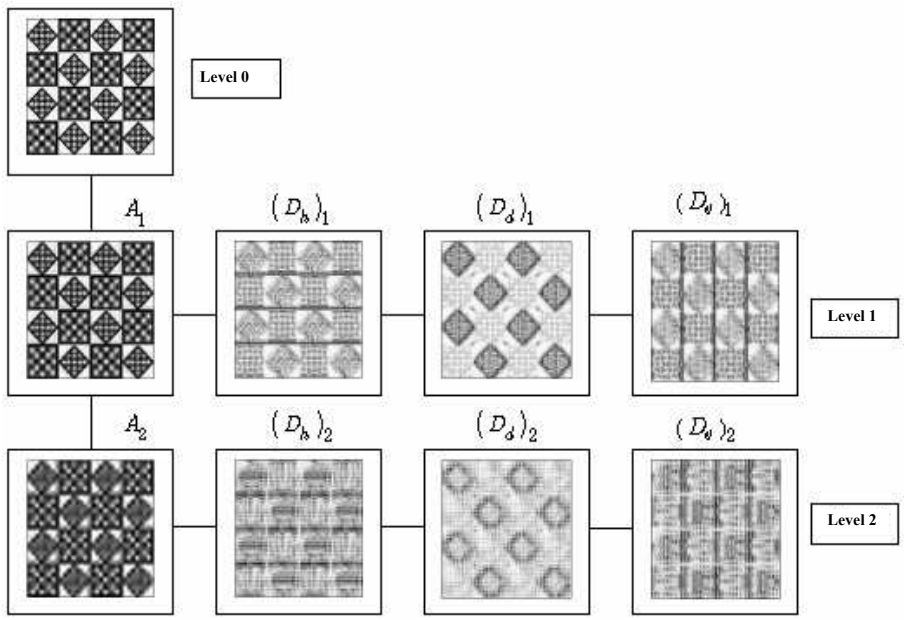


Figure 8.4. *Approximations and details by direction, of the "tartan" image analyzed at level 2 with the wavelet sym4*

Let us consider the analysis of an artificial image by the wavelet *sym4*. In the first column of Figure 8.4 we can see, from top to bottom, the original image and the level 1 and 2 approximations. The three following columns contain horizontal, diagonal and vertical details respectively.

The geometric aspects come out clearly, despite the use of a wavelet different to the Haar wavelet.

Having three details at each level is inherited from the use of tensor 1D wavelets. In the two examined theoretical cases this makes it possible to highlight the edges in each direction. It is also interesting to have a global vision of the detail, thus, making it possible to find the following interpretation: an approximation is the sum of a rougher approximation and a touching up.

To this end, let us group the details corresponding to the various orientations for each level by exploiting the equality:

$$A_{j-1} = A_j + \left[\left(D_h \right)_j + \left(D_v \right)_j + \left(D_d \right)_j \right] = A_j + D_j$$

In Figure 8.5 we find four rows. The first row shows the original image assimilated to A_0. The two following rows illustrate the equalities:

$$X = A_1 + D_1 \text{ (2}^{\text{nd}} \text{ line)}$$

$$X = A_2 + D_2 + D_1 \text{ (3}^{\text{rd}} \text{ line)}$$

This last relation breaks X up into a level 2 approximation and two final improvements capturing the differences between two successive approximations: $D_1 = A_0 - A_1$ and $D_2 = A_1 - A_2$.

Finally, in the last row we group the details D_1 and D_2. Thus, the original image is expressed as the sum of a rough approximation and the total final improvement:

$$X = A_2 + D_{1plus2} \text{ (4}^{\text{th}} \text{ line)}$$

In Figure 8.5 D_1 and D_{1plus2} highlight the contour lines of the original image. We will take advantage of this property in the following section.

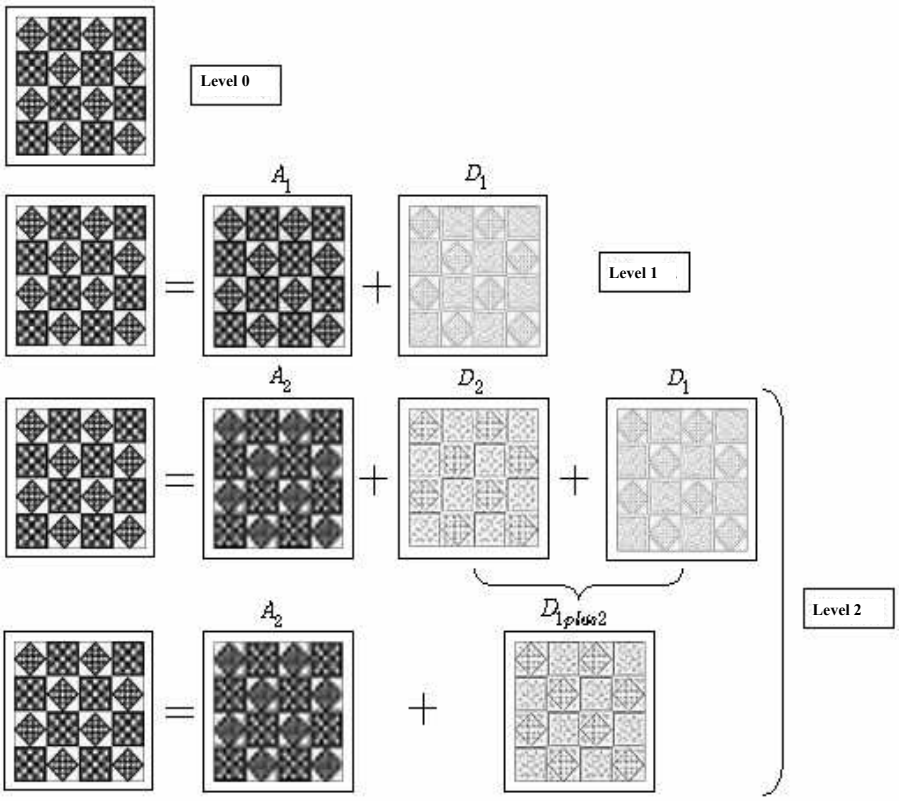


Figure 8.5. *Approximations and details of the "tartan" image analyzed at level 2 with the wavelet sym4*

8.3. Edge detection and textures

Here we deal with the edge detection using three examples: the first one, which is elementary, is associated with a synthetic image, whereas the two following ones are associated with real images.

8.3.1. *A simple geometric example*

Let us consider a grayscale image containing three simple geometric forms (see Figure 8.6).

We carry out a level 1 analysis with the wavelet *db2* and then we calculate the approximation and detail of level 1. The image X is expressed as $X = A_1 + D_1$. In Figure 8.7 we represent A_1 on the left and D_1 on the right. For the approximation A_1 we use the same palette of colors as for X. For the detail D_1 we proceed differently because the amplitude of its values implies a bad representation. We choose a two-color representation: black if $|D_1| > 0$ and white if $|D_1| = 0$.

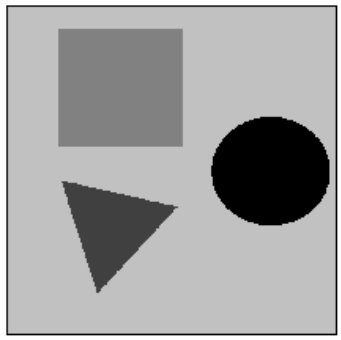


Figure 8.6. *Original image: three simple geometric forms*

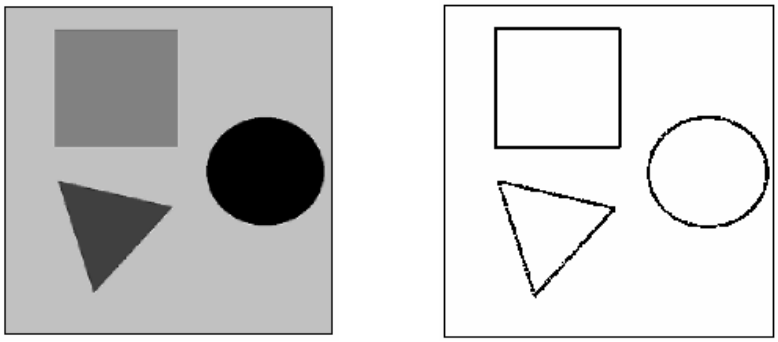


Figure 8.7. *Approximation A_1 (on the left) and detail D_1 (on the right)*

The approximation A_1 resembles the initial image. It contains the texture zones. Let us note that in a certain manner A_1 is a "compressed" version of the initial image since it is obtained with four times less information.

Via a suitable representation and subject to the use of a wavelet with small support, the details enable the edge detection, as we can see on the right side in Figure 8.7.

8.3.2. *Two real life examples*

We now analyze two real images showing certain marked geometric features. These analyses are carried out at level 1 using the Haar wavelet and the same techniques are applied to both the images.

We decompose the image as previously: $X = A_1 + D_1$. The approximation A_1, which is very similar to the initial image X, is visualized using the same color palette as X. The detail D_1, in turn, is coded in two fashions:

– $D_1^{Grayscale}$ obtained by dividing the variation band of the values of $|D_1|$ into 255 intervals. To each of them we allot a shade of gray;

– $D_1^{Two\text{-}color}$ obtained using two-color coding defined with the aid of a threshold: black if $|D_1|^3 \geq threshold$ and white if $|D_1| < threshold$.

This second representation suppresses the fluctuations of grayscale and extracts the skeleton of the large coefficients from which the edges are detected.

Each analysis is summarized by four images laid out as follows:

X	A_1
$D_1^{Grayscale}$	$D_1^{Two\text{-}color}$

In the first image analyzed (see Figure 8.8), vertical, horizontal and oblique lines clearly mark the edges of regular zones. They are well located by the $D_1^{Two\text{-}color}$ detail. The building shade zone located at the foreground is also perfectly delimited by a contour. The stone walls (in the foreground and at the back), the tree (on the left) and the paved path (in the center), which are the zones with irregular texture, yield messy details.

The second image (see Figure 8.9), representing a porch, also contains regular geometric patterns: vertical, horizontal, oblique and arcs of a circle. They are all perfectly detected in detail D_1. The edges of shady zones are also clearly visible.

These examples show that, almost without any processing, the image analysis using wavelets makes it possible to extract a new image like $D_1^{Two\text{-}color}$, from which the edges can be isolated. To improve edge detection we may change the analyzing wavelet, the level of decomposition or the threshold, but also reprocess $D_1^{Two\text{-}color}$, for example, by erasing all isolated black spots.

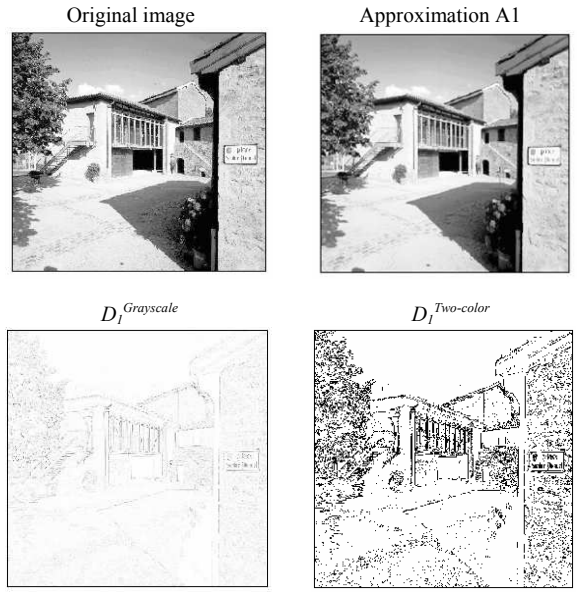


Figure 8.8. *Edge detection (a building)*

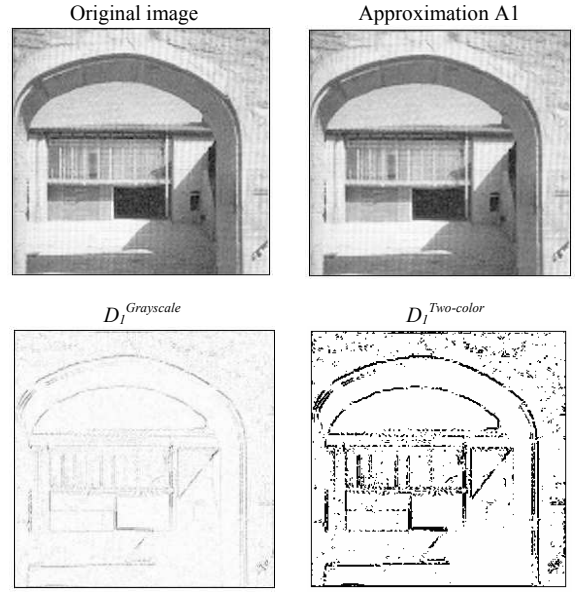


Figure 8.9. *Edge detection (a porch)*

8.4. Fusion of images

In this section, dedicated to the fusion of images by wavelets, we consider three situations. The purpose of the first, constructed on a simple example, is to present the problem and the technique used. The second, which is more realistic, amalgamates two fuzzy images. The last, a little more playful, amalgamates a mask of the Japanese Nô theater with the bronze bust of a Roman emperor.

8.4.1. *The problem through a simple example*

Using an original image represented on the left in Figure 8.10 we construct two images of the same size. The first one *I1*, in the middle, is obtained by darkening a disk in the center of the image. The second one *I2*, on the right, is "the supplement" of the previous one. In this context the problem of image fusion by wavelets consists of reconstituting the original image using the wavelet decompositions of the images *I1* and *I2*.

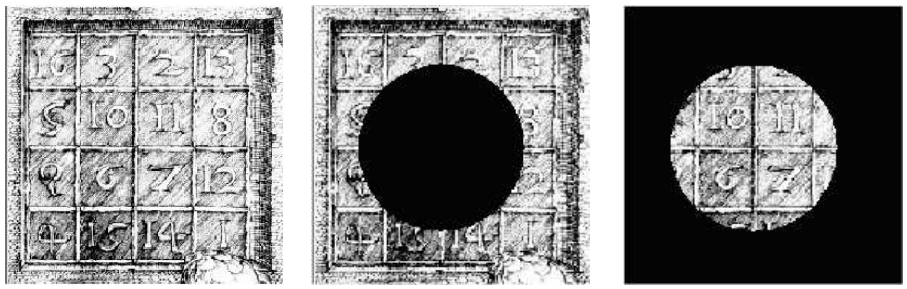


Figure 8.10. *The original image on the left and the extracted images I1 and I2*

The technique involves three phases:

– decompose the images to be fused in the same wavelet basis;

– combine the two decompositions in order to obtain a new one;

– use the inverse transform to construct a new image.

Let us apply this technique to the above example. Let us first decompose the images *I1* and *I2* at level 2, with the wavelet *db2*. The two decompositions are presented in Figure 8.11.

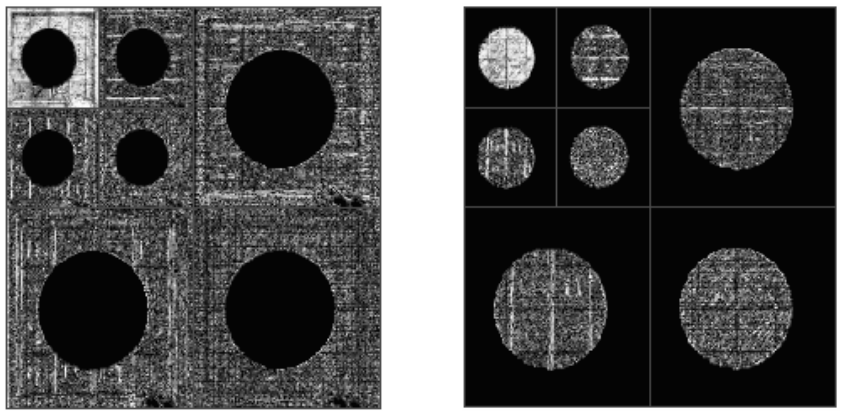


Figure 8.11. *Decomposition of the image I1 (on the left) and the image I2 (on the right)*

Since the two images are of the same size, the decompositions also have identical sizes. The idea now is "to mix" the two decompositions. We directly construct a new decomposition of the right size, for which each coefficient is obtained by combining the corresponding coefficients of the decompositions of *I1* and *I2*.

How can we perform this mixing? We may choose to use a linear combination (the average for example) or the maximum.

In this example we choose the strategy of the maximum for the level 2 approximations and all the details. It is particularly relevant *a priori*, since the two images are complementary. The wavelet analysis being local, this property is preserved in the decompositions, as Figure 8.11 reflects very well.

The origin of the detail coefficients of the new decomposition (D_1 in the first row and D_2 in the second row) is depicted in Figure 8.12 in two colors. The coefficients coming from *I1* are shown in light gray, otherwise they are shown in dark gray.

The use of the inverse transform then constructs a "mixed" image. On the left in Figure 8.13 we find the new decomposition and on the right is the fused image obtained by the inversion of the DWT. The joining "scar" of circular form is far from visible with the naked eye, but it exists and arises from the "uncertainty" zone at the border of the disk. This zone is easily distinguished in the second row of Figure 8.12.

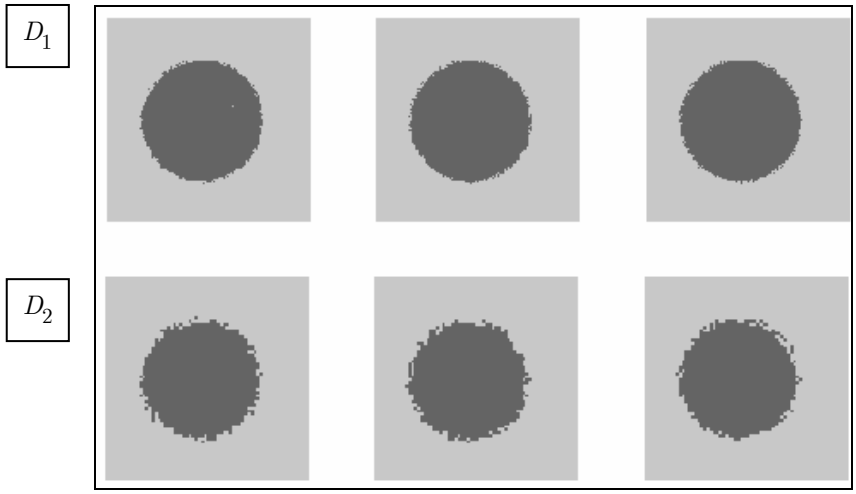


Figure 8.12. *Source of the level 1 and 2 detail coefficients*

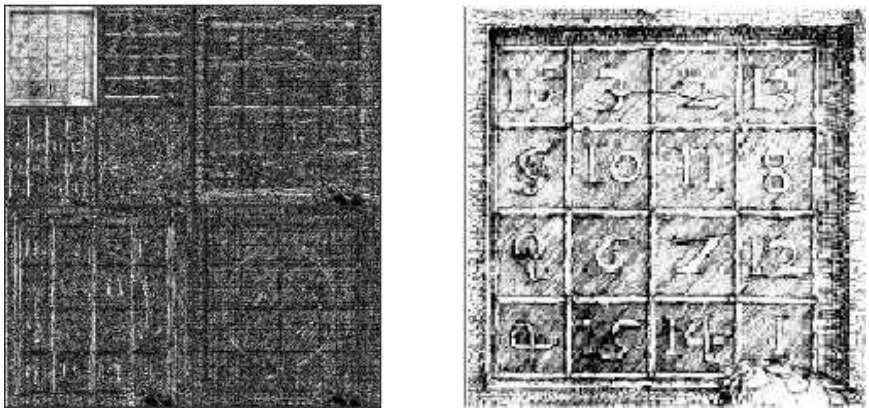


Figure 8.13. *Fused decomposition and image*

We have thus fused the images *I1* and *I2*.

The preceding procedure, with some technical adjustments, is usable to process less academic examples. We can thus amalgamate images of different size, when it becomes necessary to make them the same size and to carry out the required positioning shift.

8.4.2. *Fusion of fuzzy images*

We can also start with two corrupted versions of the same image and try to reconstitute it as well as possible.

In Figure 8.14 we find two images (noted *I1* and *I2*) with fuzzy zones whose intersection is very small.

Figure 8.14. *The corrupted images I1 (on the left) and I2 (on the right)*

The technique and the strategy for choosing coefficients are the same as those adopted in the previous example.

The two images to be fused are analyzed at level 2, with the wavelet *db2*, and the decompositions are represented in Figure 8.15.

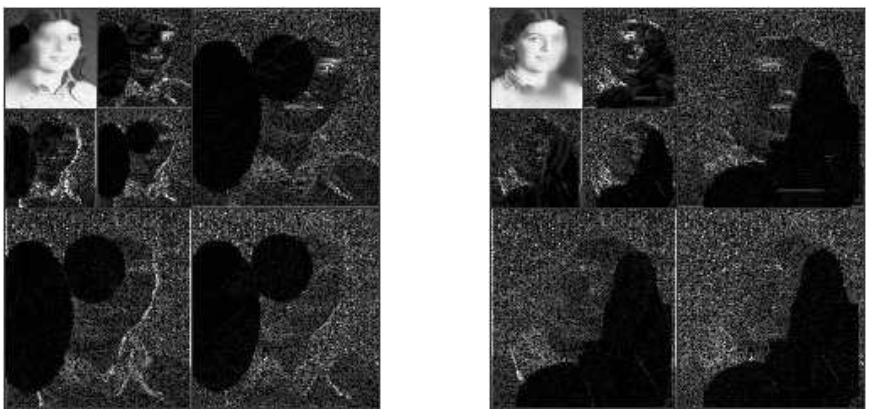


Figure 8.15. *Decomposition of the I1 image (on the left) and the I2 image (on the right)*

The fuzzy zones in the various details are clearly visible.

The situation is different from that of the preceding example where the two images to be merged were obtained from the same image by complementary screenings. Here the blur affects almost disjoined zones, but relation $X = I1 + I2$ is no longer true. Nevertheless, the strategy of maximum is to be retained, since the blur affecting one of the images, for a given pixel, stems from a local average and thus generates a zero detail coefficient. For this reason, the black zones of the two decompositions in Figure 8.15 correspond to the fuzzy zones of the images *I1* and *I2*. The same argument holds true for the approximations, since the use of the maximum will select non-regularized coefficients.

In Figure 8.16 we see the origin of the coefficients of the new decomposition represented as in section 8.4.1.

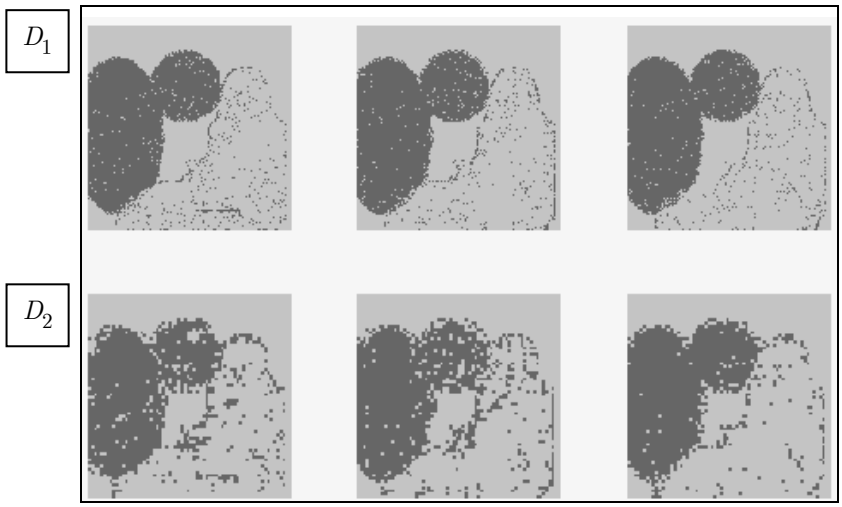


Figure 8.16. *Source of the level 1 and 2 detail coefficients*

We note here that the zones delimited by these origins are less homogenous than in the preceding example. That reflects the greatest difficulty of the problem.

The process ends by obtaining the new decomposition, on the left in Figure 8.17, and the fused image, on the right.

Figure 8.17. *Fused decomposition and image*

The result is satisfactory. We obtain a new image that at first sight no longer contains fuzzy zones. Nevertheless, an attentive examination shows that a small fuzzy zone persists in the collar of the shirt, below on the left. It is difficult to distinguish on the fused image, but it is obvious in the level 1 details of the new decomposition. It corresponds to a non-empty intersection between the fuzzy zones of the images *I1* and *I2*.

8.4.3. *Mixing of images*

In the previous examples, the two images to be fused were corrupted versions of the same original image. Here we consider a new problem: it consists of constructing a new image by combining two different images. The technique used is the same, except possibly for the manner of combining the decompositions.

The two images used in this example represent a mask of the Nô theater, on the left in Figure 8.18, and the bust of a Roman emperor, on the right.

Figure 8.18. *The two initial images I1 and I2*

These two images have the same size: 256 × 256. They are decomposed at level 2 with the *bior6.8* wavelet, often used in 2D. The two decompositions are depicted in Figure 8.19 and the details are presented "in negative". By considering the detail coefficients we notice that the *I1* image is more regular than the *I2* image.

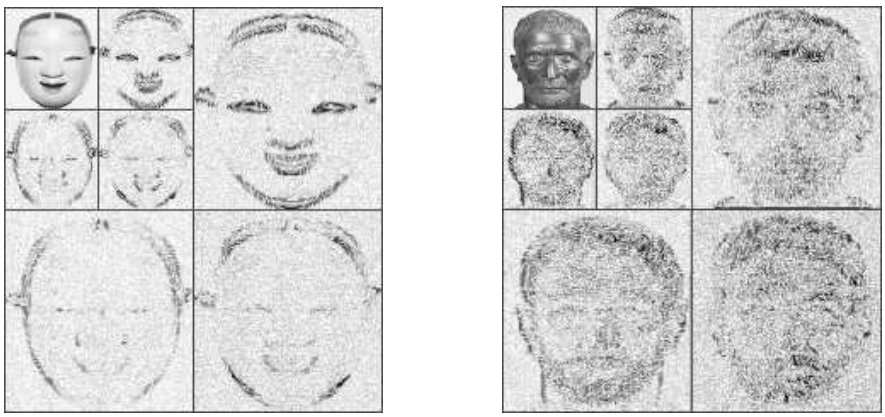


Figure 8.19. *Decomposition of the I1 image (on the left) and the I2 image (on the right)*

Initially, in order to construct the new decomposition, we adopt the strategy based on the average, simultaneously for the approximation and the details coefficients. The new decomposition and the fused image are represented in Figure 8.20.

When the detail coefficients of the two decompositions are considered, we note that the large coefficients are definitely fewer for the first image. Averaging out the details in fact leads to decreasing the marked features of the two images in the reconstructed one: beard, hair, etc. We may reduce this effect and reinforce the edges by choosing the maximum instead of the average for the detail coefficients. Figure 8.21 shows the decomposition and image obtained in this manner.

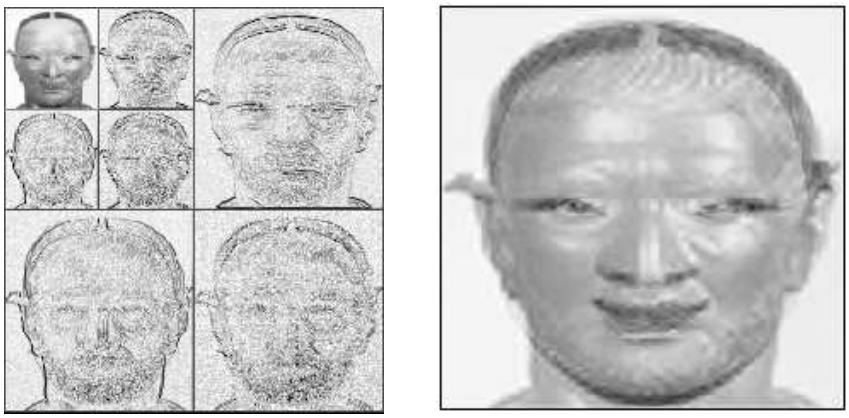


Figure 8.20. *Fused decomposition and image: average for the details and the approximation*

Figure 8.21. *Fused decomposition and image: average for the approximation and maximum for the details*

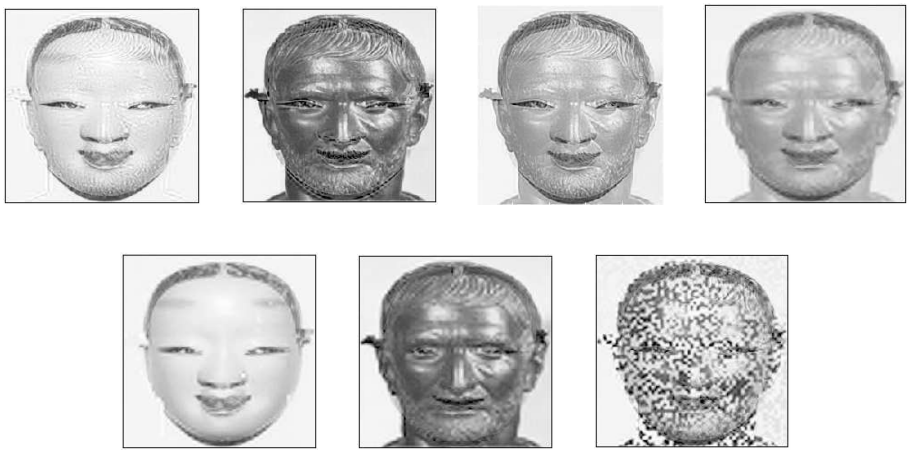


Figure 8.22. *Various fused images obtained by modifying the way of combining coefficients of approximations and details*

Obviously, we may choose many other strategies to "mix" the coefficients of the two decompositions and, thus, to obtain a gallery of portraits. If we note by *(a, d)* the couple of combinations relating to the approximations and the details, in Figure 8.22 we find the fused images corresponding to the following combinations: from left to right, at the top, *(maximum, maximum)*, *(minimum, minimum)*, *(average, minimum)*, *(average, details of the bust)* and then at the bottom *(maximum, minimum)*, *(minimum, maximum)*, *(random, random)*.

All the previous fusions are obtained by applying a global strategy for choosing the origin of the coefficients. Evidently, it is possible to use local choices.

As an example let us start with the two preceding decompositions and construct a new decomposition. Let us call C one of the decomposition coefficients matrices and A, B the corresponding matrices for the decompositions of *I1* and *I2*. The i^{th} line of C is obtained by: $C_{ij} = (1 - t_i) A_{ij} + t_i B_{ij}$. While i varies, t_i is a linear function of *i* with a value of 0 for the first row and 1 for the last. Thus, the 1st line of C is equal to the 1st line of A and the last line of C coincides with the last line of B. The decomposition and the image obtained in this manner are represented in Figure 8.23.

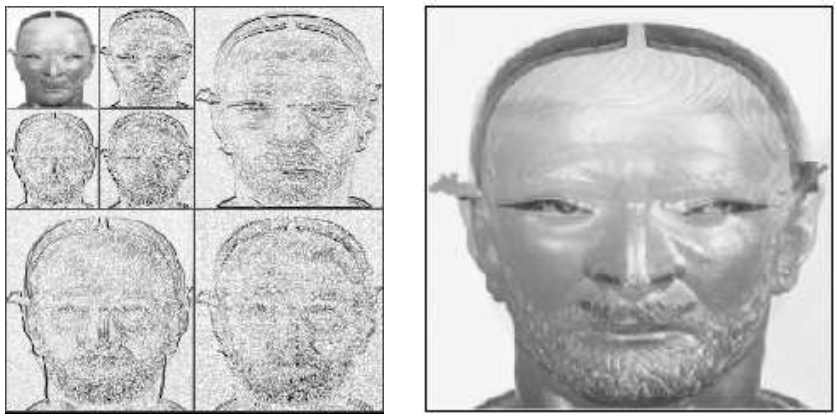


Figure 8.23. *Fused decomposition and image: combination by row*

By permuting the role of the two images and/or by modifying the weights of the linear combinations we obtain a variety of fused images, as shown, for example, by Figure 8.24.

Figure 8.24. *Various images fused by row by row combination*

Fusing images with wavelets thus constitutes a simple and easily adaptable technique to construct synthesized images.

8.5. Denoising of images

In this section we consider the problem of denoising of images. The methods detailed in the case of the 1D signals (see Chapter 7) can be extended to image denoising. However, they are in general less powerful.

In the first section we present the problem and the technique on an example of an artificially noisy image. Then, in the second section, we apply a similar technique to an image noised in an uncontrolled way.

8.5.1. *An artificially noisy image*

Figure 8.25 presents on the left the original image noted f and on the right the noisy image noted Z.

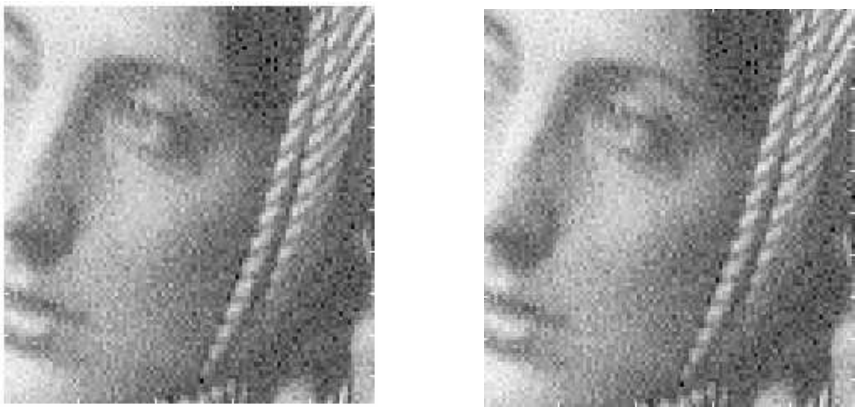


Figure 8.25. *Original image (on the left), noisy image (on the right)*

As in the 1D case we suppose that the image Z is the original image f corrupted by homogenous additive noise. That is, we consider a model of the type:

$$Z(i,j) = f(i,j) + \varepsilon_{i,j} \qquad [8.1]$$

where Z is the image to be denoised, ε is a white noise with a covariance matrix $\sigma^2 I$ and f is the image being restored.

The process of denoising by wavelets consists of three phases:

– wavelet decomposition at the J level of the image to be denoised;

– thresholding, in three directions, of the detail coefficients with smaller absolute value than a threshold dependent on σ;

– reconstruction of the denoised image, starting from the approximation coefficients of level J (unchanged) and of the modified detail coefficients.

We carry out the wavelet decomposition of the noisy image by the Haar wavelet, at level 4, and obtain the results presented in Figure 8.26.

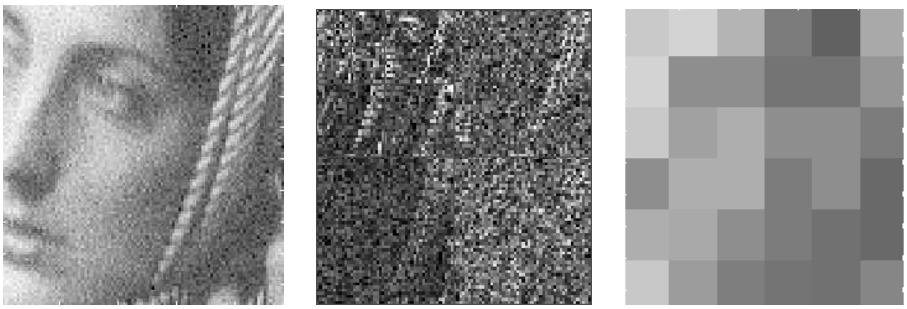


Figure 8.26. *Noisy image (on the left), level 4 decomposition (in the middle) and level 4 approximation (on the right)*

The principle of denoising is simple. We start with a coarse version of the image, here the level 4 approximation and then we select detail coefficients to construct the necessary touching up in order to improve the result.

We will present three successive tests, without specifying the choice of thresholds but focusing on the results (see [MIS 00]).

Let us begin by thresholding the Haar decomposition coefficients seen in Figure 8.26. The threshold is obtained by estimating σ and applying a penalization strategy.

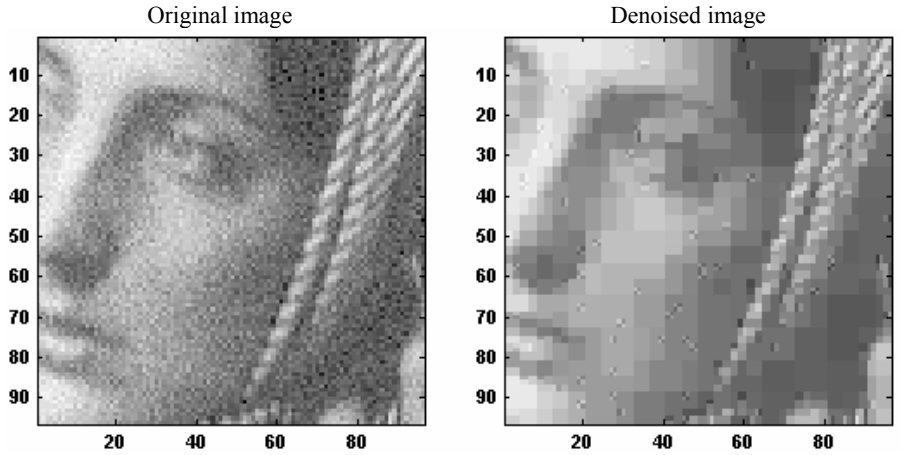


Figure 8.27. *Noisy image (on the left) and image denoised by Haar (on the right)*

In Figure 8.27 we find the noisy image on the left and on the right the denoised image. The latter is not satisfactory, because it presents serious blocking effects. It is one of the common drawbacks of the Haar wavelet. Let us take a more regular and almost symmetric wavelet *sym6* and apply a similar strategy.

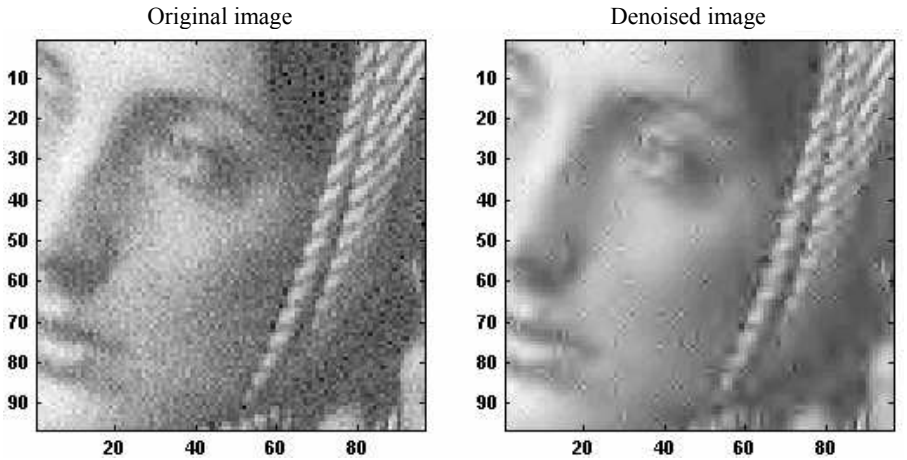


Figure 8.28. *Noisy image (on the left) and image denoised by sym6 (on the right)*

The denoised image (see Figure 8.28) no longer presents any blocking effects but now exhibits other more local imperfections.

In fact, direct adaptations of the usual strategies of the 1D denoising to the two-dimensional case do not apply to images in a satisfactory way. On the other hand, the use of the invariant transform by translation (known as SWT; see Chapter 3), whose contribution to 1D denoising (see Chapter 7) remains rather simple, improves the results for 2D in a sometimes spectacular way.

The principle, in this context, consists of finding the average of several denoised images. They are obtained using the usual scheme of denoising applied to slightly different decomposition coefficients.

Thus, we have here two competing effects: a better edge detection due to taking into account all the shifts in the analysis and a more intense smoothing effect due to averaging the images denoised on each basis.

The application of the SWT to the example considered is presented in Figure 8.29. The wavelet used is *sym6* and the decomposition is carried out at level 3.

The result is very satisfactory from the point of view of noise. On the other hand, we note a loss of definition of the edges and the denoised image is fuzzy.

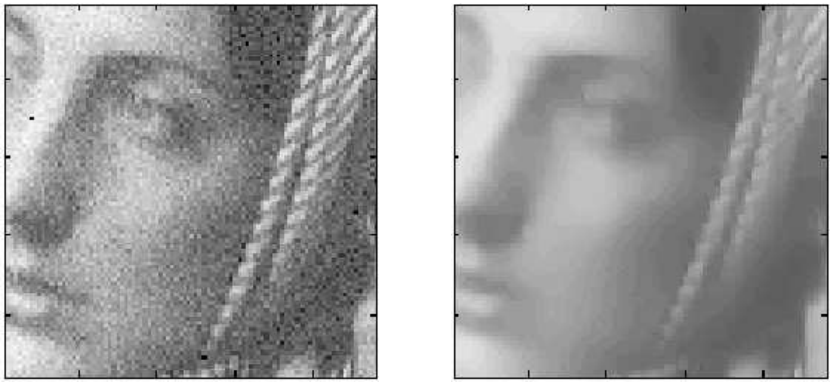


Figure 8.29. *Noisy image (on the left) and image denoised by the SWT (on the right)*

After this example in which we added "a good" noise to a good quality image, let us move on to a noisy image obtained via an uncontrolled process.

8.5.2. *A real image*

The image to be denoised is presented in Figure 8.30. It is a real image coming from the digitalization of an ID photo. It is of a rather bad quality, as can be seen from the hair, for example. An additional noise seems to have been superimposed on the original image and, moreover, white spots are deteriorating the texture of the face.

We try to improve the image by applying denoising methods similar to those seen previously. Let us consider again the model [8.1]:

$$Z(i,j) = f(i,j) + \varepsilon_{i,j}$$

where f, the image to be restored, is in this case the initial ID photograph.

Let us first apply the denoising method by SWT using the wavelet *sym6* and decomposing at level 3. The result is shown in Figure 8.31.

Figure 8.30. *Original image (Catherine)*

The image is suitably denoised, thus illustrating the robustness of the method with respect to the probabilistic structure of the noise. However, it is slightly fuzzier than the original image, as in the previous case.

Figure 8.31. *Denoising of Catherine by SWT*

8.6. Image compression

The problem of compression is often handled using image compression. For a digital image, the essential objective of compression is to minimize the length of the series of bits necessary to represent it, while storing information of acceptable quality. Wavelets contribute to efficient solutions to this problem. Of course, the complete chain of compression includes phases of quantization, coding and decoding in addition to the wavelet processing itself. If we concentrate on it, the objective consists of obtaining the sparsest representation as possible (i.e. comprising the largest number of negligible terms) of the image to be compressed. Sometimes we accept a weak degradation of the compression performance in order to satisfy complementary objectives: real-time transmission, uninterrupted reconstitution of increasingly fine versions of the same image, etc.

In this section we first present the principles of compression by wavelets and we then concentrate on the aspects specifically related to the latter. Finally, we tackle the complete compression chain concentrating on quantization and coding.

8.6.1. *Principles of compression*

Starting from the image to be compressed, we first perform decomposition on an orthogonal or biorthogonal wavelets basis by means of the discrete transform. Then we select a part of the coefficients by thresholding, while preserving intact the approximation coefficients of an appropriately selected level. The kept coefficients are then quantized and, to finish off, they are encoded for purposes of storage or transmission. Decompression consists of inversing the previous operations as far as possible. From the decoded and "dequantized" coefficients we rebuild an image by applying the inverse discrete transform. The image obtained is thus the compressed image.

All of these operations are schematized in Figure 8.32. We then speak about "true" compression. This point will be developed in section 8.6.3.

This type of compression can lead to a loss of information during one of the two following stages:

– at the thresholding step, when the value of certain coefficients is modified;

– at the quantization step, when the value of certain coefficients is truncated.

The first "error" can be suppressed if no thresholding is performed and the second can be avoided by the use of wavelets with integer or rational coefficients.

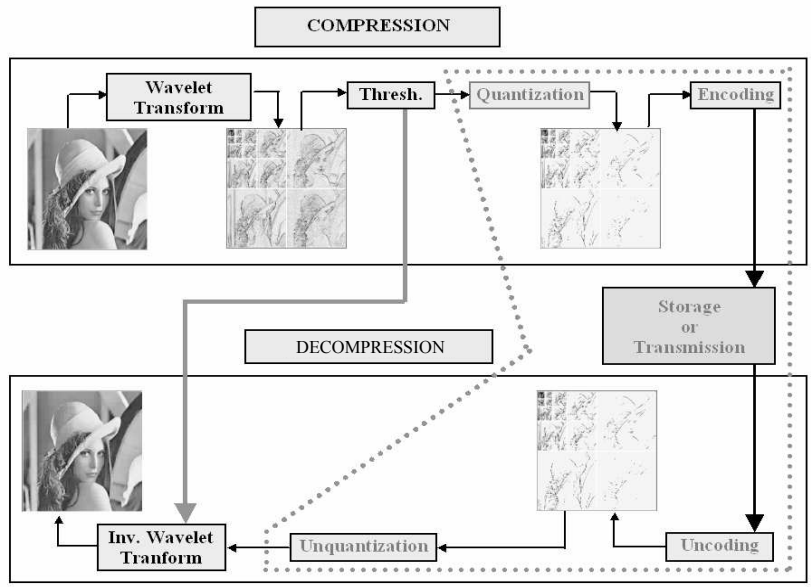


Figure 8.32. *The diagram of compression – grayed stages (in the right part in dotted lines) are not a priori directly related to wavelets*

Initially, we concentrate on the aspects related to wavelets. They are involved in the first step of compression and the last step of decompression. *A priori* the other operations do not depend on the use of wavelets; they are grayed out in Figure 8.32. However, certain methods of coding exploit the arborescent structure of the wavelet decomposition. Let us concentrate on the path which links, in the figure, the transform, the thresholding and the inverse transform.

8.6.2. *Compression and wavelets*

8.6.2.1. *Why does it work?*

The key argument used to understand the efficiency of these methods is the capacity for economic representation of broad image classes in wavelet bases. Indeed, images generally have very sparse wavelet decompositions, i.e. representations where few coefficients are significantly different from zero. They are very well represented by the coefficients of a rather rough approximation supplemented by some large detail coefficients. Compression would then consist of keeping the "good" coefficients and coding them in the most efficient way possible.

In the wavelet domain, we therefore concentrate the detail coefficients around zero: this is the key to compression. Let us illustrate this fundamental idea with three examples.

Below we present the grayscale histogram of the original image and that of the wavelet coefficients for three images: Lena, a fingerprint and a porch. The analyses are carried out at level 8 using the Haar wavelet, with a periodic extension mode (see Chapter 3).

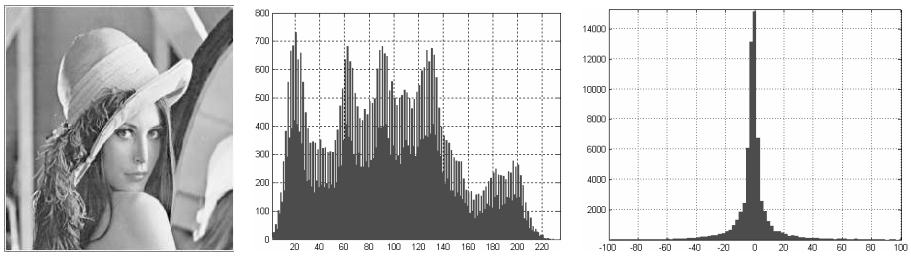


Figure 8.33. *Lena (256 × 256), histogram of values (between 1 and 235) and histogram of wavelet coefficients (zoom on the significant part)*

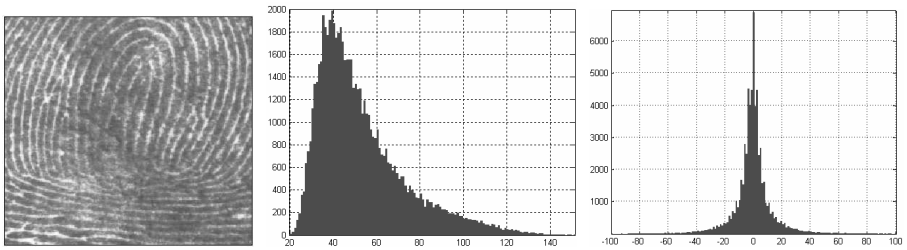


Figure 8.34. *A fingerprint (256 × 256), histogram of values (between 19 and 152) and histogram of wavelet coefficients (zoom on the significant part)*

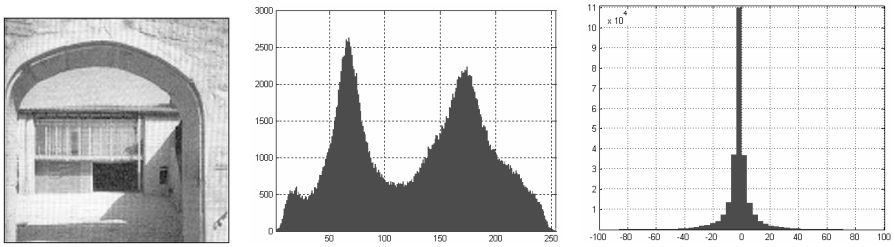


Figure 8.35. *A porch (512 × 512), histogram of values (between 1 and 255) and histogram of wavelet coefficients (zoom on the significant part)*

The remarkable fact is that, despite the very different grayscale distributions of the three images (multimodal for the first one, unimodal asymmetric for the second one and bimodal for the last one), the distributions of the wavelet coefficients values are very similar. The significant part of these distributions is symmetric, unimodal with the mode at the origin.

8.6.2.2. *Why threshold?*

The simplest framework to introduce the core of wavelet compression consists of taking a finite energy function f and considering an orthonormal wavelet basis of $L^2\left(\mathbb{R}^2\right)$ noted simply $(g_m)_{m\in\mathbb{N}}$. The 15 or so lines that follow are taken from Chapter 7 and cited here to render this chapter self-sufficient.

The representation of f in this basis is $f = \sum_{m\in\mathbb{N}} (f, g_m)\, g_m$. The coefficients of f are $\{(f, g_m)\}_{m\in\mathbb{N}}$; we will note them simply as $\alpha_m = (f, g_m)$.

Performing a compression consists of selecting a fixed number M of coefficients. Thus, generally, a compression is carried out with loss. Let us note I the set of indices of M selected coefficients and $f_M = \sum_{m\in I} \alpha_m g_m$ the approximation obtained. If we agree to judge the quality of compression by using the quadratic error, we have:

$$\left\| f - f_M \right\|^2 = \sum_{m\notin I} \alpha_m^2$$

Since $\|f\|^2 = \sum_{m\in\mathbb{N}} \alpha_m^2$, it is clear that the best choice of I consists of keeping the M longest absolute value coefficients. Noting $\left\{\alpha_{m_p}\right\}_{p\in\mathbb{N}}$ and the coefficients being sorted in a descending order, the optimal quadratic error is then:

$$\left\| f - f_M \right\|^2 = \sum_{p\geq M} \alpha_{m_p}^2 .$$

We can define the compressibility curve associated with the representation of f: on the x-axis we find the number M of kept coefficients and, on the y-axis, the proportion of retained energy $\left\| f_M \right\|^2 / \|f\|^2$, which is also written $1 - (\left\| f - f_M \right\|^2 / \|f\|^2)$ if we express it as a function of the associated relative quadratic error.

To illustrate the theoretical aspects developed above, two compressibility curves are represented in Figure 8.36. The first uses fingerprint image decomposition on the

canonical basis. The second is built on the basis of coefficients obtained by a level 8 analysis with the *db1* wavelet.

It is clear that the representation in wavelets "concentrates" energy on a very small number of coefficients. The largest image coefficient corresponds to 0.01% of energy. The largest coefficient of wavelet decomposition, which is the single coefficient of the order 8 approximation, captures 87.36% of energy. Let us note that the corresponding *A8* approximation, although dominating in energy, is not very informative visually since it is reduced to a monochromatic square.

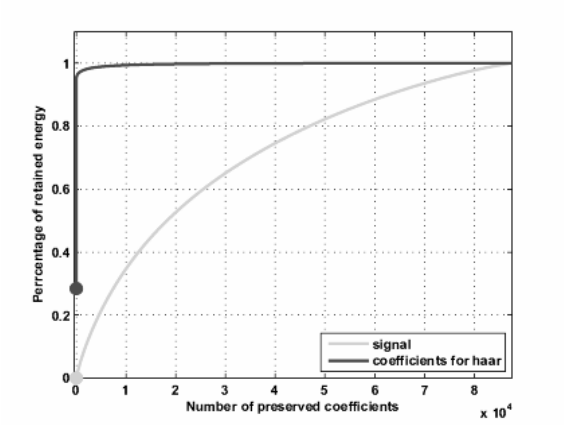


Figure 8.36. *Comparison between energy percentages for the fingerprint*

8.6.2.3. *Examples of image compression*

We decompose the fingerprint at level 5 with the *sym4* wavelet and impose 94% of zeros in the "thresholded" decomposition corresponding to a universal threshold of 28.61. In Figure 8.37 we have represented: on the top left, the original image, on the top right the image reconstructed after selection of coefficients, and below, the thresholding process performed on the histogram of the wavelet decomposition.

The result is satisfactory; however, we note a small deterioration of the edge definition in the right bottom part of the image.

Once thresholding has been performed on the 65,536 coefficients, only about 3,932 non-zero coefficients remain. In particular, the 64 coefficients of the level 5 approximation are all retained. The distribution by scale level of the kept coefficients is indicated in Figure 8.38.

Let us now decompose a more regular image, with a size of 512 × 512, at the level 5 with the *sym8* wavelet. We impose 98.50% of zeros in the thresholded

decomposition. Kept energy is then very close to 100%. The result (see Figure 8.39) is so good that we could almost permute the two images, whereas the image on the right is rebuilt with only 1.5% of the coefficients.

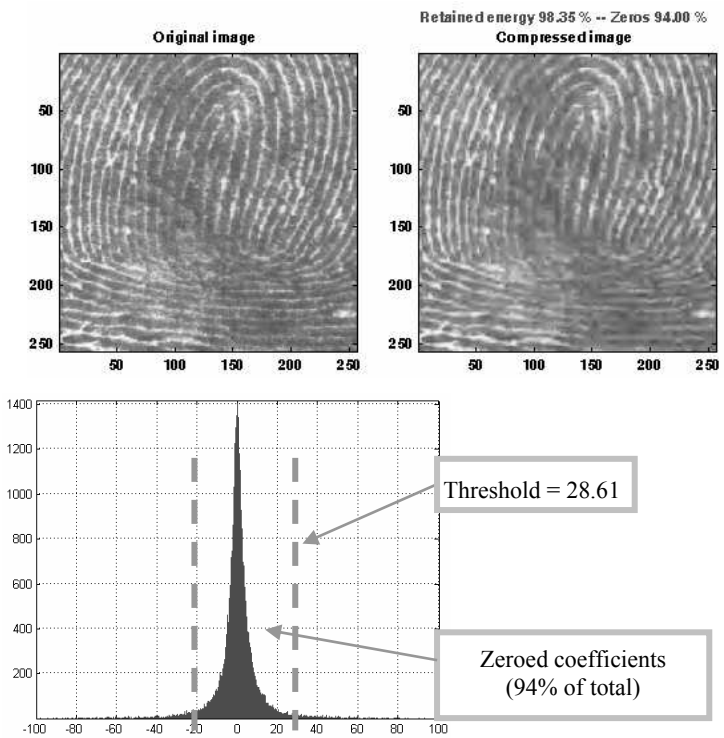
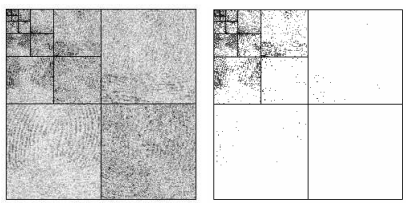


Figure 8.37. *Compression of a fingerprint*



	Number	Kept	Ratio kept/ number by level	Ratio kept/total number of kept coeff.
A5	64	64	100.00%	1.63%
D5	192	157	81.77%	3.99%
D4	768	420	54.69%	10.68%
D3	3,072	1,460	47.53%	37.13%
D2	12,288	1,704	13.87%	43.34%
D1	49,152	127	0.26%	3.23%
Total	65,536	3,932	6.00%	100.00%

Figure 8.38. *On the left, the complete decomposition, in the middle the "thresholded" decomposition and on the right the coefficients distribution for the fingerprint*

Once thresholding has been performed, from the 262,144 coefficients, only 3,932 that are non-zero remain. Again, all the level 5 approximation coefficients are kept. Let us also note that a negligible part of the level 1 detail coefficients is kept. As in the preceding example, the distribution by scale level of the kept coefficients is indicated in Figure 8.40.

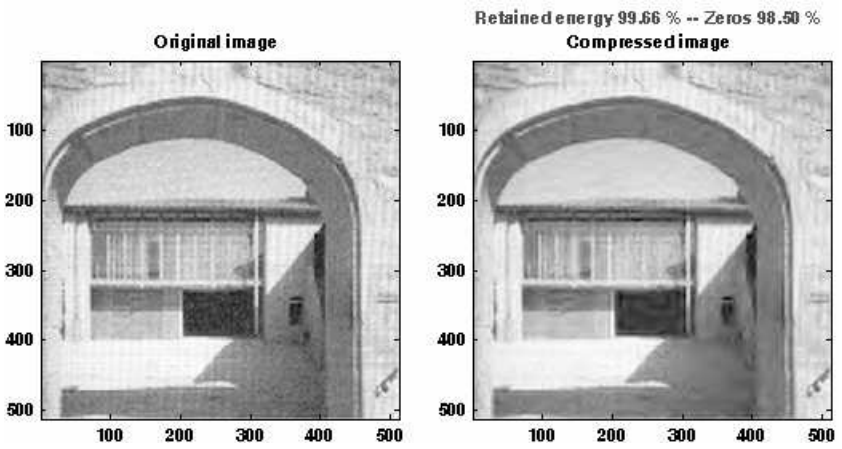
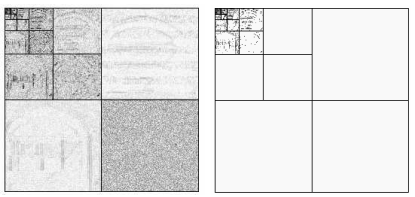


Figure 8.39. *Compression of a porch*



	Number	Kept	Ratio kept/ number by level	Ratio kept/total number of kept coeff.
A5	256	256	100.00%	6.51%
D5	768	480	62.50%	12.21%
D4	3,072	1,127	36.69%	28.66%
D3	12,288	1,881	15.31%	47.84%
D2	49,152	133	0.27%	3.38%
D1	196,608	55	0.03%	1.40%
Total	262,144	3,932	1.50%	100.00%

Figure 8.40. *On the left the complete decomposition, in the middle the "thresholded" decomposition, and on the right the coefficients distribution for the porch*

NOTE.– as detailed in Chapter 2, biorthogonal wavelets are particularly efficient for compression because the analysis part (decomposition) and the synthesis part (reconstruction) are treated by two distinct wavelets. The "sparse representation" and "rendered decompressed image" aspects are thus partially uncoupled.

We can also not restrict ourselves to wavelets and possibly use wavelet packets. Obviously, all these elements play a part in the performance of the compression: on the compression ratio, but also on rendered reconstructed image.

8.6.3. *"True" compression*

We have just dealt with aspects specifically related to compression by wavelets (see Figure 8.32). However, in addition to the algorithms related to wavelets, DWT and IDWT, it is necessary to use other ingredients concerning the quantization mode and the coding type. This set constitutes the true compression represented by Figure 8.41.

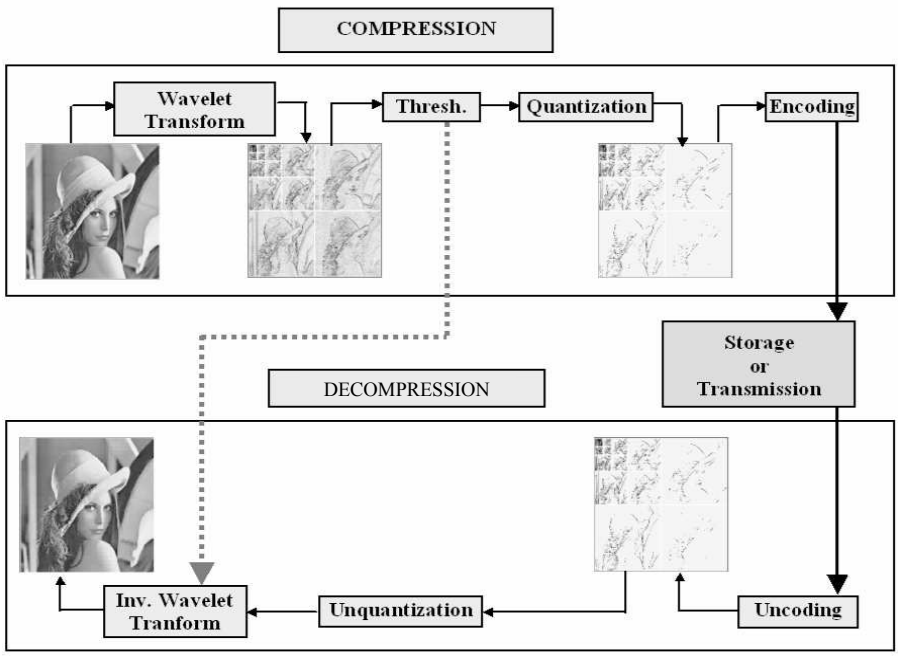


Figure 8.41. *The complete diagram of true compression*

Quantization transforms a domain containing an infinite number or even a finite but large number of values, in a finite and small set of values. This operation is, apart from the thresholding, the only step where information is lost. Coding usually consists of transforming a finite set of symbols obtained, for example, after a quantization, into a finite stream of bits, i.e. a sequence of 0 and 1. Acceptable

coding does not involve any loss of information. Now let us detail each of these two operations.

8.6.3.1. *The quantization*

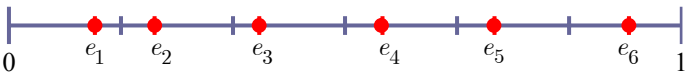
The coefficients obtained by applying the DWT and thresholding to an image are generally real numbers. Sometimes they are integers, but their variation domain is large. The quantization phase consists of approximating the set of values of the coefficients by a finite set of numbers and, in general, implies a loss of information.

Let E be a set, P a finite partition of E and F a set containing one and only one element of each class of the partition P. The quantizer q associated with (E,P,F) is the application for which $\hat{x} \in F$ is the image of $x \in E$. If $E \subset \mathbb{R}^K$ we refer to a vector quantizer and if $E \subset \mathbb{R}$, we refer to a scalar quantizer.

Let us give a simple example of a quantizer:

$$E = [0,1] \ , \ P = \left[0,\tfrac{1}{6}\right[\cup \left[\tfrac{1}{6},\tfrac{1}{3}\right[\cup \cdots \cup \left[\tfrac{2}{3},\tfrac{5}{6}\right[\cup \left[\tfrac{5}{6},1\right] = \bigcup_{k=1}^{6} I_k \text{ and}$$

$$F = \left\{e_1,e_2,e_3,e_4,e_5,e_6\right\} \text{ with } e_k \in I_k$$



How can we choose class representatives, i.e. e_i? The error of quantization for the set $X = \{x_n\}_{n=1}^N$ of given elements of E is measured by a distance between X and $\hat{X} = q(X) = \{\hat{x}_n\}_{n=1}^N$, the set of values obtained by quantization. In general, we take the average quadratic error:

$$d\left(X,\hat{X}\right)^2 = \frac{1}{N}\sum_{n=1}^{N} |x_n - \hat{x}_n|^2$$

If a probability distribution E exists, the quantization error is measured by the expectation D of the mean square error.

Let us continue the preceding example and provide $E = [0,1]$ with the uniform probability. The best choice for the set F, minimizing the distortion D, is obtained by taking the middle of each partition segment for e_i:

$$F = \left\{\tfrac{1}{12}, \tfrac{3}{12}, \tfrac{5}{12}, \tfrac{7}{12}, \tfrac{9}{12}, \tfrac{11}{12}\right\}$$

We will now show the effects of quantization on the visualization of the fingerprint. This indexed image corresponds to a matrix of integers ranging between 0 and 255. Through quantization we can decrease the number of usable colors, here 256. Figure 8.42 illustrates the passing from 256 to 16 colors. We see at the top left, the original image, at the top right, the histogram of values, then at the bottom right, the histogram of quantized values and at the bottom left, the reconstructed image. The classes are delimited by dotted lines and their center is illustrated by a dot. The effective interval of values is $[19,152]$ and it is quantized into eight classes.

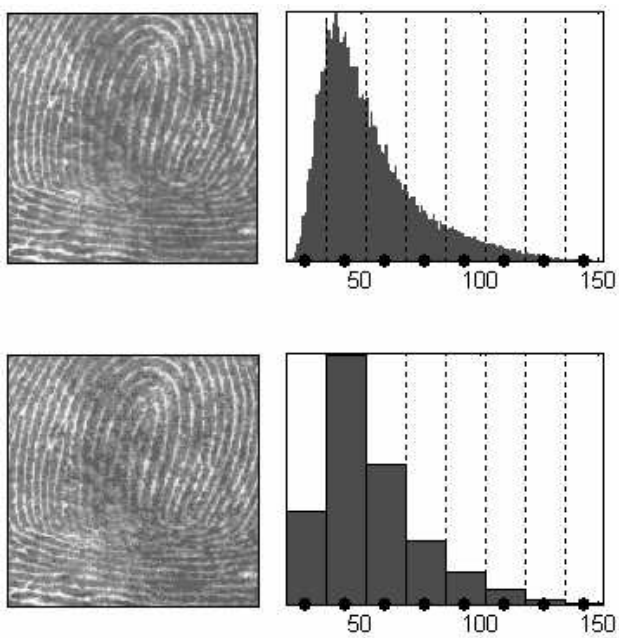


Figure 8.42. *Example of quantization of the original image: a fingerprint*

This quantization leads to a compression of the images. Indeed, with a fixed length binary code we need 8 bits per pixel to code 256 colors and 3 bits per pixel to code 8 of them. We notice that the image obtained after quantization has a good quality. However, within the framework of "true" compression, quantization is used not on the original image, but on its wavelet decomposition.

Now let us move on to the quantization of coefficients.

We decompose the fingerprint at level 4 with the Haar wavelet. The histogram of wavelet coefficients and the quantized histogram are normalized so that the values vary between –1 and +1. The 15 intervals of quantization do not all have the same length. In Figure 8.43 we see at the top left, the original image, the coefficients histogram at the top right, at the bottom right, the histogram of the quantized coefficients and, at the bottom left, the reconstructed image.

The key point is that the histogram of the quantized coefficients is massively concentrated in the class with its center in 0. Let us note that yet again the image obtained has good quality.

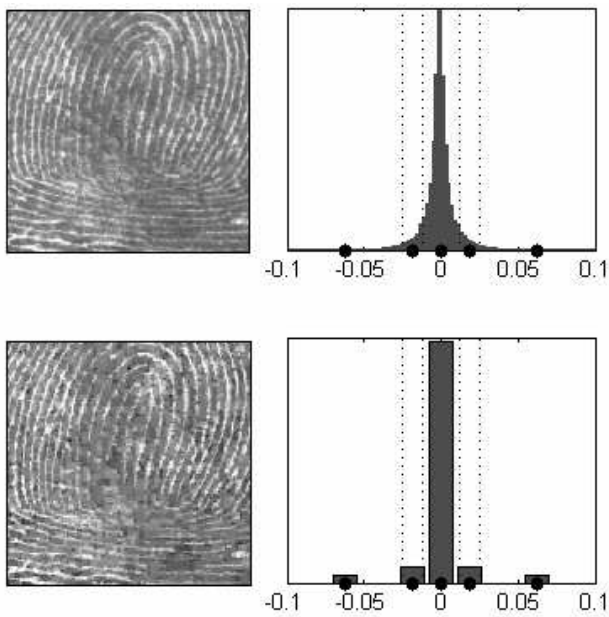


Figure 8.43. *Example of quantization on the wavelet decomposition of a fingerprint (only the five significant intervals of quantization are represented)*

8.6.3.2. *EZW coding*

Once transformation and quantization have been performed we have a finite sequence $\hat{X}$ of elements belonging to a finite set of symbols S. The objective of coding is to represent in the best possible way, for example by a sequence of 0 and 1, the sequence $\hat{X}$ according to predefined objectives.

We now have to code $\hat{X}$ in the most efficient way possible. In general, this means the most economic way possible, i.e. by minimizing the quantity of information used to describe this set. This is the main objective of data compression. Thus presented this problem does not have a direct connection with wavelets; it is a common problem and studied widely in information theory. There are many manners of building a code, let us outline some of them.

The first idea is to code each symbol of S using a binary fixed length word (*fixed length code*). Thus, we obtain a sequence of 0 and 1 coding the sequence $\hat{X}$. From the point of view of compression this method offers practically no interest.

The second idea consists of exploiting the fact that $\hat{X}$ contains a large quantity of zero values. We can then code the most frequent symbols with shorter sequences (*variable length code*). This reduces the quantity of information necessary to represent $\hat{X}$. We can also code only the non-zero values of $\hat{X}$, but their positions should then also be coded. This method, allied to *Huffman type coding* or *semi-arithmetic* or *arithmetic codings* [HUF 52], [AHO 87], [MAL 98], [WIT 87], [HOW 92], leads to satisfactory compression performances, comparable with those of JPEG 92.

The third idea, which is critical for the use of wavelets in image compression, is to fundamentally exploit the tree structure of the wavelet decomposition. Certain codes developed from 1993 to 2000 use this idea, in particular, the EZW coding algorithm introduced by Shapiro [SHA 93], which we will now consider.

The EZW algorithm is based on progressive coding of data. On the one hand, this makes it possible to obtain during decoding an image whose precision, i.e. resolution, increases gradually. In addition, it is possible to obtain a sequence of compression ratios based on the length of the retained code. This compression usually involves a loss of information, but the algorithm also enables compression without loss.

Generally, real images have a low frequency spectrum. Wavelet decomposition transforms them into a set of large (on average) approximation coefficients and detail coefficients that are on average always growing with the scale.

The multi-resolution aspect and the local aspect of wavelet decomposition make it possible to link, at least locally, the coefficients through the decomposition scales. They influence the same zone of the image because the wavelets with which they are associated have close localizations. The central idea of the EZW algorithm is to exploit the inter-scale dependence of these coefficients when coding. Coefficient trees are thus coded only once, in good cases.

More precisely, we link a detail coefficient of the coarsest level J to four children of the J–1 level, and so on down to level 1. The corresponding tree structure is represented in Figure 8.44. We see, for example, that a level 3 coefficient of the horizontal detail (H3) is connected to four level 2 horizontal detail coefficients (H2), each one of them connected to four level 1 horizontal detail coefficients (H1).

The detail coefficients of a wavelet decomposition are then grouped throughout the scales by quad trees whose roots are located at the highest scale. A *zerotree* is a quad tree, whose coefficients are, in absolute value, lower than the root coefficient.

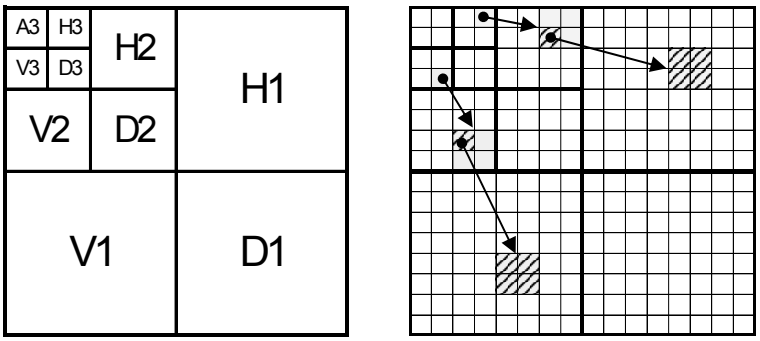


Figure 8.44. *Coefficient trees linked across scales*

The fundamental idea of the EZW algorithm is to exploit the decrease of coefficients across scales and the strong probability of having *zerotree*. When the root coefficient is lower than a fixed threshold, all the others are obviously lower than the said threshold and the whole tree can be coded by a single symbol. We will present the principle of the basic algorithm, whose performances can be improved with additional technical implementation difficulties.

Once the wavelet decomposition is done, the EZW algorithm carries out several successive approximations by quantization of the coefficients. The principle consists of scanning the sequence of coefficients several times. Each scan comprises two stages. During the first one, the principal stage, they are compared with a threshold in order to be classified and possibly coded. The second stage is a refinement stage.

With each stage the threshold decreases making it possible to refine the approximations.

The order in which the coefficients are considered is important. In order to benefit from zerotrees, it is necessary to read the coefficients starting with those of large scales. Each coefficient is thus analyzed before its children. Several possibilities exist (see Figure 8.45).

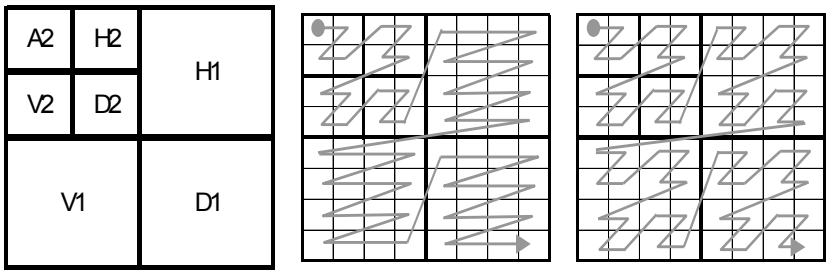


Figure 8.45. *Two possible orders for scanning the coefficients*

If the sequence of thresholds is not predetermined, it should be transmitted to the decoder. However, in general, a "binary" thresholding is used: we initialize the threshold, then for each following stage we divide the current threshold by two. This method makes it possible to work directly on the binary representation of the coefficients.

A detailed description of the EZW algorithm as well as an application example may be found in the Appendix.

In Figure 8.46 we present all the steps of the decoding phase on a real image in order to highlight the progressive improvement of the quality of the restored image. We decompose at level 5 with the *db1* wavelet an image whose size is 128×128. It comprises 255 shades of gray and its storage therefore requires 16,384 bytes. The EZW algorithm uses 4 symbols for the principal stage and 2 for the secondary stage.

Let us briefly comment on the results obtained. The images in the first row of the figure are of bad quality and unusable. The sixth and especially the seventh image already make it possible to distinguish the principal zones of the original image. They are sufficient to distinguish a seated woman. Edges, nevertheless, remain badly defined. The quality of the reconstruction then gradually improves and the image obtained after nine iterations of the algorithm is almost perfect, at least visually. In order to considerably reduce information to be stored or transmitted, we still have to efficiently code all the symbols produced.

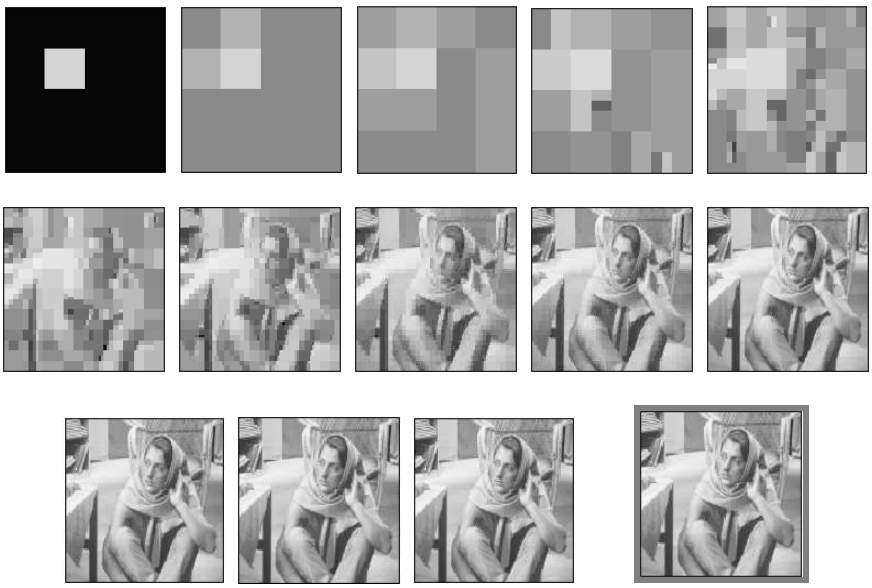


Figure 8.46. *EZW algorithm: at the top, results after 1, 2, 3, 4 and 5 iterations; in the middle, results after 6, 7, 8, 9 and 10 iterations; at the bottom on the left, results after 11, 12 and 13 iterations, and at the bottom on the right the original image*

8.6.3.3. *Comments on the JPEG 2000 standard*

In addition to operations linked to wavelets (DWT and IDWT) and the aspects associated with coding and quantization, the JPEG 2000 method comprises many technical subtleties. It is the effectiveness of the system on the whole which led to JPEG 2000 becoming the standard for image compression (see [SKO 01] and [JPE 00]); here we will only provide some indications.

The JPEG 2000 system was constructed to satisfy three major objectives. First of all, to offer a certain universality: for a broad family of color images to achieve better performances than existing standards for strong compression ratios, without degrading them for lower ratios. Then, to enable a progressive transmission of compressed images with increasing quality until a lossless compression is achieved and to make it possible to define specific areas that have to be coded with greater precision. Lastly, to ensure good data security and robustness with respect to errors during transmission.

The complete procedure to code a color image thus includes several operations; let us cite in particular:

– original image decomposition in color components;

– cutting (optional) of the components in equal rectangular zones;

– normalization of the coefficients;

– application of the DWT to each rectangular zone;

– quantization of the decomposition coefficients for each rectangular zone;

– coding of coefficients (coding can differ according to zones in order to improve quality).

For example, in order to illustrate the second point, the image of the porch (512×512) can be cut up into 16 square zones of size (64×64), as in the left part of Figure 8.47. The choice of cutting can influence compression performances. On the right, we find level 2 analyses of each "zone" (the details are displayed in negative). The EZW algorithm is then applied to each of these decompositions.

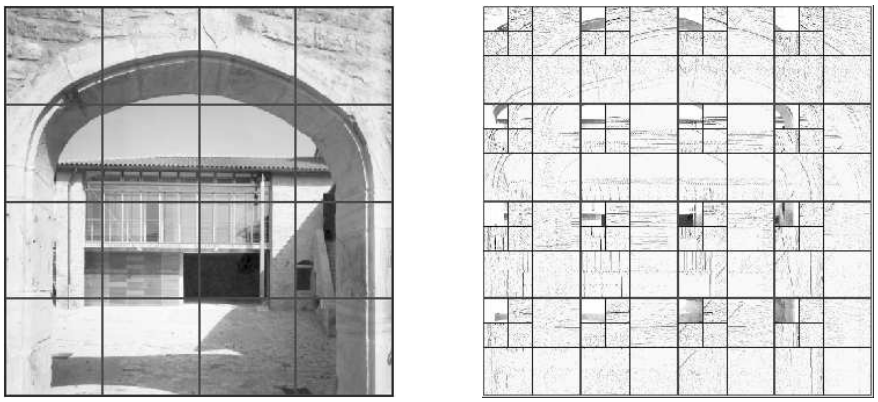


Figure 8.47. *Cutting up the porch image (512 × 512) into 16 square zones of size (64 × 64), and level 2 analyses of each "zone" (the details are displayed in negative)*

Chapter 9

An Overview of Applications

9.1. Introduction

9.1.1. *Why does it work?*

As this chapter shows, wavelets are used in a large number of fields. Let us just mention geophysics, astrophysics, quality control, biology and aural signals in medicine, imagery in all its aspects and medical imagery in particular, compressed representation of fingerprints or photographs, satellite imagery, coding of video signals, modeling of traffic in communication networks like the Internet and analysis of atmospheric or wind tunnel turbulence. Even after years of work we remain surprised by the variety of domains concerned and the problems dealt with.

A question arises from the start: how is it possible that the same tool works for so many applications? Of course, it is impossible to find a single really satisfactory answer, but several reasons appear plausible. In his book of historical presentation of wavelets, Hubbard (see [HUB 95] p. 139-173) proposes several clues on very diverse levels:

– first of all, the wavelets method is new in signal processing. It brings technical innovations and makes it possible to look at traditional information through original easily accessible tools. A new dictionary of shapes is created, connecting signal characteristics to those of wavelet transforms, enabling us to infer properties of the signals based on coefficient structures. There are, for example, visible cones in continuous decomposition, which indicate a serious signal irregularity. It is also known that zero coefficients in a discrete decomposition indicate that the signal is

smooth, that nothing changes. Fractal signals have recognizable shapes in terms of wavelet coefficients;

– the wavelets technique, moreover, constitutes a tool for local analysis, where it differs strongly from a Fourier analysis. The wavelet can focus its glance locally by inspecting the neighborhoods of a point. Then the information coded in the coefficients is entirely determined by the values of the signal close to the wavelet support. Such work based on using local zones can be found in every local technique. It is supplemented by a synthesizing work which carries out the comparison of the results obtained in various points. To some extent we globalize local analyses while locating, for example, zones of similar local behavior, thus determining various interesting zones;

– as often is the case in signal processing, little information is useful *a priori*, for analysis as well as for complex processing like denoising. This advantage which renders the model light under which processing is adapted, ensures adequate processing of large signal families. Of course, from the point of view of modeling, hypotheses on the law of noises (Gaussian hypotheses) or on the nature of the searched signal (Bayesian hypotheses) often facilitate the work. Naturally, if specific knowledge is available, there is often a more adequate technique, which is less universal and more effective than the wavelets;

– wavelets constitute bases, in the vectorial sense of the term, spaces where signals “live”. They represent all the signals of these sets. These bases are often orthogonal. The role of this characteristic should not be underestimated: by allowing an economic representation it particularly ensures the inversibility. In other words, it is not only possible to reconstruct a signal from co-ordinates, but the reconstruction also works in another way. Let us imagine that we are authorized to transform the coefficients, for example by removing some well selected ones, enough so that the effect is visible but not so much as to destroy the principal contents of the message. Then the inverse transformation reconstructs a signal which is often more comprehensible than the original signal. It is thus possible to act on the coefficients. For wavelets operating very locally, the coefficients depending on a small zone of the signal, the reconstructed modification affects and transforms only a small zone of the synthesized signal;

– wavelets analyze the signal scale by scale. They are used, as many authors say, as a microscope and mathematical and numerical zoom. They look at the signal with various resolutions, as in geography maps with different levels of detail. Each point is sounded, very finely on a small scale in a small zone, then less finely on an average scale and, finally, coarsely on a large scale. The neighborhoods observed have variable sizes, a whole continuum of sizes is treated by tools whose resolution covers a large band. The same moment is simultaneously observed from several points of view, on different scales;

– wavelets make it possible to characterize signal spaces which have been somewhat ignored before because they are difficult to handle, in particular Besov spaces. They contain functions which can present parts less regular than the derivable functions, separated by discontinuities. These signals are difficult to tackle with other currently available tools.

9.1.2. *A classification of the applications*

It is never easy to classify practices or applications because they often simultaneously use several aspects of the technique. Nevertheless, we propose a ventilation "directed by wavelets".

The first classification divides the applications into three fields. It separates the applications where the scale aspects dominate from those in which specific aspects dealing with time or space are prevalent and, finally, applications using the possibilities of representation offered by the wavelets. The tables which follow refer to the applications presented in the remainder of the chapter or to others found in other works.

	Scale aspects for signals and images	**Time or space aspects**
Problems	Determination of trends. Calculation of approximations, smoothing.	Ruptures. Edges.
	Decomposition, superposition, separation.	Fast evolution: phenomena of short duration, transitory.
Fields	Modeling of traffic of communication networks: Internet.	Detection of pathological events: epileptic crisis, evoked potentials in EEG.
	Biomedical signals: mammography.	Intermittency in physics: pressure holes in a field.
		Industrial monitoring of gears for location of ruptures.
	Scale laws in physics: turbulence.	Non-destructive control: detection of dysfunctions in process control.
		Underwater signals.

Table 9.1. *Classification of the applications according to the scale, time or space aspects*

By using all the characteristics as a whole, wavelets are useful as a representation tool.

	On the whole
Problems	Simplification. Economic representation.
	Pattern recognition. Classification.
Fields	Denoising of biomedical signals and medical images.
	Video coding, coding of animated images.
	Compression of photographs, of video images.
	Classification of star spectra. Classification of eating behaviors.
	Detection of boats, helicopters by signature recognition. Detection of seismic jolts.
	Classification of acoustic signals.
	Numerical approximation of linear operators.

Table 9.2. *Classification of applications where time and scale aspects are regarded as a whole*

9.1.3. *Two problems in which the wavelets are competitive*

Denoising and compression are two problems solved well by wavelets and are, nonetheless, still subject to development. The techniques are detailed in Chapters 7 and 8. We only summarize some useful ideas in the application files, where they have not yet been included.

Denoising

The general spirit of modeling distinguishes the useful signal from the noise. Noise reduction, which consists of, according to the beautiful expression of V. Wickerhauser quoted in [HUB 95], "cut the weeds and spare the daisies", includes:

– suppression of noise to locate great, often slow, evolutions associated with the trend or smooth zones;

– search and modeling of non-smooth parts varying quickly but stable enough, characterized by flat spectra whose most usual representative is white noise, or by textures, zones exhibiting a repeating pattern;

– suppression of the irrelevant part to preserve the interpretable part.

We come across denoising in several of the applications indexed in this chapter.

Compression

Compression compacts information to preserve or transmit it. To some extent, it bears a greater resemblance to denoising, with which it shares an end of the technical processing chain. However, apart from noise reduction, it optimizes a criterion which is a direct function of the scarcity of the coefficients preserved – in the sense that it is a question of restoring the original object in the most accurate manner possible with some information given *a priori*.

9.1.4. *Presentation of applications*

In order to conclude this introduction in the rows of Table 9.3 we find the applications detailed in the rest of this chapter and, in the columns, the topics or the fields with respect to which they are introduced.

	Fields							
Applications	Physics	Health	Industry	Signal	Image	Statistics	Denoising	Compression
Wind gusts	✓			✓		✓		
Seismic jolts	✓			✓				
Bathymetry	✓				✓		✓	
Turbulence	✓			✓		✓	✓	
Coding ECG signals		✓		✓				✓
Eating behavior		✓		✓		✓		
Fractional wavelets and imagery NMR		✓			✓			
Wavelets and biomedical sciences		✓			✓			
Transitories underwater signals			✓	✓		✓		
Statistical control of processes (SCP)			✓	✓		✓	✓	
Compression of industrial information			✓	✓		✓		✓

Table 9.3. *Distribution by field of the applications presented*

Many works assemble wavelet applications in very varied fields; references to them appear at the end of this chapter.

We have preferred, rather than to summarize these books, to which the reader may refer, to analyze other sources, according to the following framework. Each application is centered around a reference cited at the beginning of the corresponding section and presented by its object, its data, the techniques used, the results and the role of the wavelets.

9.2. Wind gusts

Objective

In [LIU 94] the objective is to characterize ocean winds during their formation phase, their growth and the period of full development.

Data

Measurements of wind and wave heights were made during the large storm on October 26, 1990 at sea (Atlantic Ocean), on the basis of 10 recordings a second, producing 100 sets of 1,024 measurement points and covering the total duration of the storm. The intensities of winds range between calm and strong winds of 18 m/s (64.8 km/h). The waves reach 7 meters in height.

Technique

The spectrum, cospectrum, coherence and quadrature are usually defined and studied by spectral analyses. These concepts adapt to the wavelets framework by replacing the base of the complex exponentials intervening in Fourier with wavelets $\psi_{a,b}$. The energy of the decomposed signal is then analyzed according to time and scale. For one-dimensional signals the analysis is traditional. For signals with two or more dimensions processed simultaneously, the study becomes multivariate and then more specific.

If X and Y are time-series and C_X and C_Y are the corresponding continuous decompositions, then for a wavelet we define:

– wavelet spectrum:

$$W_X(a,b) = \left|C_X(a,b))\right|^2$$

– wavelet cross spectrum:

$$W_{XY}(a,b) = C_X(a,b)\overline{C_Y(a,b)}$$

– coherence:

$$Coher\,(a,b) = \frac{W_{XY}(a,b)}{\sqrt{W_X(a,b)W_Y(a,b)}}$$

– quadrature spectrum:

$$\left|Coher\ (a,b)\right|^2$$

The Morlet "wavelet" seen like $\psi(t) = e^{imt}e^{-\frac{t^2}{2}}$ is not an admissible wavelet because $\hat{\psi}(0) \neq 0$, but for large enough m this condition is almost satisfied.

Results

The wavelets are used for three goals: detection of group waves effects, analysis of the development of waves due to the wind and breaks in the waves:

– Large waves arrive in sequences, making the underlying process non-stationary. To determine their length we count the number of waves whose size exceeds a preset level. From the wavelet spectrum (resulting from a continuous analysis) of the sea level, coded in frequency rather than in scale, we deduce the curves of isoenergy. High energies appear with the help of the contour lines grouped in homogenous zones, delimiting the time periods of the various groups and the bands of frequencies covered (see Figure 9.1). By synthesizing the durations, the frequencies and the number of waves, we find that a linear relation links the duration (normalized) to the energy (also normalized) within the groups of waves.

– Crossed spectra, when taken as a whole, are difficult to interpret. However, the analysis of some frequencies and of the phase of the coherence function makes it possible to draw conclusions. We find that, although the growth of waves is very complex, the strong energy components of wind and waves are in phase. This result clashes with the working theory that connects wind average to waves. Energetic waves appear in groups which constitute a basic element of understanding the wave process. Moreover, the phenomenon acquires power mainly in these regroupings.

– Wave breaks are a central phenomenon in the exchange of gas between the ocean and the atmosphere and the exchange of moment between the wind and surface. Usually we fix a limiting value on the slope of the signal associated with the wave beyond which the wave breaks. In simpler terms, we can consider that the rupture occurs when acceleration towards the bottom exceeds a preset proportion of gravity driven acceleration. This approach can be transcribed in the context of wavelets. Heuristic ideas, very technical, make it possible to locate the moments of "wave rupture" on the signal. They show a relation that is closely affine between the proportion of the breaks and wind speed, corroborating existing results.

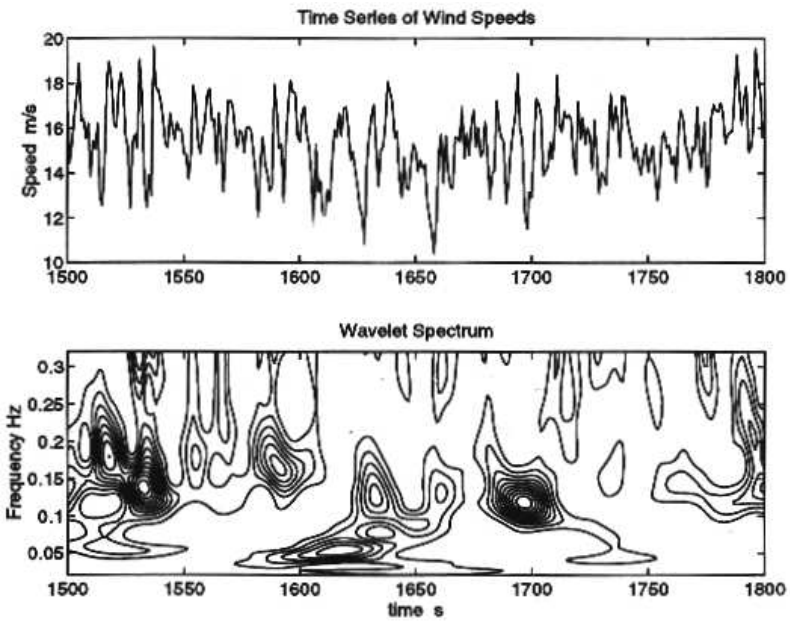


Figure 9.1. *(Extracted from [LIU 94] p. 159): wind speeds are represented in the upper part of the figure. The lower part shows the wavelet spectrum of the signal using the Morlet wavelet. Time is measured along the x-axis and frequencies are located along the y-axis. The energy is constant for each contour line. These lines delimit the time and frequency zones of strongly energetic wind*

Role of the wavelets

As usual, wavelets analyze the local distribution of the energy of a scalar signal. They also compare time shifts and phase shifts, which appear in a couple of signals. They supplement the possibilities offered by the cospectra of the Fourier analysis.

9.3. Detection of seismic jolts

Objective

In [ANA 97] the issue is to automatically identify various components, usually called P and S, of a seismic signal, short and non-stationary, using wavelets as a tool of pattern recognition.

The seismogram records motion in three directions: two in the horizontal and one vertical. Usually, the location is based on several ground detectors. We may,

however, have to use just one sensor, especially for events concerning small regional events.

The significant features are retained over several levels. The analysis deals with a tri-varied recording. The detector locates the part P by using information known as polarization through the scales. The analysis of the part S, which is a shear wave, is based on the relationship between transverse and radial amplitudes seen across the scales.

The seismic signals are modeled as the convolution of filters representing the medium and fundamental sources, which represent vibrations of the ground due to a tremor or an explosion.

Data

The available period covers 5 years and starts in 1990. The measuring site is located in Spain and has recorded 23 seismic events within a radius of slightly more than 2,500 km.

Technique

Several wavelets resemble an interesting signal, either in its part P, or in its part S. It is the case of the *db12* wavelet, of some biorthogonals and of wavelets proposed by Vetterli.

The determination of the input of P and then of S is done in two stages:

– *stage 1*. Measurements are three-dimensional. The arrival of the compression wave P produces a "transition in the rectilinearity function C_F", defined later on. The signal covariance matrix at the moment t is a 3×3 matrix noted $M(t)$. A measurement of linearity consists of $F(t) = 1 - \dfrac{\lambda_2(t)}{\lambda_1(t)}$, where λ indicates the ordered eigenvalues of $M(t)$ starting with the largest of them: λ_1. Linearity corresponds to $\lambda_2 = 0$, that is, $F = 1$. M is estimated in various time bands. The function F is calculated as well for the raw data as for the coefficients of each scale j of the wavelet decomposition. Scales 1 and 2 primarily contain noise and after scale 8, F is practically equal to 1. The functions F^j for $3 \leq j \leq 8$ are grouped into a synthetic inter-scale linearity function $C_F = \prod_j F^j$, whose maximum locates the arrival of P;

– *stage 2*. Identification of the phase S. The arrival of the wave S is detected by a function of the coefficients of wavelet decomposition on 10 scales of radial and

transverse directions. This function is calculated after a change of reference mark in the plane aligning one of the axes with the direction of the source of the wave S. Several wavelets have been used;

The results obtained are the following:

– *results for the P wave.* Algorithmic detection is compared with an analyst and a technique based on a short term/long term ratio. The algorithm is effective when compared to the analysis, the errors of the two approaches not being deeply different. It is better than the short term/long term ratio, in particular, when the energy of the signal is weak (revealing case of algorithm quality);

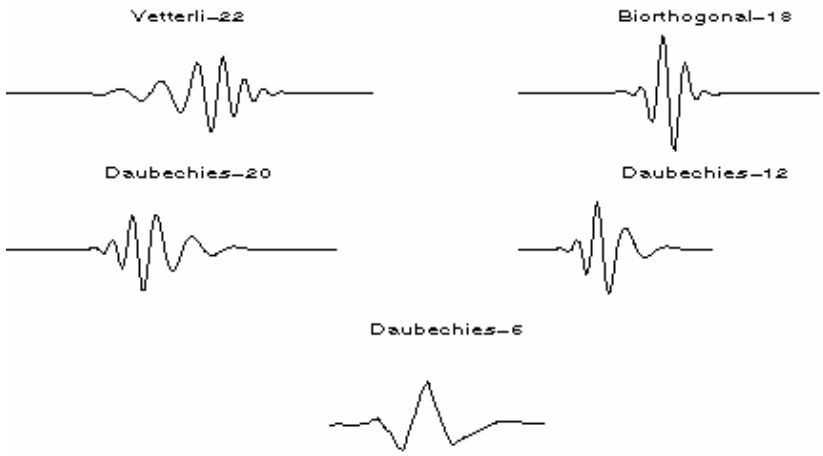


Figure 9.2. *Wavelets resembling the interesting signal*

– *results for the wave S.* The detection of the wave S is more difficult than that of the wave P and the results are worse. Nevertheless, they seem better than those of an expert. The result depends on the wavelet and the useful one resembles the most to the signal being detected. The winning wavelets are those in Figure 9.2, the Vetterli-22 gaining 9 times out of 23. The resemblance to the signal is visible in Figure 9.3.

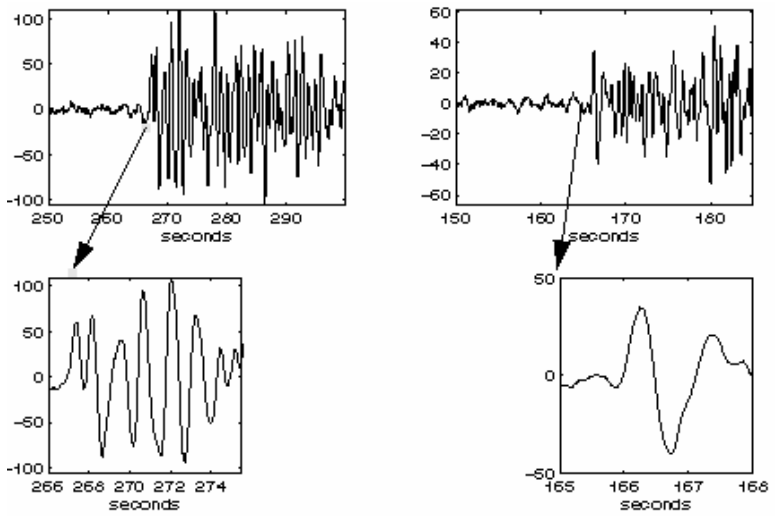


Figure 9.3. *(Extracted from [ANA 97]): two shapes of seismic signals and zooms around the event. We can compare the shapes of the start of the signals and the shapes of the wavelets in Figure 9.2 and note that they correspond well*

9.4. Bathymetric study of the marine floor

Objective

In [LIT 94] the issue is to chart marine floor areas in order to locate shapes and, in particular, to locate the strong escarpment declivities. The zone covered is in the North Atlantic and shows a central valley oriented S/E-N/W, zone of the tectonic plates separation, framed by escarpments. It measures 100 km × 70 km. The depths vary between 1,800 m and 4,000 m.

Data

An "echo sounder" mounted on a boat traversing the zone to be covered collects the signal. The typical image has as a size of 755 points along the longitude direction by 495 along the latitude direction.

Technique

It is a technique of image improvement and pattern detection of shapes by using a wavelet based on B-splines tensored with a filter. The function is rotated in the plane so that its principal direction will be that of the required edges. The derivative of the cubic B-spline provides another function, which locates the transitions between valleys and escarpments.

Results

The required shape is a long, rectilinear, narrow edge whose dominant orientation is parallel to the axis of the expansion of the earth's crust. To locate this shape, the wavelet acts mainly as an edge detector. Figure 9.4 shows the results. The required lines are parallel to the central valley. They are difficult to locate on the basic decompositions and are discovered much better through the visualization of local minima and maxima (by traversing the image from the East to the West).

Role of the wavelets

"The wavelets decomposition of bathymetric data reveals structures and patterns which are easily overlooked in the raw data". The wavelet produces: "a qualitative image enhancement and a qualitative fault scarp identifier". "Wavelet analysis offers a useful method for decomposing the texture of the seafloor to help understand processes which occur at many different scales" (quotations extracted from [LIT 94], p.167 for the first and p. 180 for the two following ones).

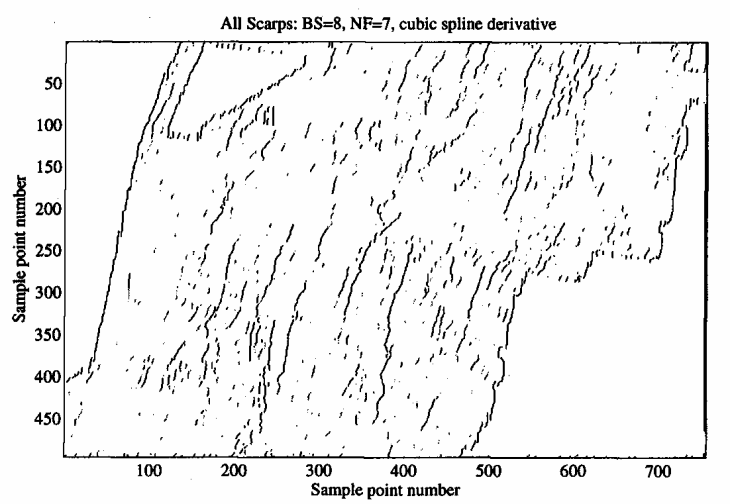


Figure 9.4. *(Extracted from [LIT 94]): detection of edges on the bathymetric image of the marine floor using a wavelet based on B-splines*

9.5. Turbulence analysis

Objective

The study of Papanicolaou and Solna [PAP 01] deals with the famous Kolmogorov turbulence law, which links energy to scale. It provides a

phenomenological description of the velocity field in the atmosphere and is expressed by a power law. The average square of the increments of the velocity V_x, for a large range of scales, is modeled by $E\left[\left(V_x\left(x_0+u\right)-V_x\left(x_0\right)\right)^2\right]=C_v^2\,|u|^{2/3}$.

The important thing in this formula is its power form and the order of magnitude of the coefficient (here $2/3$).

The two basic ingredients of this type of study are:

– on the one hand, the model of the process called fractional Brownian motion and noted MBF_H (which is the French abbreviation) which depends on a parameter H (or one of its extensions which allows the variation of H);

– on the other hand, the wavelet spectrum, which generalizes the usual spectrum based on the Fourier transform.

The wavelets harmoniously introduce the decompositions according to the scales involved in various estimation techniques.

Let us say a word on the fractional Brownian motion MBF_H. It is a continuous-time process, which resembles a usual Brownian motion whose derivative is a white noise. It is Gaussian, self-similar in distribution and the variance of the increments depends on H in the following way:

$$Var\left(MBF_H(t)-MBF_H(s)\right)=\sigma^2\,|t-s|^{2H}.$$

The extension of MBF_H to the $MBF_{H(t)}$ or $MBF_{H(\lambda)}$, making it possible to make H depend on time or frequency, is important since situations compatible with models with a constant parameter seem rare. These new objects are used, for example, in the modeling of vertical turbulence. Additional information can be found, for example, in [BOS 00] and [DOU 03].

Data

The data comes from a lidar emitting a laser beam, delivering measurements from which atmospheric temperatures at various altitudes are deduced. The recordings are long and include n = 4,200,000 vertically spaced points of 2 cm. They cover a vertical more than 80 km long.

Technique

The very first methods of decompositions of DWT type using around 20 dyadic scales show that any modeling, which would provide a power law, should allow the parameter representing power to vary with altitude. The model is then local.

The work consists of finding a partition of the vertical line (altitude) and calculating the corresponding estimates of H. The technique is as follows:

– choice of the bands of values where the parameter can be regarded as constant;

– estimation of the values of H, parameter which varies with the segment.

Results

We know that, according to the value of H, the memory of the process is short or long. The results locate $H \simeq \frac{1}{3}$ in the zone of the short memory values, corresponding to a theoretical spectrum of increments of the shape $|\lambda|^{-\frac{5}{3}}$ in the neighborhood of the 0 frequency.

An important part of statistical work consists of:

– determining segments in which H can be regarded as constant, segments which, in order to ensure a precise estimate, must be as large as possible;

– verifying that the estimates do not depend too much on the partitioning. The vertical is partitioned into segments of 655 m, a value determined following many tests. Energy analyses make it possible to notice that low altitudes present lower energy than the high ones [PAP 01].

The statistical technique is based on several ideas, including the one that connects in one MBF_H the variance of the wavelet coefficients at level j to j itself. The estimate $\hat{H}$ of H is obtained by regressing the logarithm of this variance over the logarithm of j. The study is a little more complicated and a modeling of H using a random walk is introduced.

The only wavelet used in this study is the Haar wavelet. The scaling function φ does not appear. In fact, only the multi-resolution aspect is useful, at the expense of other wavelet properties. The wavelet simultaneously carries out the calculation of the increments of the measurements and the low frequencies filtering.

Percival and Guttorp [PER 94] study a similar problem concerning vertical shearing forces in the ocean. The data are collected every 0.1 m, differentiated and pre-filtered in order to preserve the non-noised evolutions. A reasonable model

postulates the evolution into power law $S(\lambda) = |\lambda|^{\alpha}$ of energy according to depth, at least for a rather broad range of scales. The model is estimated by the use of SWT wavelet decomposition, needed for the calculation of the variance of the wavelet coefficients for each scale level. This estimator is linked to the Allen variance, famous in the non-stationary signal processing.

These studies are useful in the modeling of turbulence because they specify the link between interesting and difficult to model data, the local power law, Brownian modeling and wavelets.

9.6. Electrocardiogram (ECG): coding and moment of the maximum

Objective

In [ANA 95] the issue is the adequate coding of the wavelet analyses of long signals. This problem arises for biomedical signals for the purposes of storage and transmission. The compression ratio must be increased while maintaining the distortion at an acceptable level in order to allow a clinical examination. The essential shapes of the cardiac electric signal are the P shapes, the QRS complex and the T shape (see Table 9.4).

There are classical coding techniques. They are based on "DPCM (scalar and vector coding), entropy coding, Fourier or Walsh transformations" [ANA 95].

An ECG is a short and quickly variable signal and is, thus, a good candidate for wavelet analysis. The amplitude, duration and rhythm enable the identification of the anomalies. The operation of the cardiac muscle is described on the basis of the elements of Table 9.4.

Characteristics of the wave	Activity
P	Electric activation of Atria
QRS	Electric activation of the ventricles
P-QRS Time	Duration of Atria-Ventricle electric conduction
T	Repolarization of the ventricles

Table 9.4. *The essential shapes of the cardiac electric signal*

Data

100 ECGs are analyzed by three wavelets: the *db12*, a Vetterli-Herley wavelet and a biorthogonal wavelet.

Technique

Scalar and vectorial codings are used on different scale bands. The former for small scales with signals changing quickly; the latter for long and slowly changing signals. The wavelet decomposition goes up to level 5.

The coding dictionary is built around the sequences $a^5, d^5, ..., d^1$ by the algorithm of average points (k-means) and the average points are the codes. The number of learning vectors is large and depends on the scale.

Results and advantages of wavelets

The gain is significant compared to the usual methods. The two evaluation criteria are the bit per sample ratio and a quadratic error of reconstruction. The Vetterli-Herley wavelet produces the best results.

Supplements to an older work

The study [ANA 94] proposes an estimate of the moment of the maximum of the wave P using wavelets. We start with many cardiac cycles, which are signals of length 316 points. We construct a study data set by associating to each cycle 4 new signals constructed by moving the P wave and the QRS complex at random.

Decomposition uses a quadratic spline wavelet with compact support and wavelet coefficients of a single scale constitute the data input of the research procedure for the moment of maximum. A neural network determines the required moment. We compare the rough data processing with that of the wavelet coefficients. The qualities are similar, the wavelet analysis is used here as a preprocessor for the main treatment and saves the quantity of processed information. From this point of view the wavelet is useful.

9.7. Eating behavior

Objective

Tate [TAT 95] and [TAT 96] studies the influence of the type of fat ingested by an individual on the fat stored by the body. The basic measurement is a magnetic resonance spectrum providing information on the chemical composition of fat. The peaks of the spectrum represent nuclei in various molecular sites, resonating at

slightly different frequencies. The surface located under each peak estimates the quantity of the substance associated with the peak. *In vitro*, the peaks are pointed. *In vivo*, the identification of the peaks is much more difficult, so much so that it is desirable to avoid it.

Data

The database is a set of spectra ^{13}C of subcutaneous fat taken from patients' thighs. The volunteers are either vegans ($n = 33$), vegetarians ($n = 8$), or meat eaters ($n = 34$). A pre-processing realigns the spectra based on the largest peak; 512 points are preserved and the total energy is normalized to 1.

Technique

Each spectrum is broken up into its wavelet coefficients by the *db20*, recoded using the first 64 coefficients. A discriminant factorial analysis classifies the set of the 75 spectra without it being necessary to measure or to identify the peaks of the spectra.

Results

A jackknife technique makes it possible to select the best discriminant factorial components, thereafter used in the discrimination stage. The properties of good and bad classifications are indexed in Table 9.5.

	Good classification	**Bad classification**	**Rate of good classification**
Vegetarians excluded	63	4	94%
Vegetarians included	60	15	80%

Table 9.5. *Discriminant results obtained using wavelet-based classification*

These results improve the previous studies using spectrum peaks' energies, in which 20 cases were classified badly when the vegetarians are included and 5 when they are excluded (see [TAT 96]).

Advantage of wavelets

Although working blind, without information on the signal to be analyzed, wavelets select information by strongly reducing its quantity. The later statistical stages are simplified and the quality of classification improved.

9.8. Fractional wavelets and fMRI

Objective

The purpose of the article [FEI 00] is to discuss the determination, in a rather broad family, of the wavelet to choose in order to ensure a good analysis of the fMRI (functional magnetic resonance imaging) data. Optimization is done for known activation patterns.

Functional imagery highlights the cortical zones of a patient, activated by external stimulation or during a cognitive or motor task.

Data

The research in the large majority of works is structured by an experimental design whose factors are the subjects and the working conditions. Let us note $I_{A_\tau}(x,y)$ and $I_{B_\tau}(x,y)$ a sequence of images indexed by the time τ for the two conditions A (activation) and B (rest), and pose:

$$I_{D_\tau}(x,y) = I_{A_\tau}(x,y) - I_{B_\tau}(x,y)$$

Prior to any treatment, the data are realigned; a motion correction is enforced and drifts in time are removed.

Technique

The family of wavelets considered, called fractional splines, generalizes the B-splines and was introduced by Unser. It is constructed on the basis of power functions $\left|x-n\right|_+^\alpha$ where $x_+ = \max(x,0)$ and exists in causal and non-causal symmetrical, biorthogonal or orthogonal shapes. The associated filters do not have a finite impulse response but are, nevertheless, simple. The design parameter used for optimization is α.

To detect an activation we calculate the wavelet coefficients I_{W_τ} of I_{D_τ}. A statistical test on the coefficients makes it possible to reduce noise and then to evaluate the significance of activation.

In an fMRI the traditional technique is called SPM (statistical parameter mapping). It filters the images before the statistical test. The pre-filtering complicates the situation by making the noises of various pixels of the resulting image dependent.

Results and role of the wavelets

The zero hypothesis of the statistical activation test concerning the pixel (x, y) is $E\left(I_W(x,y)\right) = 0$. The test is based on a Student's t distribution, $t = \frac{I_{W(x,y)}}{s(x,y)}$, where $I_W(x,y) = \frac{1}{n}\sum_{\tau=1}^{n} I_{W_\tau}(x,y)$ and $s(x,y)$ is the standard deviation of $\left(I_{W_\tau}(x,y)\right)_{\tau=1,\cdots,n}$. As the test bears on each pixel, the fine-tuning of the critical value $t_{threshold}$ is done, as usual, using conservative techniques.

Optimization proves to be important. The selected criteria differ for simulated and real data. In both cases the gain is large between an optimized choice and a merely reasonable choice.

The new wavelets are better than the tested Daubechies wavelets. We also find that:

– causal wavelets are better than symmetrical ones, the former presenting a better spatial localization;

– biorthogonal and orthogonal wavelets are better than the Daubechies wavelets, at least when the noises are large;

– the optimal degree α differs from one situation to another. It is close to 1 in the following expression of the biorthogonal filter of the two scales relation for the scaling function $H(z) = \sqrt{2}\left(\frac{1+z^{-1}}{2}\right)^{\alpha+1}$. To implement the test it is rare that the wavelet decomposition requires more than two levels.

Detection is considered to be, on the whole, effective and robust with respect to the different sizes of the activated zone and to the noise. "Because of its multi-resolution nature, our method is especially suited for detecting shapes with various frequency contents, such as mixtures of large and tiny activation zones" [FEI 00].

9.9. Wavelets and biomedical sciences

Objective

We focus here on a very interesting synthesis paper [UNS 96] which, although published in 1996, still gives a broad, very clear and relevant review of the applications of wavelets in biomedical sciences.

The objective is thus to analyze the place of wavelet processing in the biological and medical domains. The problems of the biomedical sector are, in general, badly posed and require the development of solutions that are more robust than optimal. Wavelets enter a process aiming to discover a dysfunction, comparing the normal and the abnormal, and improving the collected information to ensure a clearer job.

Data

The biomedical sector is characterized by the large variability of available signals. The recordings are often complex: mixtures of signals localized as spikes in EEG and diffused signals, like the cardiac background noise or the blood flow.

Technique

The first partition of applications separates the work on biomedical signal from image processing. 1D applications cover acoustics, ECG and EEG. 2D applications complete *ad hoc* work, such as tomography, NMR or functional imagery as well as rather traditional work, such as noise reduction, improvement of image contrast or detection of malformations (mammography).

All the types of wavelet analyses find an application: continuous or discrete, redundant (continuous or in frame) or economic (orthogonal, bi- or semi-orthogonal wavelets) or wavelet packages analysis.

Results and role of the wavelets

Each characteristic of the wavelets is associated to one or more uses. Let us cite some of them:

– as filter banks wavelets separate the frequency bands and evaluate the distribution of energies to recognize a distribution. It is the case of the analysis of the turbulent heart sounds in order to identify coronary artery disease;

– as filter banks wavelets are used for denoising and improvement of image contrast;

– as bases, tools of representation in which each coefficient counts, wavelets are used as an economic coding tool, since nothing is in excess and that there is enough information to describe exactly what is important. In contrast, compression admits that any description of an object is partially redundant or useless; it determines the important information and then codes it. The wavelet is then useful, in this context, to build the hierarchy of the information in order of importance;

– image processing is also based on the classical representation, introduced by Mallat, of signals by the lines of maximum of the continuous analysis;

– the analysis by level of scale provides a technique of detection of shape and, thus, constitutes a zooming tool (see Chapter 5). Here is the reason: if a signal in an unknown t_0 point and on an unknown a_0 scale resembles the shape of the wavelet $\psi\left(\frac{t-t_0}{a_0}\right)$, the associated wavelet coefficient is large and makes it possible to estimate t_0 and a_0. This property is used to study spikes in the EEG of epileptic patients as well as in the detection of complex QRS from the ECG;

– the wavelets having zero spectra in the vicinity of the 0 frequency and even "very" zero (the zero is of an order higher than one) make it possible to make the colored noises whiter. By filtering the noise by convolution, the wavelet seen as a filter multiplies the signal spectrum by that of the filter and cancels the energy at low frequencies. Thus, filtering renders nearly independent the noises with a long memory whose spectrum $f(\lambda) \sim \lambda^{-(2H+1)}$ is infinite, with $\frac{1}{2} < H < 1$ in the vicinity of $\lambda = 0$;

– wavelets ensure a simultaneous time-scale analysis and compete with time-frequency analysis tools, which are the short-time Fourier transform, also called the Gabor transform (see Chapter 2), or the Wigner-Ville analyses.

9.9.1. *Analysis of 1D biomedical signals*

9.9.1.1. *Bioacoustic signals*

The analysis of aural signals emitted by the heart, classified into "sounds and murmurs" constitutes a basic monitoring technique. The wavelets supplement the usual time-frequency techniques, separating components that are difficult to dissociate (like "aortic and pulmonary valve components" [UNS 96] p. 630). The more diffuse sounds are modified by vasodilator drugs and the changes are appraisable by wavelets, thus enabling the experiments on the classification of subjects suffering from coronary disease.

9.9.1.2. *Electrocardiogram (ECG)*

The cardiac signal breaks up into sequences, the most important of which is called the QRS complex. The location of shape by wavelets makes it possible to discriminate between normal and abnormal patterns. Wavelets tested on a usual database prove to have an excellent rate of detection. Although the other methods could not locate them, the wavelets clearly detect small signals with rather high frequency appearing in the coronary diseases, infarctions or ventricular arrhythmias.

9.9.1.3. *Electroencephalogram (EEG)*

Wavelets have proven reliable in the location of transitory characteristic signals mixed with normal signals. As the seizure increases, "the transient activity slowly develops into more nearly regular high amplitude quasi-periodic oscillations" [UNS 96]. Very different depending on the patients, complex signals remain quite analyzable. Studies for neurosurgery with implanted electrodes developed fast algorithms for continuous analysis. When the electrodes are on the scalp, the background electromyogenic noise interferes with the signal. The wavelets have also been used in the works of discrimination on the functioning of the fetus's brain.

The action of an external stimulation to the patient is visible on the EEG, whether this stimulation is acoustic, visual or somatic. The noises are averaged across a great number of repetitions, typically a few hundred. Stimulation is locatable after a time of latency. The evaluation of this time, based on the distribution of the coefficients energy, enables a reliable estimate of the time evolution of a neurological attack such as cerebral anoxia.

9.9.2. *2D biomedical signal analysis*

The most traditional application in biomedical signal analysis is noise reduction. Another is to increase image contrasts. In mammography, where a very small change of fabric structure can reveal the presence of a tumor or a micro calcification, wavelets are effective to detect such events.

9.9.2.1. *Nuclear magnetic resonance (NMR)*

Subjected to an external magnetic field of controlled frequency, the spins of the nuclear cores respond and increase their energy. The signal records the emissions coming from the return of the spins to the realignment position. The recording is done in a prescribed portion of space made up, for example, by a section. The measurement of the space density of the resonating spins provides information on the state of the excited medium.

An integral formula that has to be inversed connects the measured variable to the space density. Usually the inversion is done in the Fourier base. The wavelets enable faster calculations and reduce the artifacts ascribable to the movements of the patients.

In tomography the fMRI uses the multi-resolution aspects of the wavelets. "The information at different scales is updated at different rates; in particular, low frequency components can be reconstructed almost instantaneously and used to

estimate the motion of the object" [UNS 96], an estimate which, in turn, contributes to the determination of the fine structures.

9.9.2.2. *fMRI and functional imagery*

fMRI is a research tool in neurosciences. It studies "the neural activity of the brain *in vivo*" and records, for example, spatial distribution of radiotracers whose local changes are being followed. It is integrated in an experimental planning whose factors are the subjects and the conditions.

The images are extremely noisy and variable. A multi-resolution analysis based on splines ensures the representation of the images and volumes.

In order to distinguish activation from non-activation, the comparison is made pixel by pixel after control and compensation for the anatomical differences between subjects and the movements of each subject. Once realigned, the data is submitted to statistical tests carried out in the wavelet domain, i.e. on the coefficients, thus avoiding the noise and concentrating on the smoothest parts which are more meaningful for the image.

The advantages of wavelets over the more traditional methods are as follows:

– wavelets recode information in a small number of large value coefficients;

– the signal-to-noise ratio of these coefficients is good if we find ourselves in the rather common situation where the noise is about the same everywhere in the measurement space. A good signal-to-noise ratio ensures a good detection rate;

– finally, wavelets often have the capacity to decrease space dependences enabling the use of simple statistical tests.

9.10. Statistical process control

Objective

The statistical process control (SPC) is a set of methods, integrated into quality control, enabling the development of procedures for designing, controlling and monitoring industrial and manufacturing processes. They are based on the modeling of normal operation, called controlled (or under control) operation, and on the definition of situations of abnormal operation indicated by the term out of control. A monitoring statistic is calculated and controlled regularly during successive samplings. As long as it remains in the controlled zone, the following sampling is carried out. As soon as it enters the out of control zone, production is stopped and the search for causes which provoked the dysfunction is initiated. The monitoring

statistics most usually employed are Student's t or its multivariate version called Hotelling's T^2. Successive values are visualized on a chart exhibiting the index of the batch along the x-axis and the value of t or of T^2 along the y-axis.

When several simultaneous measurements are available, the problem is multivariate nature and the investigation into the causes becomes crucial and difficult. Despite their apparent complexity, the majority of multidimensional recordings depend only on a small number of dimensions: often two, sometimes three, seldom more, although the theoretical approach does not limit this number. The principal component analysis makes it possible to reduce the overall dimension by highlighting relevant combinations of the initial variables. It produces new variables, called principal factors, uncorrelated (whereas the initial variables are correlated). The calculation of the values of the monitoring statistics is then simple.

When measurements are signals recorded in time, the significant components are usually corrupted by white or even colored noise. Wavelets are used, first of all, for noise reduction, but also to locate the dysfunction signals, which are transitory signals hidden on only several scales and occurring at moments unknown *a priori*.

Hereafter we analyze two articles by Bakshi [BAK 98], [BAK 99].

Technique

For 1D signals, the basic technique consists of creating several control charts, one for each scale, i.e. for each level of wavelet decomposition. The threshold of each chart is calibrated in an adapted manner but fine-tuning remains difficult. The levels out of control are then located, by reconstructing the associated signal, and the state of the process and the causes of rupture of normal operation are determined.

For multidimensional signals the situation is as follows. Let us consider an example in which the time evolutions of the four process variables are as follows: temperature, pressure, water content and the quality of the product. The raw data is grouped in a $n \times 4$ matrix noted by $X = \left(X_i^j\right)$, with $1 \leq i \leq n$ and $1 \leq j \leq 4$, where i indexes time and j locates the variable.

The technique is as follows:

– decompose using wavelets on 3 levels; each X^j produces 3 details and an approximation denoted by D_1^j, D_2^j, D_3^j, A_3^j;

– locate the levels containing mainly noise;

– other levels (let us imagine that these are levels 2 and 3) are preserved. The three matrices including respectively the four details D_2^j, $j = 1,\ldots,4$, the four details D_3^j and the four approximations A_3^j are reconstructed;

– at each preserved level, perform a PCA. A fundamental dimension r is identified, which is able to be different across the scales, and only the r principal components are retained;

– apply the inverse wavelet transform that reconstructs a denoised signal capturing essential information;

– finally, the monitoring statistics is calculated using this last signal.

Results

Since the change from the initial basis of signal representation to an orthonormal wavelet basis is carried out very easily and since angles and distances are preserved, the calculations of APC can be performed well both before as well as after wavelet decomposition. One of the current orientations is based on the idea of carrying out thresholding on multi-scale decompositions of the principal components.

Advantages

Most of the SPC procedures include a significant part of know-how, built-in for example, in the often complicated preliminary transformations (*ad hoc* "preprocessing"). Wavelets avoid preprocessing, which constitutes a substantial advantage. Moreover, PCA decorrelates the various variables in space, while wavelets almost ensure a decorrelation in time.

9.11. Online compression of industrial information

Objective

The objective is "to characterize, control and predict properties of a product, such as its humidity, composition and the distribution of fiber in the case of this application" ([TRY 01] p. 311) and to show that wavelets allow a better statistical control of the production process, which is continuous here.

Data

The near infrared spectra (NIR) are measured online in the upper part of a conveyor belt filled with wood chips for a Swedish company.

The data noted by $\left(X\left(i,\lambda_j\right)\right)_{i=1,\ldots,1000}$ includes N = 1,000 spectra recorded successively at one minute intervals and rendered discrete over 900 frequencies ($j = 1,\ldots,900$) covering the band 400 – 2,000 *nm*. The time is indexed by *i*. The difficulties stem from the existence of large variations of moisture, temperature, movement of the carpet and the height of the chips.

Technique

The technique is quite traditional except for a small number of particular characteristics. The detection of the failure events is based on the following idea: the events related to an abrupt change of the machine or a "shutdown" (collapse of the process) appear in the small scales and are definitely distinguishable from the derivatives and periodic variations, which, of course, appear on the large scales. The technique is as follows.

The data are divided into two sets: the calibration and the prediction sets. The compression step is performed in two steps:

1) row use compression of the NIR spectra using wavelet transforms;

2) column use wavelet transform on the compressed first step result.

A PCA model is calculated on the second step result.

The failure detection is performed by comparing the local energy of the residuals to a reference.

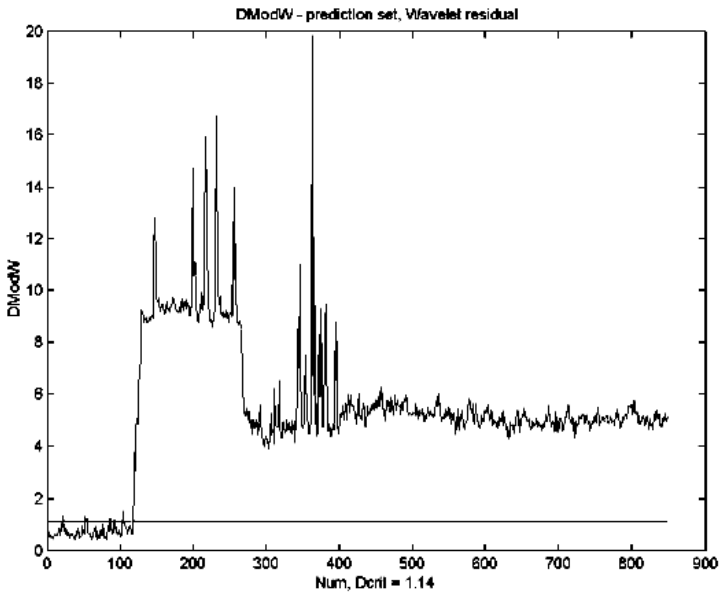


Figure 9.5. *(Extracted from [TRY 01]): variations of the detection statistics (noted DModW in the original graph) in the residual table of information for the entire prediction set. The residuals deviate strongly from the normal behavior*

Results

Figure 9.5 presents the variations of the DW_i statistics (noted *DModW* in the original graph) in the residual table of information for the entire prediction set. Around time 100 a large projection locates a rupture explained by the engineers and resulting from an empty carpet. The visible fast variations between moments 110 and 270 are ascribable to a plant shut-down resulting in an empty conveyor belt. The spectra are measured every minute and the graph visualizes the large differences. After instant 400 there seems to be a return to normality. Nevertheless, the statistics retains oscillations still revealing an abnormal variation. In fact a shift of the spectra continues to be present. This is due to a rise in temperature. A direct approach using statistics similar to DW_i, without wavelets and based directly on a PCA, had not been able to locate the spectral shift.

Role of the wavelets

The failures that we seek occur on different scales: they can be located in the form of details of the spectra or have much more global consequences. The wavelets are well adapted to make these distinctions.

9.12. Transitories in underwater signals

Objective

The objective of [BAI 98] is to detect transient signals against background underwater noise. "The interesting underwater signals are acoustic waves superimposed on an environment" (p. 73). Typically, they consist of a "whistle" or short-term clicks followed by several clicks. They resemble human words but present a greater interaction with the background. Using wavelets instead of Fourier-based techniques stems from the idea that wavelets provide a better time-frequency localization for transitories.

Data

The very long recording is partitioned into 0.8 second time blocks. Each block is sampled at 40.96 Khz producing 2^{15} = 32,768 measurement points.

Technique

A transitory signal appears as a point deviating with respect to the distribution of the non-transitory signal characteristic. A statistical test compares the energy of the block with an energy reference. Each block (indexed by t) is analyzed using wavelets over about 15 levels (we note the level j detail coefficients by d_k^j).

The average energy of the first 7 scales noted x_{tj} is the basic ingredient:

$$x_{tj} = \frac{1}{2^{(8-j)}} \sum_{k=1}^{2^{(8-j)}} d_{k(t)}^{j\ 2} \qquad j = 1,...,7$$

The time index t is linked to the specific indices at each scale, wherefrom stems the index $k(t)$ in the right-hand member of the preceding equality. Various adjustments show that only three or sometimes even the first two scales are enough: the useful vector appearing in the test does not use large scales corresponding to the low frequency parts of the signal. The useful part is: $X_t = \begin{bmatrix} x_{t3} & x_{t2} & x_{t1} \end{bmatrix}^T$.

The test statistic initially estimates the density f distribution of X_t in the absence of transitory phenomena and in the presence of the background noise. Calculation is carried out over a long record taken during a reference period comprising only background noise. The traditional core estimator is:

$$\hat{f}(X) / \frac{\det(S)^{-1/2}}{(2\pi)^{3/2} T h^3} \sum_{t=1}^{T} \exp\left(-\frac{(X - X_t)^T S^{-1} (X - X_t)}{2h^2} \right)$$

In this expression $X \in \mathbb{R}^3$ and S is a robust estimator of the covariance matrix of (X_t) and h is the width of the suitable bandwidth. We admit that the successive (X_t) $(1 \le t \le T)$ are time-independent. Even if this assumption is unreasonable for the basic observations, it is more reasonable for wavelet coefficients and, thus, for the (X_t) vectors. The detection rule is based on the comparison between the estimator $\hat{f}$ of f, calculated for all the observations noted x_j made between the moments 1 and T, and the estimator $\hat{f}_t$ obtained without using X_t. In fact, at every moment j we remove from x_j the observations which have been determined as deviating.

Results

Two signals are tested: one is synthetic and the other is real. The synthetic signal is constructed by adding a spoken word to a real signal recreated on the basis of marine background noise by an inverse short-time Fourier transform. The signal-to-noise ratio is an adjustment parameter. The technique based on wavelets using only two levels of decomposition is compared to a detection based on the short-time Fourier transform. We split the record in time windows. In each window, the frequencies are partitioned into classes, for which the energy distribution is

compared to the average on the entire signal. We decide that a transitory is present in a frequency band if the energy of the band exceeds a pre-calculated threshold. For 100 simulations the technique of wavelets detects a present transient signal well. However, it causes more false alarms than the short-time Fourier transform.

The real signal includes whistles and clicks. For the 1920 blocks, the technique detects 727 signals. A short-time Fourier transform identifies 578 signals in accordance with those detected by the wavelet approach. The difference is difficult to explain. Nevertheless, it seems that by analyzing the recordings again, by ear, wavelets are right more often than the short-time Fourier transform approach.

The technique of detection based on wavelets is new. It is, moreover, capable of adapting to slowly variable background noises.

9.13. Some applications at random

Let us, finally, mention some additional applications at random, less detailed than the previous ones.

9.13.1. *Video coding*

In [HEI 01] and [MAR 99] animated video images are coded by wavelets. On the one hand, a realignment technique makes it possible to reduce the time redundancies of a sequence of images and, on the other hand, an interpolation of texture improves the prediction. We de-correlate the wavelet prediction errors by using pre-coding procedures called PACC (partitioning, aggregation and conditional coding).

The technique is as follows: time realignment uses a bilinear procedure on a quadrangular grid based on a model of continuous motion. The following stages are distinguished:

- estimation and compensation of movement;
- wavelet representation and quantization;
- pre-coding (which is the original stage);
- arithmetic coding.

The results are definitely better than the H 263 techniques, in particular, for the very small rates where coders containing blocks present blocking effects.

9.13.2. *Computer-assisted tomography*

An x-ray is emitted at a body or an object to be analyzed. Its intensity measured at the exit point codes a balanced integral of the body's density. Using several rays it is possible to reconstruct the density and to distinguish healthy zones from sick ones. The Radon's integral transformation constitutes a basic tool of computer-assisted tomography, positron tomography and NMR (nuclear magnetic resonance).

Traditional tomography is global; the reconstruction in a point requires the knowledge of the values of the integrals on the straights located far from the point. Local methods use only local information. The approaches with wavelets fall under this category. The inversion of the transform can make use of a 2D wavelet base, for which the nullity of moments and localization constitute an advantage (see [WAL 02] p. 412 and [RAS 97]).

More generally, wavelets ensure fast calculation of integral operators also appearing in turbulent flows.

9.13.3. *Producing and analyzing irregular signals or images*

The irregularity of a signal manifests itself in several ways. A rupture constitutes a form of irregularity, whereas a fractal signal presents another form: it can be continuous but not derivable anywhere. Undoubtedly, a good approach is to think of the surface, covered by an aluminum paper sheet that has been ruffled and then unfolded. The surface presents very small asperities, in a great number. It is not derivable anywhere. We can construct irregularity indicators.

Before inventing fractals, Mandelbrot introduced and popularized fractional Brownian processes $MBF_H(t), t \in \mathbb{R}$ and $0 < H < 1$. They generalize the Brownian movement whose derivative is white noise. The increases $MBF_H(t) - MBF_H(s)$ are centered, Gaussian, with a variance $\sigma^2 |t-s|^{2H}$, auto-resembling and not independent except for $H = \frac{1}{2}$. The value of H is an indicator of the irregularity of the trajectories.

According to many authors, new signals are being introduced: in finance, in the modeling of rough surfaces, in the description of the borders of the cancerous cells, in the study of edges of clouds, etc. These signals present irregularities with varied scales. Wavelets are used for the analysis and synthesis of such signals and appear as tools for the creation of signals and surfaces as well as tools for estimating their characteristics (see, for example, the syntheses in the two articles by Bardet *et al.* in [DOU 03]).

The generators are based on random series representations of wavelet coefficients whose properties are rather simple. For the estimators, two ideas are involved. The first stems from the behavior of the wavelet Fourier transform at very low frequencies, which tends towards 0. Since the wavelet filters the signal, it destroys the spectrum of process increments, which tends towards infinity according to a frequency like $|\omega|^{-2H}$ for a wavelet with a sufficient number of zero moments. The coefficients of decomposition, level by level, have very regular spectra and without bizarre behavior in the vicinity of frequency 0. These are much more pleasant values to handle than the trajectory itself.

The second idea (see the article of Abry *et al.* in [DOU 03]) stems from the expression of the variance of the coefficients according to the level. The law of evolution is such that the logarithm of the variance is a function affine of the unknown parameter H. Estimating the variances for several scales is used to build a good estimator of H.

9.13.4. *Forecasting*

Wavelets are used in [REN 01] to construct a predictor of signals recorded in time.

The Haar wavelet is the basic tool and ensures the decomposition of the SWT type. Each scale is forecast independently after the adjustment of an AR, a statistical linear equation model of order 1 recurrence. These models are forecast easily and well. More precisely, in each scale j we adjust an AR model (to detail and approximation coefficients, in fact with a sub-sampled family) whose order $p(j)$ can depend on the scale, we forecast the coefficients and then reconstruct the signals. The total forecast is the sum of the forecasts by scale.

The wavelet transform presents the following interests:

– it decorrelates quite well even the signals whose correlation presents a long range effect;

– the Haar wavelet can be used without pre-empting future measurements.

The authors regard this method as effective.

9.13.5. *Interpolation by kriging*

By efficiently coding different scale characteristics, wavelets take part in the resolution of interpolation problems. Although they do not play the central role

therein, they are used efficiently. Kriging (from the name of a South-African geologist) is one of the statistical procedures of interpolation of surfaces (functions of $\mathbb{R}^2$ in $\mathbb{R}$) in geo-statistics [DEM 01] or for functions of $\mathbb{R}^p$ ($p \geq 2$) in $\mathbb{R}$. When information contains a lot of noise or when the variations are hardly known, kriging is used competitively with other methods, such as splines.

It is the case considered in the article by Jin *et al.* [JIN 00], which presents a work with values of p of around 10 in industrial applications. This is about the designing of the shapes of the nose of an airplane, car engines or calculating crashes. The object to be designed is represented by a computer program associating the value of a characteristic $f(x)$ to the design parameters x. The time needed for one calculation is enormous, for example, around a day. An experimental design determines, say, 125 points x_i, in which the $Z(x_i)$ calculations are carried out.

The model is $Z(x_i) = f(x_i) + \varepsilon(x_i)$. All in all, some x_i and $Z(x_i)$ are measured or calculated but f and ε are not. Moreover, for different x_i, the model dictates that the associated $\varepsilon(x_i)$ are not independent, making it difficult to estimate f. It also goes through a stage of modeling of the dependences. The technique, in a very summarized manner, is effective when the deterministic part has the shape $f(x) = \sum_{j,k} a_{j,k}\psi_{j,k}(x)$, where the summation relates to a sufficiently small set of well selected indices [DEM 01]. The estimate $\hat{a}_{j,k}$ of the $a_{j,k}$ coefficients using $Z(x_i)$ leads to the estimate $\hat{f}(x) = \sum_{j,k} \hat{a}_{j,k}\psi_{j,k}(x)$.

Appendix

The EZW Algorithm

This appendix supplements section 8.6.3.2 dedicated to the EZW algorithm used in wavelet image compression. It details the operation of the algorithm and presents an example of application. We successively tackle coding and decoding (we note by $\lfloor x \rfloor$ the integer part of x).

For other insights on the EZW algorithm, see [CRE 97], [USE 01] and [VAL 99]. In the book by Strang and Nguyen (see [STR 96], p. 362-383), we will find a detailed presentation of the various operations related to compression: quantization, coding, etc.

A.1. Coding

A.1.1. *Detailed description of the EZW algorithm (coding phase)*

(1) Initialization. All the coefficients are placed on the principal list and the threshold is initialized by $T_0 = 2^{\lfloor \log 2(C_{\max}) \rfloor}$, where $C_{\max}$ is the maximum of the absolute value of the coefficients.

(2) Principal stage. Go through the coefficients on the principal list in an appropriate order (see Figure 8.45) and compare each of them with the current threshold T_n. Attribute one of the four following symbols to each coefficient:

– *P*, if it is positive and has an absolute value higher than the threshold;

– *N*, if it is negative and has an absolute value higher than the threshold;

– *Z*, if its absolute value is lower than the threshold but if one of its children has an absolute value higher than the threshold;

– *T* if its absolute value is lower than the threshold and if all its children have absolute values lower than the threshold.

Thanks to the zerotree property each coefficient that is the child of a coefficient already coded by a *T* is not coded because the decoder can determine its value.

Each significant coefficient, i.e. of the *P* or *N* type is placed on the secondary list and is replaced by 0 in the principal list.

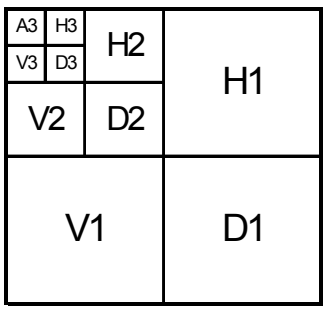
The list of symbols is coded in the most efficient manner possible (for example, by a Huffman coding).

(3) Secondary stage. For each coefficient *C* of the secondary list produce a 0 or a 1. To this end write $d = C - \sum_{j \le n} T_j$ and compare d to the middle of the interval $\left[T_n, T_n + T_{n-1}\right]$. The produced code is 0 if $d \in \left[T_n, T_n + \frac{1}{2}T_{n-1}\right[$ and 1 if $d \in \left[T_n + \frac{1}{2}T_{n-1}, T_n + T_{n-1}\right]$. When binary thresholding is applied, testing the position of d is reduced to a comparison of bits. If the paths are numbered starting from 1 and at the p^{th} path of the coefficients it is necessary to extract the $(p+1)^{th}$ bit (starting from the left) from the binary expression of *C*.

(4) New threshold. We read or calculate the new threshold with $T_{n+1} = \left\lfloor \frac{1}{2}T_n \right\rfloor$, for example. If the minimum threshold or the desired compression ratio is attained, we stop; if not, we repeat stages 2, 3 and 4.

A.1.2. *Example of application of the EZW algorithm (coding phase)*

Let us consider the following simple example:



57	-35	52	8	5	12	-10	5
-29	25	14	-14	3	1	5	-2
15	15	2	-9	8	-10	6	12
-10	-6	-11	7	7	-2	4	5
-2	12	-1	47	6	6	-2	3
0	3	-2	1	1	-4	3	1
0	-4	8	-4	4	5	3	3
5	14	4	3	-2	5	-4	1

(1) Initialization. Cmax = 57, thus $T = T_0 = 2^5 = 32$.

(2) Principal stage. Path no. 1

– A3 (1,1) = 57 > 32 gives the symbol *P* and 57 is placed on the secondary list and replaced by a zero;

– H3 (1,1) = –35 gives the symbol *N* since |–35| > 32 and 35 is placed on the secondary list and replaced by a zero;

– V3 (1,1) = –29 gives *Z* because a child (V1 (1,4) = 47) is higher than the threshold;

– D3 (1,1) = 25 gives *T* because no child (in D2 and D1) is higher than the threshold. It is thus useless to code D2 and D1.

Thus, [A3, H3, V3, D3] gives ***PNZT*.**

Continuing in a similar manner, H2 gives *PTTT* (*P* corresponds to the value H2 (1,1) = 52), V2 gives *TZTT* (*Z* corresponds to the value V2 (1,2) = 15 since its child V1 (1,4) = 47) and D2 is not coded since D3 is a zerotree). The coding process continues and the complete code obtained is:

PNZT PTTT TZTT TTTT TPTT

At the end of this main stage, the secondary list contains the values [57 35 52 47].

(3) Secondary stage. Path no. 1

The elements of the secondary list are compared to the middle of the interval $[T, 2T] = [32, 64]$, i.e. at 48. For the 1st path of the coefficients it is necessary to extract the 2nd bit (starting from the left) from the binary expression of *C*.

Secondary list		
symbol	binary	2nd bit
57	111001	1
35	100011	0
52	110100	1
47	101111	0

produced code = **1 0 1 0**

After the first loop of the algorithm, the matrix of the coefficients (principal list) thus becomes:

0	0	0	8	5	12	-10	5
-29	25	14	-14	3	1	5	-2
15	15	2	-9	8	-10	6	12
-10	-6	-11	7	7	-2	4	5
-2	12	-1	0	6	6	-2	3
0	3	-2	1	1	-4	3	1
0	-4	8	-4	4	5	3	3
5	14	4	3	-2	5	-4	1

(4) New threshold. $T = \frac{1}{2}T = 16$.

(5) Principal stage. Path no. 2

We obtain the codes ZTNPTTTTTTTT and the secondary list becomes:

[57 35 52 47 29 25].

(6) Secondary stage. Path no. 2

We compare (see section A.1.1 point 3) the elements [57 35 52 47 29 25] to 16. We obtain *100111*. They are the 3rd bits (starting from the left) of the binary expressions.

(7) End of the algorithm.

The complete application of the coding phase of the algorithm in this example is given below. It should be noted that the secondary stage (S6) is not present because the threshold is already lower than 1.

(P1): PNZTPTTTTZTTTTTTTPTT
(S1): 1010
(P2): ZTNPTTTTTTTT

(S2): 100111
(P3):ZZZZZPPNPPNZTNNTTPTTNTTTPNTTTPTTTPTTTTTTTTTPPTTTTTTTTTTT
(S3): 001110011110001000111 0
(P4):ZZZZZZZZTTZNZTZPPTTTTPPTTTPTPTPPTNPTTNPTPPTNPPTPTTNT
(S4): 0101000111110101010010110001100000001100000
(P5):ZZZZZZZTZZTZPZZZTTPTTTTNTTTNNTTPTTNTTTTPTTTTNPPTTTNTPPTT
(S5): 1101110001101100000000001111100101000000110010001010110 11
(P6): ZZZZZTTTTZTTZZTZTTTPNTTPTTPTTTTPTTTP

The sequence of codes [P1], [P1, S1], [P1, S1, P2], [P1, S1, P2, S2], etc. constitutes a succession of increasingly precise quantized coefficient approximations.

A.2. Decoding

A.2.1. *Detailed description of the EZW algorithm (decoding phase)*

The decoding phase consists of reversing the above stages.

(1) Initialization. Initialize the values of the coefficients by zeros, the list of coefficients processed by an empty list and the threshold by $T_0 = 2^{\lfloor \log 2(C_{\max}) \rfloor}$.

(2) Principal stage. Read the codes of the principal list one by one. Each code is one of the four symbols [*P, N, Z, T*]:

– if the code is *P*, assign T_n as the value of the coefficient and add the coefficient (its position) to the list of processed coefficients;

– if the code is *N*, assign $-T_n$ as the value of the coefficient and add the coefficient to the list of processed coefficients;

– if the code is *Z* or *T*, do nothing.

(3) Secondary stage. Read the codes of the secondary list one by one. It corresponds to the list of already processed coefficients. For each of these coefficients, if the corresponding term of the secondary list is 0, do nothing, whereas if this term is 1 modify its value. If the coefficient is positive, $\frac{1}{2}T_{n-1}$ is added to it whereas if it is negative the same quantity is subtracted from it.

(4) New threshold. We read or calculate the new threshold with $T_{n+1} = \left\lfloor \frac{1}{2} T_n \right\rfloor$, for example, and repeat stages 2, 3 and 4 until the list of codes is exhausted.

A.2.2. *Example of application of the EZW algorithm (decoding phase)*

The decoding phase in the example used for coding gives:

48	-32	48	0	0	0	0	0
-16	16	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	32	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

P1 + S1 + P2

56	-32	48	0	0	0	0	0
-24	24	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	40	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

P1 + S1 + P2 + S2

32	-32	32	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	32	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

P1

48	-32	48	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	32	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

P1 + S1

A.3. Visualization on a real image of the algorithm's decoding phase

In Figure A.1 we visualize some stages of the decoding phase on a real image in order to progressively highlight the improvement of the resolution with decoding. We decompose an image of size 64×64 at level 5 with the *bior4.4* wavelet. The grayscale image contains 255 levels, its storage thus requires 4,096 bytes. The EZW algorithm uses 4 symbols for the principal stage and 2 for the secondary stage. After its use, the effective coding of all the produced symbols notably reduces the information to be stored or transmitted.

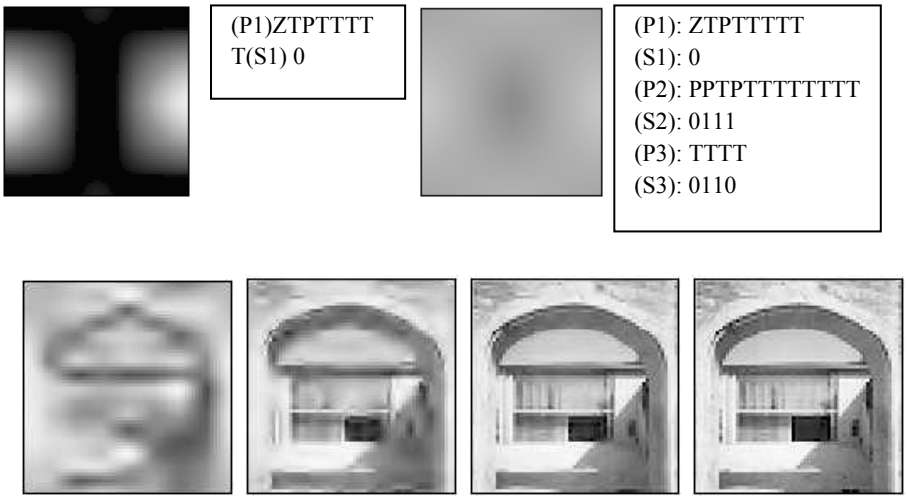


Figure A.1. *EZW algorithm: at the top, the results after 1 and 3 iterations, at the bottom, the results after 5, 7 and 9 iterations (number of code symbols 266, 2,160 and 8,179 respectively) and at the bottom on the right is the original image*

Bibliography

[ABR 95] ABRY P., ALDROUBI A., "Designing multiresolution analysis-type wavelets and their fast algorithms", *The Journal of Fourier Analysis and Applications*, vol. 2, no. 2, 1995.

[ABR 96] ABRY P., SELLAN F., "The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer – remarks and fast implementation", *Applied and Comp. Harmonic An*., vol. 3, p. 377-383, 1996.

[ABR 97] ABRY P., *Ondelettes et turbulence – Multirésolutions, algorithmes de décompositions, invariance d'échelle et signaux de pression*, Diderot Editeur, 1997.

[ABRA 00] ABRAMOVIC F., BAILEY T.C., SAPATINAS T., "Wavelet analysis and its statistical applications", *The Statistician*, vol. 49, p. 1-29, 2000.

[ADD 02] ADDISON P.S., *The Illustrated Wavelet Transform Handbook*, IoP, 2002.

[AHO 87] AHO A., HOPCROFT J., ULLMAN J., *Structures de données et algorithmes*, IIA InterEditions, p. 94-102, 1987.

[ALD 93] ALDROUBI A., UNSER M., "Families of multiresolution and wavelet spaces with optimal properties", *Numer. Funct. Anal. and Optimiz*., vol. 14, no. 5 and 6, p. 417-446, 1993.

[ANA 94] ANANT K., DOWLA F., RODRIGUE G., "Detection of the electrocardiogram P-wave using wavelet analysis", *Proceedings of the SPIE*, vol. 2 242, p. 745-749, 1994.

[ANA 95] ANANT K., DOWLA F., RODRIGUE G., "Vector Quantization of ECG Wavelet Coefficients", *IEEE Signal Processing Letters*, vol. 2, no. 7, p. 129-131, 1995.

[ANA 97] ANANT K., DOWLA F., "Wavelet Transform methods for phase identification in three-component seismograms", *Bulletin of the Seismological Society of America*, vol. 87, no. 6, p. 1598-1612, 1997.

[ANT 95] ANTONIADIS A., OPPENHEIM G. (eds.), *Wavelets and Statistics*, Springer Verlag, Lecture Notes in Statistics, no. 103, 1995.

[ANT 97] ANTONIADIS A., "Wavelets in statistics: a review (with discussion)", *J. Ital. Statist. Soc*., 6, p. 97-144, 1997.

[ANT 01] ANTONIADIS A., BIGOT J., SAPATINAS T., "Wavelet estimators in nonparametric regression: a comparative simulation study", *Journal of Statistical Software*, vol. 6, no. 6, p. 1-83, 2001.

[ARN 95] ARNEODO A., ARGOUL F., BACRY E., ELEZGARAY J., MUZY J.F., *Ondelettes, multifractales et turbulences, de l'ADN aux croissances cristallines*, Diderot Editeur, 1995.

[AZE 84] AZENCOTT R., DACUNHA-CASTELLE D., *Séries d'observations irrégulières*, Masson, 1984.

[BAI 98] BAILEY T.C., SAPATINAS T., POWELL K.J., KRZANOWSKI J., "Signal detection in underwater sound using wavelets", *JASA*, vol. 93, no. 441, p. 73-83, 1998.

[BAK 98] BAKSHI B.R., "Multiscale PCA with application to MSPC monitoring", *AIChE Journal*, vol. 44, no. 77, p. 1596-1610, 1998.

[BAK 99] BAKSHI B.R., "Multiscale Analysis and Modeling using Wavelets", *Journal of Chemometrics*, vol. 13, p. 415-434, 1999.

[BIR 97] BIRGÉ L., MASSART P., "From model selection to adaptive estimation", D. POLLARD (ed.), *Festchrift for L. Le Cam*, Springer, p. 55-88, 1997.

[BOS 00] BOSQ D., KUTOYANTS Y., *Statistical Inference for Stochastic Processes*, vol. 3, no. 1 and 2, 2000.

[BRI 95] BRISLAWN C.M., "Fingerprints go digital", *Notices of the AMS*, vol. 42, p. 1278-1283, 1995.

[BUR 96] BURKE-HUBBARD B., *The World according to Wavelets*, A.K. Peters, Wellesley, 1996.

[COH 90] COHEN A., "Ondelettes, analyses multirésolutions et traitement numérique du signal", PhD Thesis, Paris IX-Dauphine University, 1990.

[COH 91] COHEN A., FROMENT J., ISTAS J., *Analyse multirésolution de signaux aléatoires*, CRAS, 1991.

[COH 92a] COHEN A., *Ondelettes et Traitement numérique du signal*, Masson, 1992.

[COH 92b] COHEN A., DAUBECHIES I., FEAUVEAU J.C., "Biorthogonal bases of compactly supported wavelets", *Commun. on Pure and Appl. Math*., no. 45, p. 485-560, 1992.

[COH 92c] COHEN A., *Biorthogonal Wavelets in Wavelets: A Tutorial in Theory and Applications*, C.K. CHUI (ed.), Academic Press, 1992.

[COH 93] COHEN A., DAUBECHIES I., VIAL P., "Wavelets bases on the interval and fast algorithms", *J. of Appl. And Comput. Harmonic Analysis*, no. 1, p. 54-81, 1993.

[COH 95] COHEN A., *Wavelets and Multiscale Signal Processing*, Chapman and Hall, 1995.

[COI 92] COIFMAN R.R., WIRCKERHAUSER M.V., "Entropy-based algorithms for best basis selection", *IEEE Trans. on Inf. Theory*, vol. 38, no. 2, p. 713-718, 1992.

[COI 95] COIFMAN R.R., DONOHO D.L., *Translation Invariant Denoising*, Springer Verlag, Lecture Notes in Statistics, no. 103, p. 125-150, 1995.

[CRE 97] CREUSERE C.D., "A new method of robust image compression based on the embedded zerotree wavelet algorithm", *IEEE Trans. Image Processing*, vol. 6, no. 10, p. 1436-1442, 1997.

[DAU 88] DAUBECHIES I., "Orthonormal basis of compactly supported wavelets", *Comm. Pure Appl. Math.*, vol. XLI, p. 909-996, 1988.

[DAU 92] DAUBECHIES I., *Ten Lectures on Wavelets*, SIAM, 1992.

[DAU 97] DAUBECHIES I., SWELDENS W., "Factoring wavelet transforms into lifting steps", *J. Fourier Anal. Appl*., vol. 4, no. 3, p. 245-267, 1998.

[DEM 01] DEMYANOV V., SOLTANI S., KANEVSKI M., CANU S., MAIGNAN M., SAVELIEVA E., TIMONIN V., PASARENKO V., "Wavelet analysis residual kriging vs. neural network residual kriging", *Stochastic Environmental Research and Risk Assessment*, vol. 15, p. 18-32, 2001.

[DEV 92] DE VORE R.A., JAWERTH B., LUCIER B.J., "Image compression through wavelet transform coding", *IEEE Trans. on Inf. Theory*, vol. 38, no. 2, p. 719-746, 1992.

[DON 94] DONOHO D.L., JOHNSTONE I.M., *Ideal Spatial Adaptation by Wavelet Shrinkage*, Biometrika, vol. 81, p. 425-455, 1994.

[DON 95a] DONOHO D.L., "Denoising by soft thresholding", *IEEE Trans. on Inf. Theory*, vol. 41, no. 3, p. 613-627, 1995.

[DON 95b] DONOHO D.L., JOHNSTONE I.M., KERKYACHARIAN G., PICARD D., "Wavelet shrinkage: asymptotia", *JRSS*, ser. B, vol. 57, 2, p. 301-369, 1995.

[DOU 03] DOUKHAN P., OPPENHEIM G., TAQQU M. (eds.), *Theory and Applications of Long-Range Dependence*, Birkhäuser, 2003.

[EAL 00] EALCZAK B. (ed.), *Wavelets in Chemistry*, Elsevier, 2000.

[EST 77] ESTEBAN D., GALLAND C., "Application of quadrature mirror filters to split-band voice coding schemes", *Proc. IEEE. Int. Conf. Signal Speech Process*., Hartford, Connecticut, p.191-195, 1977.

[FEI 00] FEILNER M., BLU T., UNSER M., "Optimizing wavelets for the analysis of MRI data", *Proc. SPIE*, vol. 4 119, *Wavelet Applications in Signal and Image Processing VIII*, San Diego, USA, July-August, 2000.

[FLA 91] FLANDRIN P., "Fractional brownian motion and wavelets", in M. FARGE, J.C.R. HUNT, J.C. VASSILICOS (eds.), *Wavelets, Fractals and Fourier Transforms, New Developments and New Applications*, Oxford Univ. Press, 1991.

[FOU 94] FOUFOULA-GEORGIOU E., KUMAR P. (eds.), *Wavelets: Theory, Algorithms and Application*, in C. CHUI (ed.), *Wavelets in Geophysics*, vol. 4, Academic Press, Wavelet Analysis and Its Applications, 1994.

[FRA 99] FRAZIER M., *An Introduction to Wavelets through Linear Algebra*, Springer Verlag, 1999.

[GEN 01] GENCAY R., SELCUK F., WHITCHER B., *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press, 2001.

[HAR 98] HÄRDLE W., KERKYACHARIAN G., PICARD D., TSYBAKOV A., *Wavelets, approximation and statistical applications*, Lecture Notes in Statistics, no. 129, Springer Verlag, 1998.

[HEI 01] HEISING G., MARPE D., CYCON H., PETUKHOV A., "Wavelet-based very low bit-rate video coding using image warping and overlapped block motion compensation", *IEEE Proceedings-Vision, Image and Signal Processing*, vol. 148, no. 2, p. 93-101, April 2001.

[HOL 90] HOLSCHNEIDER M., TCHAMITCHIAN P., "Régularité locale de la fonction 'non-différentiable de Riemann'", in [LEM 90], p. 102-124, 1990.

[HOW 92] HOWARD P.G., WITTER G.S., "Practical Implementations of Arithmetic coding in Image and Text Compression", in J.A. STORER (ed.), Kluwer Academic Publishing, p. 85-112, 1992.

[HUB 95] HUBBARD B.B., *Ondes et ondelettes: la saga d'un outil mathématique*, Pour la Science, Diffusion Belin, 1995.

[HUF 52] HUFFMAN D.A., "A method for the construction of minimum redundancy codes", *Proc. of IRE*, no. 40, p. 1098-1101, 1952.

[JIN 00] JIN R., CHEN W., SIMPSON T.W., "Comparitive studies of metamodeling techniques under multiple modeling criteria", *Journal of Structural Optimization*, vol. 23, no. 1, p.1-13, 2001.

[JPE 00] JPEG 2000, BOLIEK M., CHRISTOPOULOS C., MAJANI E. (eds.), *Coding of Still Pictures*, 2000.

[KAH 98] KAHANE J.P., LEMARIE-RIEUSSET P.G., *Séries de Fourier et ondelettes*, Nouvelle bibliothèque Mathématique Cassini, 1998.

[KAI 94] KAISER G., *A Friendly Guide to Wavelets*, Birkhäuser, 1994.

[KOB 98] KOBAYASHI M. (ed.), *Applications of Wavelets: Case Studies*, SIAM, 1998.

[KOL 74] KOLMOGOROV V.A., FOMINE S., *Eléments de la théorie des fonctions et de l'analyse fonctionnelle*, Editions de Moscou, 1974.

[LAV 99] LAVIELLE M., "Detection of multiple changes in a sequence of dependent variables", *Stoch. Proc. And their Applications*, vol. 83, no. 2, p.79-102, 1999.

[LEM 90] LEMARIE P.G. (eds.), *Les ondelettes en 1989*, p. 1438, Springer Verlag, Lecture Notes in Math., 1990.

[LIT 94] LITTLE S.A., "Wavelet Analysis of Seafloor Bathymetry: an example", in *Wavelets in Geophysics*, FOUFOULA-GEORGIOU E., KUMAR P. (eds.), Academic Press, p.167-182, 1994.

[LIU 94] LIU P.C., "Wavelet spectrum analysis and ocean wind", in *Wavelets in Geophysics*, FOUFOULA-GEORGIOU E., KUMAR P. (ed.) Academic Press, p. 151-166, 1994.

[MAD 92] MADYCH W.R., "Some elementary properties of multiresolution analyses of $L^2(\mathbb{R}^n)$", in *Wavelets: A tutorial in Theory and Applications*, C.K. CHUI (ed.), Academic Press, 1992.

[MAL 89] MALLAT S., "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Trans. on PAMI*, vol. 2, no. 7, p. 674-693, 1989.

[MAL 98] MALLAT S., *A Wavelet Tour of Signal Processing*, Academic Press, 2000.

[MAN 68] MANDELBROT B., VAN NESS J., "Fractional brownian motions, fractional noises and applications", *SIAM Review*, no. 10, p. 422-437, 1968.

[MAR 99] MARPE D., HEISING G., CYCON H., "An efficient very low bit rate video coding scheme using warping prediction and wavelet-based residual coding", *Picture Coding Symposium (PCS'99)*, p. 21-24, Portland, Oregon, USA, April 1999.

[MAT 90] MATLAB, MATHWORKS INC., *MATLAB Reference Manual*, 1990.

[MES 05] MESA H., "Adapted wavelets for pattern detection", in A. SANFELIU, M. LAZO CORTEZ (eds.), *CIARP, Lecture Notes in Computer Science*, p. 933-944, Springer, 2005.

[MET 01] METIN AKAY M. (ed.), *Time-Frequency and Wavelets in Biomedical Signal Processing,* IEEE Press, Series one Biomedical Engineering, 2001.

[MEY 90] MEYER Y., "Ondelettes et opérateurs", vol. 1, Hermann, *Actualités mathématiques*, 1990.

[MEY 93] MEYER Y., *Les ondelettes. algorithmes et applications*, Paris, Colin, 1993.

[MEY 93] MEYER Y., ROQUES S. (eds.), *Progress in Wavelet Analysis and Applications*, Editions Frontières, 1993.

[MEY 99] MEYER Y., SELLAN F., TAQQU M., "Wavelets, generalized white noise and fractional integration: the synthesis of fractional brownian motion", *Journal of Fourier Analysis and Applications*, vol. 5, no. 5, p. 465-494, 1999.

[MIN 85] MINTZER F., "Filters for distorsion-free two-band multirate filter banks", *IEEE Trans. Acous. Signal Process.*, no. 33, p. 626-630, 1985.

[MIS 93] MISITI M., MISITI Y., OPPENHEIM G., POGGI J.-M., "Analyse de signaux classiques par décomposition en ondelettes", *Revue de Statistique Appliquée*, vol. XLI, no. 4, p. 5-32, 1993.

[MIS 94] MISITI M., MISITI Y., OPPENHEIM G., POGGI J.-M., "Décomposition en ondelettes et méthodes comparatives: étude d'une courbe de charge électrique", *Revue de Statistique Appliquée*, vol. XLII, no. 2, p. 57-77, 1994.

[MIS 98] MISITI M., MISITI Y., OPPENHEIM G., POGGI J.-M., "Méthodes d'ondelettes en statistiques: introduction et exemples", *Journal de la SFDS*, vol. 139, no. 4, p. 3-29, 1998.

[MIS 00] MISITI M., MISITI Y., OPPENHEIM G., POGGI J.-M., *Wavelet Toolbox User's Guide*, The MathWorks, 2000.

[NAS 95] NASON G.P., SILVERMAN B.W., *The Stationary Wavelet Transform and Some Statistical Applications*, Springer Verlag, Lecture Notes in Statistics, no. 103, p. 281-299, 1995.

[OGD 97] OGDEN T., *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser, 1997.

[PAP 01] PAPANICOLAOU G., SOLNA K., "Wavelet-based estimation of local Kolmogorov turbulence", in *Long Range Dependence*, DOUKHAN P., OPPENHEIM G., TAQQU M. (eds.), Birkhäuser, 2001.

[PER 94] PERCIVAL D.B., GUTTORP P., "*Long-memory processes, the Allen Variance and Wavelets*", in *Wavelets in Geophysics*, FOUFOULA-GEORGIOU E. and KUMAR P. (eds.), p.325-344, 1994.

[PER 00] PERCIVAL D.B., WALDEN A.T., *Wavelet Methods for Time Series Analysis*, Cambridge University Press, 2000.

[PES 96] PESQUET J.C., KRIM H., CARFATAN H., "Time-invariant orthonormal wavelet representations", *IEEE Trans. Sign. Proc.*, vol. 44, no. 8, p. 1964-1970, 1996.

[RAS 97] RASHID-FARROKHI K.J.R., LIU C., BERENSTEIN A., WALNUT D., "Wavelet-based multiresolution local tomography", *IEEE Transactions on Image Processing*, no. 6, p. 1412-1430, 1997.

[REN 01] RENAUD O., STARCK J.-L., MURTAGH F., *Prediction based on a Multiscale Decomposition*, Preprint p. 1-21, submitted, 2001.

[SHA 93] SHAPIRO J., "Embedded image coding using zerotrees of wavelet coefficients", *IEEE Trans. Signal Processing*, vol. 41, p. 3445-3462, December 1993.

[SKO 01] SKODRAS A.N., CHRISTOPOULOS C., EBRAHIMI T., "JPEG 2000 still image compression standard", *IEEE Signal Processing Magazine*, vol. 18, no.°5, p. 36-58, September 2001.

[SMI 86] SMITH M.J.T., BARNWELL T.P., "Exact reconstruction techniques for tree-structured suband coders", *IEEE Trans. Acous. Signal Speech Process*., no. 34, p. 434-441, 1986.

[STR 89] STRANG G., "Wavelets and dilations equations: a brief introduction", *SIAM Review*, vol. 31, p. 614-627, 1989.

[STR 96] STRANG G., NGUYEN T., *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Boston, 1996.

[SWE 98] SWELDENS W., "The lifting scheme: a construction of second generation of wavelets", *SIAM J. Math Anal*., vol. 29, no. 2, p. 511-546, March 1998.

[TAT 95] TATE R., WATSON D. EGLEN S., "Using wavelets for classifying human in vivo magnetic resonance spectra", in ANTONIADIS A., OPPENHEIM G. (eds.), *Wavelets and Statistics*, Springer Verlag, Lecture Notes in Statistics, no. 103, p. 377-383, 1995.

[TAT 96] TATE R., "Pattern recognition analysis of *in vivo* magnetic resonance spectra", PhD Thesis, University of Sussex, 1996.

[TEO 98] TEOLIS A., *Computational Signal Processing with Wavelets*, Birkhäuser, 1998.

[TOP 98] TOPIWALA P.N., *Wavelet Image and Video Compression*, Kluwer, 1998.

[TOR 95] TORRÉSANI B., *Analyse continue par ondelettes*, CNRS Edition, 1995.

[TRI 92] TRIEBEL H., *Theory of Function Spaces II*, Birkhäuser Verlag, Basel, 1992.

[TRY 01] TRYGG J., KETTANEH-WOLD N., WALLBÄCKS L., "2D Wavelet analysis and compression of on-line industrial process data", *J. Chemometrics* , no. 15, p. 299-319, 2001.

[UNS 96] UNSER M., ALDROUBI A., "A review of wavelets in biomedical applications", *Proceedings of the IEEE*, vol. 84, no. 4, p. 626-638, 1996.

[USE 01] USEVITCH B.E., "A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000", *IEEE Signal Processing Magazine*, vol. 18, no. 5, p. 22-35, September 2001.

[VAL 99] VALENS C., *Embedded Zerotree Wavelet Encoding*, c.valens@mindless.com 1999.

[VAN 99] VAN DEN BERG J.C. (ed.), *Wavelets in Physics*, Cambridge University Press, 1999.

[VET 86] VETTERLI M., "Filter banks allowing perfect reconstruction", *Signal Process*., no. 10, p. 219-244, 1986.

[VID 99] VIDAKOVIC B., *Statistical Modeling by Wavelets*, Wiley, 1999.

[WAL 02] WALNUT D.F., *An Introduction to Wavelet Analysis*, Birkhäuser, 2002.

[WIC 94] WICKERHAUSER M.V., *Adapted Wavelet Analysis from Theory to Software*, A.K. Peters, Wellesley, Massachusetts, 1994.

[WIT 87] WITTENS I., NEAL R., CLEARY J., "Arithmetic coding for data compression", *Comm. of the ACM*, vol. 30, no. 6, p. 519-540, 1987.

[ZEE 98] ZEEUW P.M., "Wavelets and Image Fusion", *CWI*, Amsterdam, March 1998, http://www.cwi.nl/~pauldz.

Index